



Nuclei Microcontroller Software Interface Standard

**NMSIS**

***Release 1.0.2***

**Nuclei**

**Sep 09, 2021**



## CONTENTS:

<b>1</b>	<b>Nuclei MCU Software Interface Standard(NMSIS)</b>	<b>1</b>
1.1	About NMSIS	1
1.2	NMSIS Components	1
1.3	NMSIS Design	1
1.4	How to Access	2
1.5	Coding Rules	2
1.6	Validation	3
1.7	License	3
<b>2</b>	<b>NMSIS Core</b>	<b>5</b>
2.1	Overview	5
2.1.1	Introduction	5
2.1.2	Processor Support	5
2.1.3	Toolchain Support	6
2.2	Using NMSIS in Embedded Applications	6
2.2.1	Introduction	6
2.2.2	Basic NMSIS Example	9
2.2.3	Using Interrupt and Exception/NMI	10
2.2.4	Using NMSIS with generic Nuclei Processors	10
2.2.5	Create generic Libraries with NMSIS	11
2.3	NMSIS-Core Device Templates	12
2.3.1	Introduction	12
2.3.2	NMSIS-Core Processor Files	12
2.3.3	Device Examples	13
2.3.4	Template Files	13
2.3.5	Adapt Template Files to a Device	13
2.3.6	Device Templates Explanation	14
2.4	Register Mapping	57
2.5	NMSIS Core API	57
2.5.1	Version Control	57
2.5.2	Compiler Control	59
2.5.3	Core CSR Register Access	60
2.5.4	Core CSR Encoding	62
2.5.5	Register Define and Type Definitions	78
2.5.6	CPU Intrinsic Functions	93
2.5.7	Intrinsic Functions for SIMD Instructions	99
2.5.8	Peripheral Access	356
2.5.9	Systick Timer(SysTimer)	357
2.5.10	Interrupts and Exceptions	363
2.5.11	FPU Functions	382

2.5.12	PMP Functions . . . . .	385
2.5.13	Cache Functions . . . . .	387
2.5.14	System Device Configuration . . . . .	405
2.5.15	ARM Compatible Functions . . . . .	409
<b>3</b>	<b>NMSIS DSP . . . . .</b>	<b>413</b>
3.1	Overview . . . . .	413
3.1.1	Introduction . . . . .	413
3.1.2	Using the Library . . . . .	413
3.1.3	Examples . . . . .	413
3.1.4	Toolchain Support . . . . .	414
3.1.5	Building the Library . . . . .	414
3.1.6	Preprocessor Macros . . . . .	414
3.2	Using NMSIS-DSP . . . . .	414
3.2.1	Preparation . . . . .	414
3.2.2	Tool Setup . . . . .	414
3.2.3	Build NMSIS DSP Library . . . . .	414
3.2.4	How to run . . . . .	415
3.3	NMSIS DSP API . . . . .	416
3.3.1	Examples . . . . .	416
3.3.2	Basic Math Functions . . . . .	430
3.3.3	Bayesian estimators . . . . .	451
3.3.4	Complex Math Functions . . . . .	452
3.3.5	Controller Functions . . . . .	461
3.3.6	Distance functions . . . . .	471
3.3.7	Fast Math Functions . . . . .	480
3.3.8	Filtering Functions . . . . .	485
3.3.9	Interpolation Functions . . . . .	557
3.3.10	Matrix Functions . . . . .	563
3.3.11	Quaternion Math Functions . . . . .	588
3.3.12	Statistics Functions . . . . .	592
3.3.13	Support Functions . . . . .	610
3.3.14	SVM Functions . . . . .	622
3.3.15	Transform Functions . . . . .	630
3.4	Changelog . . . . .	673
3.4.1	V1.0.2 . . . . .	673
3.4.2	V1.0.1 . . . . .	674
3.4.3	V1.0.0 . . . . .	674
<b>4</b>	<b>NMSIS NN . . . . .</b>	<b>675</b>
4.1	Overview . . . . .	675
4.1.1	Introduction . . . . .	675
4.1.2	Block Diagram . . . . .	675
4.1.3	Examples . . . . .	675
4.1.4	Pre-processor Macros . . . . .	676
4.2	Using NMSIS-NN . . . . .	676
4.2.1	Preparation . . . . .	676
4.2.2	Tool Setup . . . . .	676
4.2.3	Build NMSIS NN Library . . . . .	676
4.2.4	How to run . . . . .	677
4.3	NMSIS NN API . . . . .	678
4.3.1	Neural Network Functions . . . . .	678
4.3.2	Neural Network Data Conversion Functions . . . . .	724
4.3.3	Basic Math Functions for Neural Network Computation . . . . .	726

4.3.4	Convolutional Neural Network Example . . . . .	733
4.3.5	Gated Recurrent Unit Example . . . . .	734
4.4	Changelog . . . . .	735
4.4.1	V1.0.2 . . . . .	735
4.4.2	V1.0.1 . . . . .	735
4.4.3	V1.0.0 . . . . .	735
<b>5</b>	<b>Changelog</b>	<b>737</b>
5.1	V1.0.2 . . . . .	737
5.2	V1.0.2-RC2 . . . . .	737
5.3	V1.0.2-RC1 . . . . .	738
5.4	V1.0.1 . . . . .	738
5.5	V1.0.1-RC1 . . . . .	738
5.6	V1.0.0-beta1 . . . . .	739
5.7	V1.0.0-beta . . . . .	739
5.8	V1.0.0-alpha.1 . . . . .	740
5.9	V1.0.0-alpha . . . . .	740
<b>6</b>	<b>Glossary</b>	<b>741</b>
<b>7</b>	<b>Appendix</b>	<b>743</b>
<b>8</b>	<b>Indices and tables</b>	<b>745</b>
	<b>Index</b>	<b>747</b>



---

## NUCLEI MCU SOFTWARE INTERFACE STANDARD(NMSIS)

### 1.1 About NMSIS

The **NMSIS** is a vendor-independent hardware abstraction layer for micro-controllers that are based on [Nuclei Processors](#)<sup>1</sup>.

The **NMSIS** defines generic tool interfaces and enables consistent device support. It provides simple software interfaces to the processor and the peripherals, simplifying software re-use, reducing the learning curve for micro-controller developers, and reducing the time to market for new devices.

### 1.2 NMSIS Components

**NMSIS CORE** All Nuclei N/NX Class Processors Standardized API for the Nuclei processor core and peripherals.

**NMSIS DSP** All Nuclei N/NX Class Processors DSP library collection with a lot of functions for various data types: fixed-point (fractional q7, q15, q31) and single precision floating-point (32-bit). Implementations optimized for the Nuclei Processors which has RISC-V SIMD instruction set.

**NMSIS NN** All Nuclei N/NX Class Processors Collection of efficient neural network kernels developed to maximize the performance and minimize the memory footprint Nuclei processor cores.

### 1.3 NMSIS Design

**NMSIS** is designed to help the Nuclei N/NX Class Processors processors in standardization. It enables consistent software layers and device support across a wide range of development tools and micro-controllers.

**NMSIS** is a lightweight software interface layer that tried to standardize common Nuclei processor-based SOC, and it didn't define any standard peripherals. The silicon industry can therefore support the wide variations of Nuclei processor-based devices with this common standard.

We can achieve the following benefits of **NMSIS**:

- **NMSIS** reduces the learning curve, development costs, and time-to-market. Developers can write software quicker through a variety of easy-to-use, standardized software interfaces.
- Consistent software interfaces improve the software portability and re-usability. Generic software libraries and interfaces provide consistent software framework.
- It provides interfaces for debug connectivity, debug peripheral views, software delivery, and device support to reduce time-to-market for new micro-controller deployment.

---

<sup>1</sup> [https://doc.nucleisys.com/nuclei\\_spec](https://doc.nucleisys.com/nuclei_spec)



Fig. 1: NMSIS Design Diagram

- Being a compiler independent layer, it allows to use the compiler of your choice. Thus, it is supported by mainstream compilers.
- It enhances program debugging with peripheral information for debuggers.

## 1.4 How to Access

If you want to access the code of **NMSIS**, you can visit our opensource [NMSIS Github Repository](https://github.com/Nuclei-Software/NMSIS)<sup>2</sup>.

## 1.5 Coding Rules

The **NMSIS** uses the following essential coding rules and conventions:

- Compliant with ANSI C (C99) and C++ (C++03).
- Uses ANSI C standard data types defined in **stdint.h**.
- Variables and parameters have a complete data type.
- Expressions for *#define* constants are enclosed in parenthesis.

In addition, the **NMSIS** recommends the following conventions for identifiers:

- **CAPITAL** names to identify Core Registers, Peripheral Registers, and CPU Instructions.
- **CamelCase** names to identify function names and interrupt functions.
- **Namespace\_** prefixes avoid clashes with user identifiers and provide functional groups (i.e. for peripherals, RTOS, or DSP Library).

The **NMSIS** is documented within the source files with:

<sup>2</sup> <https://github.com/Nuclei-Software/NMSIS>



- Comments that use the C or C++ style.
- Doxygen compliant comments, which provide:
  - brief function, variable, macro overview.
  - detailed description of the function, variable, macro.
  - detailed parameter explanation.
  - detailed information about return values.

## 1.6 Validation

Nuclei uses RISC-V GCC Compiler in the various tests of **NMSIS**, and if more compiler is added, it could be easily supported by following the **NMSIS** compiler independent layer. For each component, the section **Validation** describes the scope of the various verifications.

**NMSIS** components are compatible with a range of C and C++ language standards.

As **NMSIS** defines API interfaces and functions that scale to a wide range of processors and devices, the scope of the run-time test coverage is limited. However, several components are validated using dedicated test suites.

## 1.7 License

This **NMSIS** is modified based on open-source project **CMSIS** to match Nuclei requirements.

This **NMSIS** is provided free of charge by Nuclei under the [Apache 2.0 License](http://www.apache.org/licenses/LICENSE-2.0)<sup>3</sup>.

---

<sup>3</sup> <http://www.apache.org/licenses/LICENSE-2.0>



## NMSIS CORE

## 2.1 Overview

### 2.1.1 Introduction

**NMSIS-Core** implements the basic run-time system for a Nuclei N/NX Class Processors based device and gives the user access to the processor core and the device peripherals. In detail it defines:

- **Hardware Abstraction Layer (HAL)** for Nuclei processor registers with standardized definitions for the **CSR Registers, TIMER, ECLIC, PMP Registers, DSP Registers, FPU registers, and Core Access Functions**.
- **Standard core exception/interrupt names** to interface to system exceptions or interrupts without having compatibility issues.
- **Methods to organize header files** that makes it easy to learn new Nuclei micro-controller products and improve software portability. This includes naming conventions for device-specific interrupts.
- **Methods for system initialization** to be used by each Device vendor. For example, the standardized *SystemInit()* (page 406) function is essential for configuring the clock system of the device.
- **Intrinsic functions** used to generate CPU instructions that are not supported by standard C functions.
- A variable *SystemCoreClock* (page 406) to determine the **system clock frequency** which simplifies the setup the timer.

The following sections provide details about the **NMSIS-Core**:

- *Using NMSIS in Embedded Applications* (page 6) describes the project setup and shows a simple program example
- *NMSIS-Core Device Templates* (page 12) describes the files of the *NMSIS Core* (page 5) in detail and explains how to adapt template files provided by Nuclei to silicon vendor devices.
- *NMSIS Core API* (page 57) describe the features and functions of the *Device Header File <device.h>* (page 47) in detail.
- *Register Define and Type Definitions* (page 78) describe the data structures of the *Device Header File <device.h>* (page 47) in detail.

### 2.1.2 Processor Support

NMSIS have provided support for all the Nuclei N/NX Class Processors.

**Nuclei ISA Spec:**

- Nuclei Process Core Instruction Set Architecture Spec<sup>4</sup>

**Nuclei N Class Processor Reference Manuals:**

- N200 series<sup>5</sup>
- N300 series<sup>6</sup>
- N600 series<sup>7</sup>

**Nuclei NX Class Processor Reference Manuals:**

- NX600 series<sup>8</sup>

## 2.1.3 Toolchain Support

The *NMSIS-Core Device Templates* (page 12) provided by Nuclei have been tested and verified using these toolchains:

- GNU Toolchain for RISC-V modified by Nuclei

## 2.2 Using NMSIS in Embedded Applications

### 2.2.1 Introduction

To use the **NMSIS-Core**, the following files are added to the embedded application:

- *Startup File* `startup_<device>.S` (page 14), which provided asm startup code and vector table.
- *Interrupt and Exception Handling File*: `intexc_<device>.S` (page 21), which provided general exception handling code for non-vector interrupts and exceptions.
- *Device Linker Script*: `gcc_<device>.ld` (page 31), which provided linker script for the device.
- *System Configuration Files* `system_<device>.c` and `system_<device>.h` (page 36), which provided general device configuration (i.e. for clock and BUS setup).
- *Device Header File* `<device.h>` (page 47) gives access to processor core and all peripherals.

---

**Note:** The files *Startup File* `startup_<device>.S` (page 14), *Interrupt and Exception Handling File*: `intexc_<device>.S` (page 21), *Device Linker Script*: `gcc_<device>.ld` (page 31) and *System Configuration Files* `system_<device>.c` and `system_<device>.h` (page 36) may require application specific adaptations and therefore should be copied into the application project folder prior configuration.

The *Device Header File* `<device.h>` (page 47) is included in all source files that need device access and can be stored on a central include folder that is generic for all projects.

---

The *Startup File* `startup_<device>.S` (page 14) is executed right after device reset, it will do necessary stack pointer initialization, exception and interrupt entry configuration, then call `SystemInit()` (page 406), after system initialization, will return to assemble startup code and do c/c++ runtime initialization which includes data, bss section initialization, c++ runtime initialization, then it will call `main()` function in the application code.

---

<sup>4</sup> [https://doc.nucleisys.com/nuclei\\_spec](https://doc.nucleisys.com/nuclei_spec)

<sup>5</sup> <https://www.nucleisys.com/product.php?site=n200>

<sup>6</sup> <https://www.nucleisys.com/product.php?site=n300>

<sup>7</sup> <https://www.nucleisys.com/product.php?site=n600>

<sup>8</sup> <https://www.nucleisys.com/product.php?site=nx600>

In the *Interrupt and Exception Handling File: `intexc_<device>.S`* (page 21), it will contain all exception and interrupt vectors and implements a default function for every interrupt. It may also contain stack and heap configurations for the user application.

The *System Configuration Files `system_<device>.c` and `system_<device>.h`* (page 36) performs the setup for the processor clock. The variable `SystemCoreClock` (page 406) indicates the CPU clock speed. *Systick Timer(`SysTimer`)* (page 357) describes the minimum feature set. In addition the file may contain functions for the memory BUS setup and clock re-configuration.

The *Device Header File `<device.h>`* (page 47) is the central include file that the application programmer is using in the C source code. It provides the following features:

- *Peripheral Access* (page 356) provides a standardized register layout for all peripherals. Optionally functions for device-specific peripherals may be available.
- *Interrupts and Exceptions* (page 363) can be accessed with standardized symbols and functions for the **ECLIC** are provided.
- *CPU Intrinsic Functions* (page 93) allow to access special instructions, for example for activating sleep mode or the NOP instruction.
- *Intrinsic Functions for SIMD Instructions* (page 99) provide access to the DSP-oriented instructions.
- *Systick Timer(`SysTimer`)* (page 357) function to configure and start a periodic timer interrupt.
- *Core CSR Register Access* (page 60) function to access the core csr registers.
- *Cache Functions* (page 387) to access the I-CACHE and D-CACHE unit
- *FPU Functions* (page 382) to access the Floating point unit.
- *PMP Functions* (page 385) to access the Physical Memory Protection unit
- *Version Control* (page 57) which defines NMSIS release specific macros.
- *Compiler Control* (page 59) is compiler agnostic `#define` symbols for generic C/C++ source code

The NMSIS-Core system files are device specific.

In addition, the *Startup File `startup_<device>.S`* (page 14) is also compiler vendor specific, currently only GCC version is provided. The versions provided by NMSIS are only generic templates. The adopted versions for a concrete device are typically provided by the device vendor through the according device family package.

For example, the following files are provided by the **GD32VF103** device family pack:



Fig. 1: NMSIS-Core User Files

Table 1: Files provided by GD32VF103 device family pack

File	Description
./Device/Source/GCC/startup_gd32vf103.S	Startup File startup_<device>.S for the GD32VF103 device variants.
./Device/Source/GCC/intexc_gd32vf103.S	Exception and Interrupt Handling File intexc_<device>.S for the GD32VF103 device variants.
./Device/Source/GCC/gcc_gd32vf103.ld	Linker script File gcc_<device>.ld for the GD32VF103 device variants.
./Device/Source/system_gd32vf103.c	System Configuration File system_<device>.c for the GD32VF103 device families
./Device/Include/system_gd32vf103.h	System Configuration File system_<device>.h for the GD32VF103 device families
./Device/Include/gd32vf103.h	Device Header File <device.h> for the GD32VF103 device families.

**Note:** The silicon vendors create these device-specific NMSIS-Core files based on *NMSIS-Core Device Templates* (page 12) provided by Nuclei.

Thereafter, the functions described under *NMSIS Core API* (page 57) can be used in the application.

## 2.2.2 Basic NMSIS Example

A typical example for using the NMSIS layer is provided below. The example is based on a GD32VF103 Device.

Listing 1: gd32vf103\_example.c

```

1  #include <gd32vf103.h>                                // File name depends on device used
2
3  uint32_t volatile msTicks;                             // Counter for millisecond Interval
4  #define SysTick_Handler    eclic_mtip_handler
5  #define CONFIG_TICKS       (SOC_TIMER_FREQ / 1000)
6
7  void SysTick_Handler (void) {                          // SysTick Interrupt Handler
8      SysTick_Reload(CONFIG_TICKS);

```

(continues on next page)

(continued from previous page)

```

9      msTicks++;                                // Increment Counter
10  }
11
12  void WaitForTick (void) {
13      uint32_t curTicks;
14
15      curTicks = msTicks;                        // Save Current SysTick Value
16      while (msTicks == curTicks) {             // Wait for next SysTick Interrupt
17          __WFI ();                             // Power-Down until next Event/
18          ↪Interrupt
19      }
20  }
21
22  void TIMER0_UP_IRQHandler (void) {            // Timer Interrupt Handler
23      ;                                          // Add user code here
24  }
25
26  void timer0_init(int frequency) {             // Set up Timer (device specific)
27      ECLIC_SetPriorityIRQ (TIMER0_UP_IRQn, 1); // Set Timer priority
28      ECLIC_EnableIRQ (TIMER0_UP_IRQn);         // Enable Timer Interrupt
29  }
30
31  void Device_Initialization (void) {           // Configure & Initialize MCU
32      if (SysTick_Config (CONFIG_TICKS)) {
33          ; // Handle Error
34      }
35      timer0_init ();                           // setup device-specific timer
36  }
37
38  // The processor clock is initialized by NMSIS startup + system file
39  void main (void) {                            // user application starts here
40      Device_Initialization ();                 // Configure & Initialize MCU
41      while (1) {                               // Endless Loop (the Super-Loop)
42          __disable_irq ();                     // Disable all interrupts
43          Get_InputValues ();                   // Read Values
44          __enable_irq ();                      // Enable all interrupts
45          Calculation_Response ();               // Calculate Results
46          Output_Response ();                   // Output Results
47          WaitForTick ();                       // Synchronize to SysTick Timer
48      }
49  }

```

### 2.2.3 Using Interrupt and Exception/NMI

Nuclei processors provide **NMI(Non-Maskable Interrupt)**, **Exception**, **Vector Interrupt** and **Non-Vector Interrupt** features.

### 2.2.4 Using NMSIS with generic Nuclei Processors

Nuclei provides NMSIS-Core files for the supported Nuclei Processors and for various compiler vendors. These files can be used when standard Nuclei processors should be used in a project. The table below lists the folder and device names of the Nuclei processors.



Table 2: Folder and device names of the Nuclei processors

Folder	Processor	RISC-V	Description
./Device/Nuclei/NUCLEI_N	<ul style="list-style-type: none"> <li>• N200</li> <li>• N300</li> <li>• N600</li> </ul>	RV32	Contains Include and Source template files configured for the Nuclei N200/N300/N600 processor. The device name is NUCLEI_N and the name of the Device Header File <device.h> is <NUCLEI_N.h>.
./Device/Nuclei/NUCLEI_NX	NX600	RV64	Contains Include and Source template files configured for the Nuclei NX600 processor. The device name is NUCLEI_NX and the name of the Device Header File <device.h> is <NUCLEI_NX.h>.

## 2.2.5 Create generic Libraries with NMSIS

The NMSIS Processor and Core Peripheral files allow also to create generic libraries. The **NMSIS-DSP** Libraries are an example for such a generic library.

To build a generic library set the define `__NMSIS_GENERIC` and include the *nmsis\_core.h* NMSIS CPU & Core Access header file for the processor.

The define `__NMSIS_GENERIC` disables device-dependent features such as the **SysTick timer** and the **Interrupt System**.

### Example

The following code section shows the usage of the *nmsis\_core.h* header files to build a generic library for N200, N300, N600, NX600.

One of these defines needs to be provided on the compiler command line.

By using this header file, the source code can access the functions for *Core CSR Register Access* (page 60), *CPU Intrinsic Functions* (page 93) and *Intrinsic Functions for SIMD Instructions* (page 99).

Listing 2: core\_generic.h

```

1 #define __NMSIS_GENERIC    // Disable Ecltic and Systick functions
2 #include <nmsis_core.h>

```

## 2.3 NMSIS-Core Device Templates

### 2.3.1 Introduction

Nuclei supplies NMSIS-Core device template files for the all supported Nuclei N/NX Class Processors and various compiler vendors. Refer to the list of *supported toolchain* (page 6) for compliance.

**These NMSIS-Core device template files include the following:**

- Register names of the Core Peripherals and names of the Core Exception/Interrupt Vectors.
- Functions to access core peripherals, special CPU instructions and SIMD instructions
- Generic startup code and system configuration code.

The detailed file structure of the NMSIS-Core device templates is shown in the following picture.

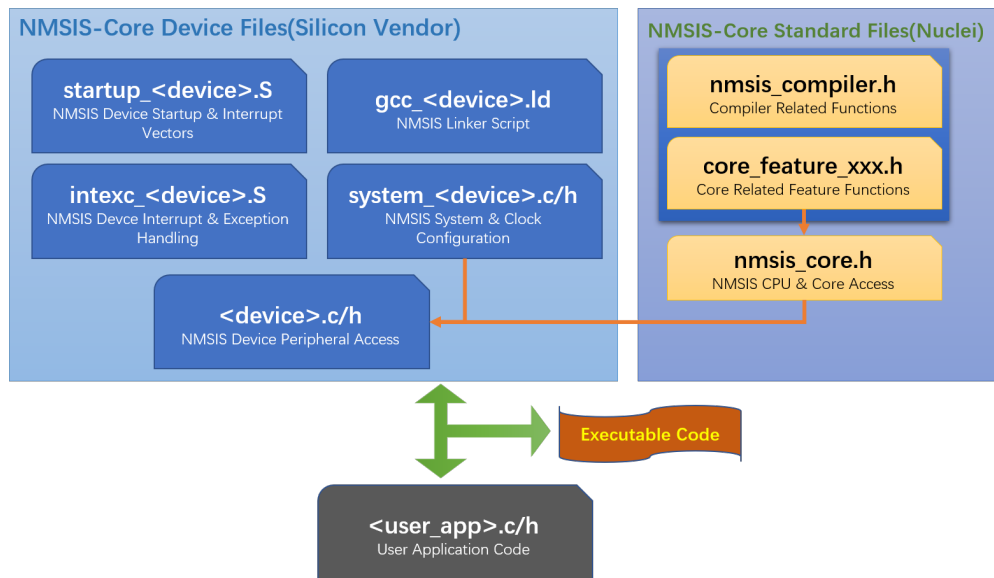


Fig. 2: NMSIS-Core Device Templates

### 2.3.2 NMSIS-Core Processor Files

The NMSIS-Core processor files provided by Nuclei are in the directory *NMSIS/Core/Include*.

These header files define all processor specific attributes do not need any modifications.

The *nmsis\_core.h* defines the core peripherals and provides helper functions that access the core registers.

### 2.3.3 Device Examples

The NMSIS Software Pack defines several devices that are based on the Nuclei N/NX processors.

The device related NMSIS-Core files are in the directory *Device/Nuclei* and include NMSIS-Core processor file explained before.

The following sample devices are defined as below:

Table 3: Device Examples of Nuclei Processor

Family	Device	Description
Nuclei N	NUCLEI N Class	Nuclei N Class based device
Nuclei NX	NUCLEI NX Class	Nuclei NX Class based device

### 2.3.4 Template Files

To simplify the creation of NMSIS-Core device files, the following template files are provided that should be extended by the silicon vendor to reflect the actual device and device peripherals.

**Silicon vendors add to these template files the following information:**

- **Device Peripheral Access Layer** that provides definitions for device-specific peripherals.
- **Access Functions for Peripherals** (optional) that provides additional helper functions to access device-specific peripherals.
- **Interrupt vectors** in the startup file that are device specific.

Table 4: NMSIS-Core Device Template Files

Template File (Under <i>./Device/_Template_Vendor/Vendor/</i> )	Description
<i>Device/Source/GCC/startup_Device.S</i>	Startup file template for GNU GCC RISC-V Embedded Compiler.
<i>Device/Source/GCC/gcc_Device.ld</i>	Link Script file template for GNU GCC RISC-V Embedded Compiler.
<i>Device/Source/GCC/intexc_Device.S</i>	Exception and Interrupt handling file template for GNU GCC RISC-V Embedded Compiler.
<i>Device/Source/system_Device.c</i>	Generic <i>system_Device.c</i> file for system configuration (i.e. processor clock and memory bus system).
<i>Device/Include/Device.h</i>	Generic device header file. Needs to be extended with the device-specific peripheral registers. Optionally functions that access the peripherals can be part of that file.
<i>Device/Include/system_Device.h</i>	Generic system device configuration include file.

**Note:** The template files for silicon vendors are placed under *./Device/\_Template\_Vendor/Vendor/*.

Please goto that folder to find the file list in the above table.

### 2.3.5 Adapt Template Files to a Device

The following steps describe how to adopt the template files to a specific device or device family.

**Copy the complete all files in the template directory and replace:**

- directory name `Vendor` with the abbreviation for the device vendor e.g.: **GD**.
- directory name `Device` with the specific device name e.g.: **GD32VF103**.
- in the file names `Device` with the specific device name e.g.: **GD32VF103**.

Each template file contains comments that start with **TODO**: that describe a required modification.

The template files contain place holders:

Table 5: Placeholders of Template files

Placeholder	Replaced with
<Device>	the specific device name or device family name; i.e. GD32VF103.
<DeviceInterrupt>	a specific interrupt name of the device; i.e. TIM1 for Timer 1.
<DeviceAbbreviation>	short name or abbreviation of the device family; i.e. GD32VF.
Nuclei-N#	the specific Nuclei Class name; i.e. Nuclei N or Nuclei NX.

## 2.3.6 Device Templates Explanation

The device configuration of the template files is described in detail on the following pages:

### Startup File `startup_<device>.S`

The Startup File `startup_<device>.S` contains:

- The reset handler which is executed after CPU reset and typically calls the `SystemInit()` (page 406) function.
- The setup values for the stack pointer SP.
- Exception vectors of the Nuclei Processor with weak functions that implement default routines.
- Interrupt vectors that are device specific with weak functions that implement default routines.

The processor level start flow is implemented in the `startup_<device>.S`. Detail description as below picture:

#### Stage1: Interrupt and Exception initialization

- Disable Interrupt
- Initialize GP, stack
- Initialize NMI entry and set default NMI handler
- Initialize Exception entry and set default exception handler
- Initialize vector table entry and set default interrupt handler
- Initialize Interrupt mode as ECLIC mode. (ECLIC mode is proposed. Default mode is CLINT mode)

#### Stage2: Hardware initialization

- Enable FPU if necessary
- Call user defined `SystemInit()` (page 406) for system clock initialization.

#### Stage3: Section initialization

- Copy section, e.g. data section, text section if necessary.
- Clear Block Started by Symbol (BSS) section
- Call `__libc_fini_array` and `__libc_init_array` functions to do C library initialization

- Call `_premain_init` function to do initialization steps before main function
- Jump Main

The file exists for each supported toolchain and is the only toolchain specific NMSIS file.

To adapt the file to a new device only the interrupt vector table needs to be extended with the device-specific interrupt handlers.

The naming convention for the interrupt handler names are `eclic_<interrupt_name>_handler`.

This table needs to be consistent with `IRQn_Type` that defines all the IRQ numbers for each interrupt.

The following example shows the extension of the interrupt vector table for the GD32VF103 device family.

```

1      .section .vtable
2
3      .weak  eclic_msip_handler
4      .weak  eclic_mt看ip_handler
5      .weak  eclic_pmaf_handler
6      /* Adjusted for GD32VF103 interrupt handlers */
7      .weak  eclic_wwdgt_handler
8      .weak  eclic_lvd_handler
9      .weak  eclic_tamper_handler
10     :      :
11     :      :
12     .weak  eclic_can1_ewmc_handler
13     .weak  eclic_usbfs_handler
14
15     .globl vector_base
16     .type vector_base, @object
17 vector_base:
18     /* Run in FlashXIP download mode */
19     j _start                                     /* 0: Reserved, Jump to _
20     ↪start when reset for vector table not remapped cases.*/
21     .align LOG_REGBYTES                         /*      Need to align 4_
22     ↪byte for RV32, 8 Byte for RV64 */
23     DECLARE_INT_HANDLER    default_intexc_handler    /* 1: Reserved */
24     DECLARE_INT_HANDLER    default_intexc_handler    /* 2: Reserved */
25     DECLARE_INT_HANDLER    eclic_msip_handler        /* 3: Machine software_
26     ↪interrupt */
27     :          :
28     :          :
29     /* Adjusted for Vendor Defined External Interrupts */
30     DECLARE_INT_HANDLER    eclic_wwdgt_handler      /* 19: Window watchDog_
31     ↪timer interrupt */
32
33     DECLARE_INT_HANDLER    eclic_lvd_handler        /* 20: LVD through EXTI_
34     ↪line detect interrupt */
35     DECLARE_INT_HANDLER    eclic_tamper_handler     /* 21: tamper through_
36     ↪EXTI line detect */
37     :          :
38     :          :
39     DECLARE_INT_HANDLER    eclic_can1_ewmc_handler  /* 85: CAN1 EWMC_
40     ↪interrupt */
41     DECLARE_INT_HANDLER    eclic_usbfs_handler     /* 86: USBFS global_
42     ↪interrupt */

```

## startup\_Device.S Template File

Here provided a riscv-gcc template startup assemble code template file as below. The files for other compilers can slightly differ from this version.

```

1  /*
2  * Copyright (c) 2019 Nuclei Limited. All rights reserved.
3  *
4  * SPDX-License-Identifier: Apache-2.0
5  *
6  * Licensed under the Apache License, Version 2.0 (the License); you may
7  * not use this file except in compliance with the License.
8  * You may obtain a copy of the License at
9  *
10 * www.apache.org/licenses/LICENSE-2.0
11 *
12 * Unless required by applicable law or agreed to in writing, software
13 * distributed under the License is distributed on an AS IS BASIS, WITHOUT
14 * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
15 * See the License for the specific language governing permissions and
16 * limitations under the License.
17 */
18 /*****
19 * \file      startup_<Device>.S
20 * \brief     NMSIS Nuclei N/NX Class Core based Core Device Startup File for
21 *           Device <Device>
22 * \version   V1.10
23 * \date      30. July 2021
24 *
25 *****/
26
27 #include "riscv_encoding.h"
28
29 .macro DECLARE_INT_HANDLER INT_HDL_NAME
30 #if defined(__riscv_xlen) && (__riscv_xlen == 32)
31     .word \INT_HDL_NAME
32 #else
33     .dword \INT_HDL_NAME
34 #endif
35 .endm
36
37 /*
38 * Put the interrupt vectors in this section according to vector remapped or not:
39 * .vtable: vector table's LMA and VMA are the same, it is not remapped
40 * .vtable_ilm: vector table's LMA and VMA are different, it is remapped, and
41 *             VECTOR_TABLE_REMAPPED need to be defined
42 */
43 #if defined(VECTOR_TABLE_REMAPPED)
44     .section .vtable_ilm
45 #else
46     .section .vtable
47 #endif
48
49     .weak eclic_msip_handler
50     .weak eclic_mtip_handler
51     /* TODO: Adjust vendor interrupt handlers */
52     .weak eclic_irq19_handler

```

(continues on next page)

(continued from previous page)

```

53  .weak  eclic_irq20_handler
54  .weak  eclic_irq21_handler
55  .weak  eclic_irq22_handler
56  .weak  eclic_irq23_handler
57  .weak  eclic_irq24_handler
58  .weak  eclic_irq25_handler
59  .weak  eclic_irq26_handler
60  .weak  eclic_irq27_handler
61  .weak  eclic_irq28_handler
62  .weak  eclic_irq29_handler
63  .weak  eclic_irq30_handler
64  .weak  eclic_irq31_handler
65  .weak  eclic_irq32_handler
66  .weak  eclic_irq33_handler
67  .weak  eclic_irq34_handler
68  .weak  eclic_irq35_handler
69  .weak  eclic_irq36_handler
70  .weak  eclic_irq37_handler
71  .weak  eclic_irq38_handler
72  .weak  eclic_irq39_handler
73  .weak  eclic_irq40_handler
74  .weak  eclic_irq41_handler
75  .weak  eclic_irq42_handler
76  .weak  eclic_irq43_handler
77  .weak  eclic_irq44_handler
78  .weak  eclic_irq45_handler
79  .weak  eclic_irq46_handler
80  .weak  eclic_irq47_handler
81  .weak  eclic_irq48_handler
82  .weak  eclic_irq49_handler
83  .weak  eclic_irq50_handler
84
85  .globl vector_base
86  .type vector_base, @object
87  vector_base:
88      j _start                                     /* 0: Reserved, Jump to _
↳start when reset for vector table not remapped cases.*/
89      .align LOG_REGBYTES                          /*      Need to align 4_
↳byte for RV32, 8 Byte for RV64 */
90
91      DECLARE_INT_HANDLER    default_intexc_handler    /* 1: Reserved */
92      DECLARE_INT_HANDLER    default_intexc_handler    /* 2: Reserved */
93      DECLARE_INT_HANDLER    eclic_msip_handler         /* 3: Machine software_
↳interrupt */
94
95      DECLARE_INT_HANDLER    default_intexc_handler    /* 4: Reserved */
96      DECLARE_INT_HANDLER    default_intexc_handler    /* 5: Reserved */
97      DECLARE_INT_HANDLER    default_intexc_handler    /* 6: Reserved */
98      DECLARE_INT_HANDLER    eclic_mtip_handler        /* 7: Machine timer_
↳interrupt */
99
100     DECLARE_INT_HANDLER    default_intexc_handler    /* 8: Reserved */
101     DECLARE_INT_HANDLER    default_intexc_handler    /* 9: Reserved */
102     DECLARE_INT_HANDLER    default_intexc_handler    /* 10: Reserved */
103     DECLARE_INT_HANDLER    default_intexc_handler    /* 11: Reserved */
104
105     DECLARE_INT_HANDLER    default_intexc_handler    /* 12: Reserved */

```

(continues on next page)

(continued from previous page)

```

106 DECLARE_INT_HANDLER default_intexc_handler /* 13: Reserved */
107 DECLARE_INT_HANDLER default_intexc_handler /* 14: Reserved */
108 DECLARE_INT_HANDLER default_intexc_handler /* 15: Reserved */
109
110 DECLARE_INT_HANDLER default_intexc_handler /* 16: Reserved */
111 DECLARE_INT_HANDLER default_intexc_handler /* 17: Reserved */
112 DECLARE_INT_HANDLER default_intexc_handler /* 18: Reserved */
113 /* TODO: Adjust Vendor Defined External Interrupts */
114 DECLARE_INT_HANDLER eclic_irq19_handler /* 19: Interrupt 19 */
115
116 DECLARE_INT_HANDLER eclic_irq20_handler /* 20: Interrupt 20 */
117 DECLARE_INT_HANDLER eclic_irq21_handler /* 21: Interrupt 21 */
118 DECLARE_INT_HANDLER eclic_irq22_handler /* 22: Interrupt 22 */
119 DECLARE_INT_HANDLER eclic_irq23_handler /* 23: Interrupt 23 */
120
121 DECLARE_INT_HANDLER eclic_irq24_handler /* 24: Interrupt 24 */
122 DECLARE_INT_HANDLER eclic_irq25_handler /* 25: Interrupt 25 */
123 DECLARE_INT_HANDLER eclic_irq26_handler /* 26: Interrupt 26 */
124 DECLARE_INT_HANDLER eclic_irq27_handler /* 27: Interrupt 27 */
125
126 DECLARE_INT_HANDLER eclic_irq28_handler /* 28: Interrupt 28 */
127 DECLARE_INT_HANDLER eclic_irq29_handler /* 29: Interrupt 29 */
128 DECLARE_INT_HANDLER eclic_irq30_handler /* 30: Interrupt 30 */
129 DECLARE_INT_HANDLER eclic_irq31_handler /* 31: Interrupt 31 */
130
131 DECLARE_INT_HANDLER eclic_irq32_handler /* 32: Interrupt 32 */
132 DECLARE_INT_HANDLER eclic_irq33_handler /* 33: Interrupt 33 */
133 DECLARE_INT_HANDLER eclic_irq34_handler /* 34: Interrupt 34 */
134 DECLARE_INT_HANDLER eclic_irq35_handler /* 35: Interrupt 35 */
135
136 DECLARE_INT_HANDLER eclic_irq36_handler /* 36: Interrupt 36 */
137 DECLARE_INT_HANDLER eclic_irq37_handler /* 37: Interrupt 37 */
138 DECLARE_INT_HANDLER eclic_irq38_handler /* 38: Interrupt 38 */
139 DECLARE_INT_HANDLER eclic_irq39_handler /* 39: Interrupt 39 */
140
141 DECLARE_INT_HANDLER eclic_irq40_handler /* 40: Interrupt 40 */
142 DECLARE_INT_HANDLER eclic_irq41_handler /* 41: Interrupt 41 */
143 DECLARE_INT_HANDLER eclic_irq42_handler /* 42: Interrupt 42 */
144 DECLARE_INT_HANDLER eclic_irq43_handler /* 43: Interrupt 43 */
145
146 DECLARE_INT_HANDLER eclic_irq44_handler /* 44: Interrupt 44 */
147 DECLARE_INT_HANDLER eclic_irq45_handler /* 45: Interrupt 45 */
148 DECLARE_INT_HANDLER eclic_irq46_handler /* 46: Interrupt 46 */
149 DECLARE_INT_HANDLER eclic_irq47_handler /* 47: Interrupt 47 */
150
151 DECLARE_INT_HANDLER eclic_irq48_handler /* 48: Interrupt 48 */
152 DECLARE_INT_HANDLER eclic_irq49_handler /* 49: Interrupt 49 */
153 DECLARE_INT_HANDLER eclic_irq50_handler /* 50: Interrupt 50 */
154 /* Please adjust the above part of interrupt definition code
155  * according to your device interrupt number and its configuration */
156
157
158 /*** Startup Code Section ***/
159 .section .init
160
161 .globl _start
162 .type _start,@function

```

(continues on next page)



(continued from previous page)

```

163
164 /**
165  * Reset Handler called on controller reset
166  */
167 _start:
168     /* ===== Startup Stage 1 ===== */
169     /* Disable Global Interrupt */
170     csrc CSR_MSTATUS, MSTATUS_MIE
171
172     /* Initialize GP and Stack Pointer SP */
173     .option push
174     .option norelax
175     la gp, __global_pointer$
176
177     .option pop
178     la sp, _sp
179
180     /*
181      * Set the the NMI base mnvec to share
182      * with mtvec by setting CSR_MMISC_CTL
183      * bit 9 NMI_CAUSE_FFF to 1
184      */
185     li t0, MMISC_CTL_NMI_CAUSE_FFF
186     csrs CSR_MMISC_CTL, t0
187
188     /*
189      * Intialize ECLIC vector interrupt
190      * base address mtvt to vector_base
191      */
192     la t0, vector_base
193     csrwr CSR_MTVT, t0
194
195     /*
196      * Set ECLIC non-vector entry to be controlled
197      * by mtvt2 CSR register.
198      * Intialize ECLIC non-vector interrupt
199      * base address mtvt2 to irq_entry.
200      */
201     la t0, irq_entry
202     csrwr CSR_MTVT2, t0
203     csrs CSR_MTVT2, 0x1
204
205     /*
206      * Set Exception Entry MTVEC to exc_entry
207      * Due to settings above, Exception and NMI
208      * will share common entry.
209      */
210     la t0, exc_entry
211     csrwr CSR_MTVEC, t0
212
213     /* Set the interrupt processing mode to ECLIC mode */
214     li t0, 0x3f
215     csrc CSR_MTVEC, t0
216     csrs CSR_MTVEC, 0x3
217
218     /* ===== Startup Stage 2 ===== */
219

```

(continues on next page)

(continued from previous page)

```

220 #if defined(__riscv_flen) && __riscv_flen > 0
221     /* Enable FPU */
222     li t0, MSTATUS_FS
223     csrs mstatus, t0
224     csrwr fcsr, x0
225 #endif
226
227     /* Enable mcycle and minstret counter */
228     csrwr CSR_MCOUNTINHIBIT, 0x5
229
230     /*
231      * Call vendor defined SystemInit to
232      * initialize the micro-controller system.
233      * TODO: You need to comment this code when run in Flash download mode.
234      * then you need to put this line of code
235      * after data/bss section initialization and before main
236      */
237     call SystemInit
238
239     /* ===== Startup Stage 3 ===== */
240     /*
241      * Load code section from FLASH to ILM
242      * when code LMA is different with VMA
243      */
244     la a0, _ilm_lma
245     la a1, _ilm
246     /* If the ILM phy-address same as the logic-address, then quit */
247     beq a0, a1, 2f
248     la a2, _eilm
249     bgeu a1, a2, 2f
250
251 1:
252     /* Load code section if necessary */
253     lw t0, (a0)
254     sw t0, (a1)
255     addi a0, a0, 4
256     addi a1, a1, 4
257     bltu a1, a2, 1b
258 2:
259     /* Load data section */
260     la a0, _data_lma
261     la a1, _data
262     la a2, _edata
263     bgeu a1, a2, 2f
264 1:
265     lw t0, (a0)
266     sw t0, (a1)
267     addi a0, a0, 4
268     addi a1, a1, 4
269     bltu a1, a2, 1b
270 2:
271     /* Clear bss section */
272     la a0, __bss_start
273     la a1, _end
274     bgeu a0, a1, 2f
275 1:
276     sw zero, (a0)

```

(continues on next page)

(continued from previous page)

```

277     addi a0, a0, 4
278     bltu a0, a1, 1b
279 2:
280     /* TODO: Uncomment this code, if you run in Flash download mode */
281     // call SystemInit
282
283     /* Call global constructors */
284     la a0, __libc_fini_array
285     call atexit
286     /* Call C/C++ constructor start up code */
287     call __libc_init_array
288
289     /* do pre-init steps before main */
290     call _premain_init
291     /* ===== Call Main Function ===== */
292     /* argc = argv = 0 */
293     li a0, 0
294     li a1, 0
295     call main
296     /* do post-main steps after main */
297     call _postmain_fini
298
299 1:
300     j 1b

```

## Interrupt and Exception Handling File: intexc\_<device>.S

The intexc File intexc\_<device>.S contains:

- Macro to save caller register.
- Macro to restore caller register.
- Default Exception/NMI routine implementation.
- Default Non-Vector Interrupt routine implementation.

Nuclei processors provide **NMI(Non-Maskable Interrupt)**, **Exception**, **Vector Interrupt** and **Non-Vector Interrupt** features.

### NMI(Non-Maskable Interrupt)

Click [NMI<sup>9</sup>](https://doc.nucleisys.com/nuclei_spec/isa/nmi.html) to learn about Nuclei Processor Core NMI in Nuclei ISA Spec.

NMI is used for urgent external HW error. It can't be masked and disabled.

When NMI happened, bit 9 of CSR MMSIC\_CTL will be checked. If this bit value is 1, then NMI entry address will be the same as exception(CSR\_MTVEC), and exception code for NMI will be 0xFFFF, otherwise NMI entry will be same as reset\_vector.

In NMSIS-Core, the bit 9 of CSR MMISC\_CTL is set to 1 during core startup, so NMI will be treated as Exception and handled.

<sup>9</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/nmi.html](https://doc.nucleisys.com/nuclei_spec/isa/nmi.html)

## Exception

Click [Exception<sup>10</sup>](#) to learn about Nuclei Processor Core Exception in Nuclei ISA Spec.

For CPU exception, the entry for exception will be `exc_entry`, in this entry code, it will call default exception handler `core_exception_handler()` (page 408).

In the common exception routine(`exc_entry`) to get more information like exception code. Exception handle flow show as below picture:

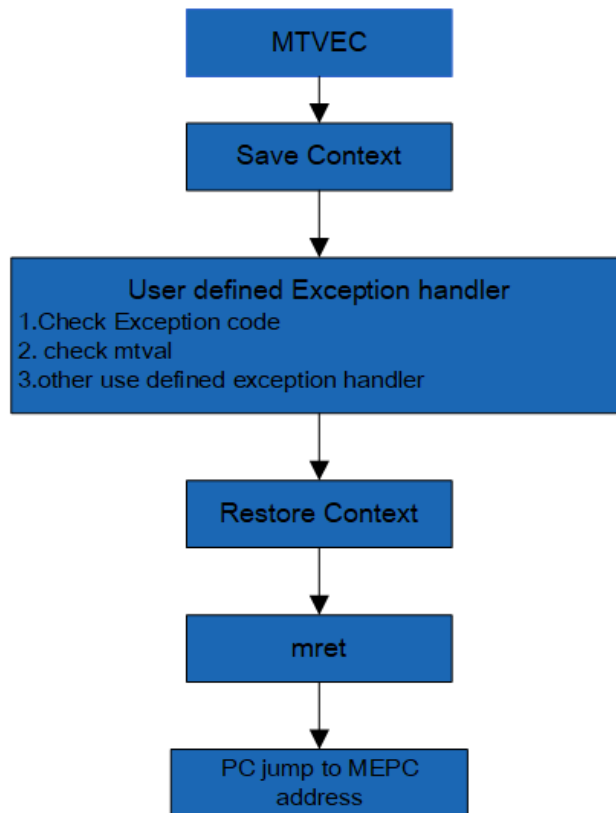


Fig. 3: Exception Handling Flow

NMI and exception could support nesting. Two levels of NMI/Exception state save stacks are supported.

We support three nesting mode as below:

- NMI nesting exception
- Exception nesting exception
- Exception nesting NMI

For software, we have provided the common entry for NMI and exception. Silicon vendor only need adapt the interface defined in *Interrupt Exception NMI Handling* (page 407).

Context save and restore have been handled by `exc_entry` interface.

When exception exception return it will run the instruction which trigger the exception again. It will cause software dead loop. So in the exception handler for each exception code, we propose to set CSR MEPC to be MEPC+4, then it will start from next instruction of MEPC.

<sup>10</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/exception.html](https://doc.nucleisys.com/nuclei_spec/isa/exception.html)

## Interrupt

Click [Interrupt<sup>11</sup>](#) to learn about Nuclei Processor Core Interrupt in Nuclei Spec.

Interrupt could be configured as **CLINT** mode or **ECILC** mode.

In NMSIS-Core, Interrupt has been configured as **ECLIC** mode during startup in *startup\_<Devices>.S*, which is also recommended setting using Nuclei Processors.

ECLIC managed interrupt could be configured as **vector** and **non-vector** mode.

Detail interrupt handling process as below picture:

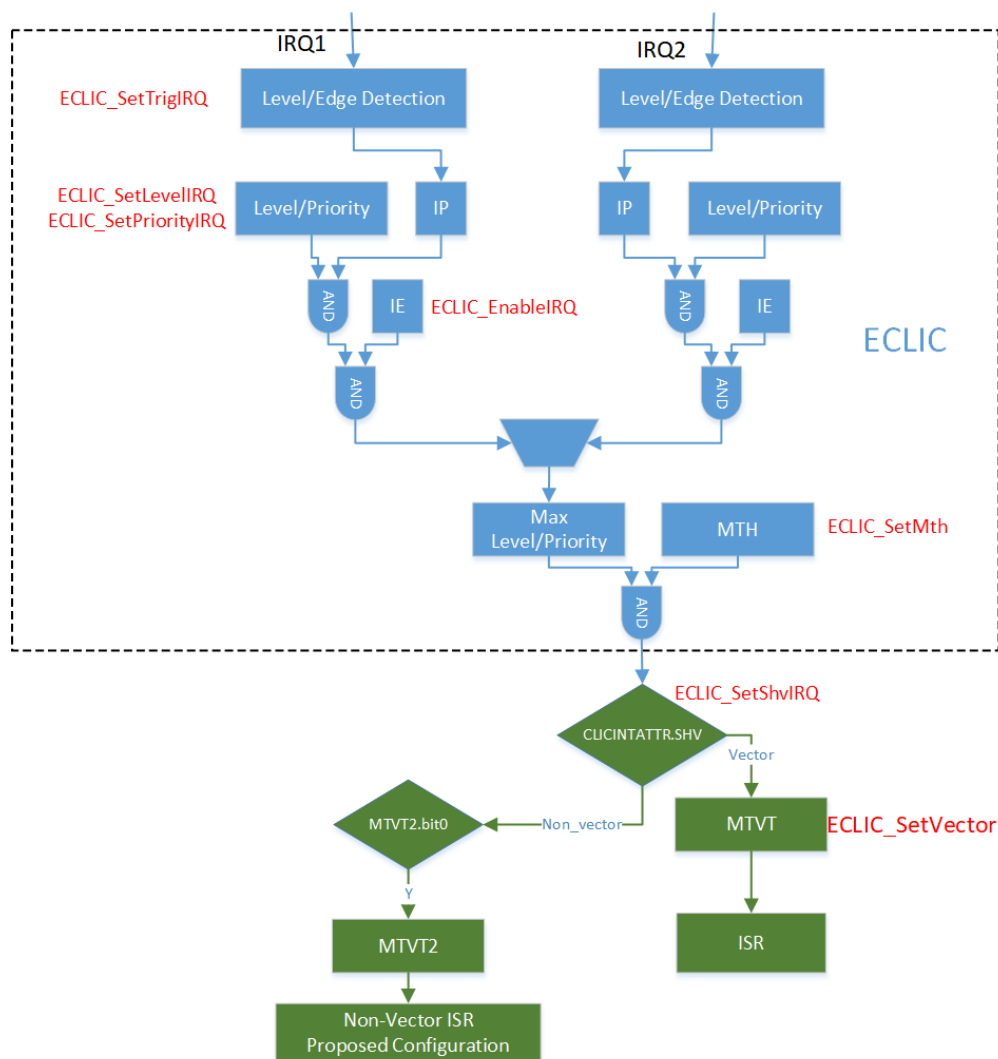


Fig. 4: Interrupt Handling Flow

To get highest priority interrupt we need compare the interrupt level first. If level is the same then compare the priority. High level interrupt could interrupt low level ISR and trigger interrupt nesting. If different priority with same level interrupt pending higher priority will be served first. Interrupt could be configured as vector mode and non-vector mode by vendor. For non-vector mode interrupt handler entry get from MTVT2 and exception/NMI handler entry get

<sup>11</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/interrupt.html](https://doc.nucleisys.com/nuclei_spec/isa/interrupt.html)

from MTVEC. If Vendor need set non vector mode interrupt handler entry from MTVVEC you need set MTVT2.BIT0 as 0.

### Non-Vector Interrupt SW handler

For **non-vector** mode interrupt it will make the necessary CSR registers and context save and restore. Non-vector mode software handle flow show as below picture:

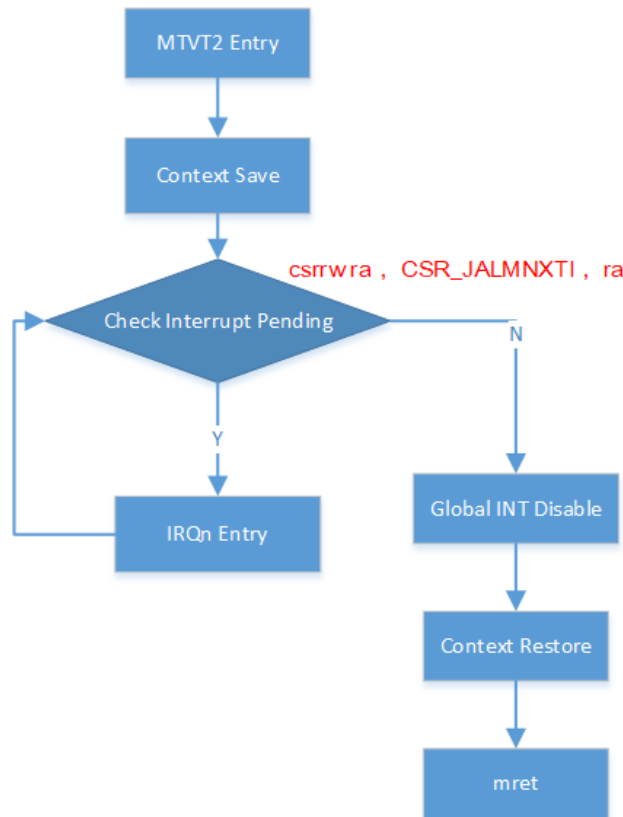


Fig. 5: Non-vector mode interrupt software handle flow

#### Detail description for non-vector mode interrupt handler as below steps:

1. Get non-vector mode handler entry from MTVT2 if MTVT2.BIT0 is 1(proposed configuration).
2. Context save to stack for cpu registers.
3. Save CSR registers MEPC/MCAUSE/MSUBM to stack.
4. Run instruction `csrrw ra, CSR_JALMNXTI, ra`. It will enable interrupt, check interrupt pending. If interrupt is pending then get highest priority interrupt and jump to interrupt handler entry in the vector table, otherwise it will go to step 6.
5. Execute the interrupt handler routine, when return from isr routine it will jump to step 4.
6. Global interrupt disable.
7. Restore CSR registers MEPC/MCAUSE/MSUBM.
8. Context restore from stack for cpu registers.
9. Execute `mret` to return from handler.

For **non-vector** mode interrupt it could support **interrupt nesting**.

**Interrupt nesting** handle flow show as below picture:

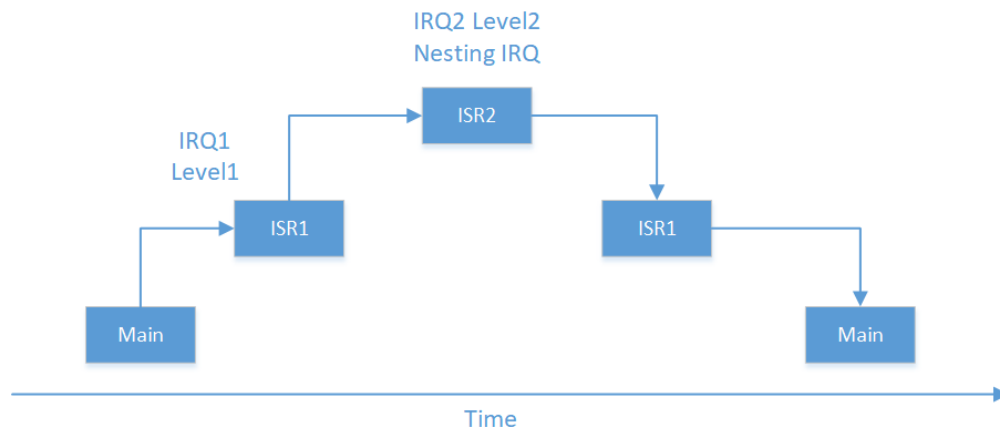


Fig. 6: Nesting interrupt handling flow

### Vector Interrupt SW handler

If vector interrupt handler need support nesting or making function call Vector mode software handling flow show as below picture:

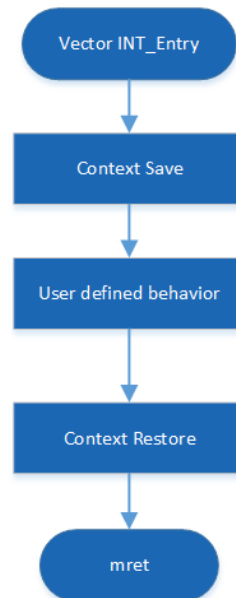


Fig. 7: Vector mode nesting interrupt handling flow

**Detail description for nested vector mode interrupt handler as below steps:**

1. Get vector mode handler from address of vector table entry MTVT added offset.
2. Context save to stack for cpu registers, done in each vector interrupt handler via `__INTERRUPT` (page 60)

3. Save CSR registers MEPC/MCAUSE/MSUBM to stack, done in each vector interrupt handler by read and save these CSRs into variables.
4. Execute the interrupt handling.
5. Restore CSR registers MEPC/MCAUSE/MSUBM from stack.
6. CSR registers restore from saved variables used in step 3.
7. Execute `mret` to return from handler

Here is sample code for above nested vector interrupt handling process:

```

1 // Vector interrupt handler for on-board button
2 __INTERRUPT void SOC_BUTTON_1_HANDLER(void)
3 {
4     // save mepc,mcause,msubm enable interrupts
5     SAVE_IRQ_CSR_CONTEXT();
6
7     printf("%s", "----Begin button1 handler----Vector mode\r\n");
8
9     // Green LED toggle
10    gpio_toggle(GPIO, SOC_LED_GREEN_GPIO_MASK);
11
12    // Clear the GPIO Pending interrupt by writing 1.
13    gpio_clear_interrupt(GPIO, SOC_BUTTON_1_GPIO_OFS, GPIO_INT_RISE);
14
15    wait_seconds(1); // Wait for a while
16
17    printf("%s", "----End button1 handler\r\n");
18
19    // disable interrupts, restore mepc,mcause,msubm
20    RESTORE_IRQ_CSR_CONTEXT();
21 }

```

#### Detail description for non-nested vector mode interrupt handler as below

To improve the software response latency for vector mode vendor could remove context save/restore and **MEPC/MCAUSE/MSUBM** save/restore.

If so vector mode interrupt will not support nesting and interrupt handler can only be a leaf function which doesn't make any function calls.

#### Then the vector mode interrupt software flow will be described as below:

1. Get vector mode handler from address of vector table entry MTVT added offset.
2. Execute the interrupt handler(leaf function).
3. Execute `mret` to return from handler

Here is sample code for above non-nested vector interrupt handler which is a leaf function handling process:

```

1 static uint32_t btn_pressed = 0;
2 // Vector interrupt handler for on-board button
3 // This function is an leaf function, no function call is allowed
4 __INTERRUPT void SOC_BUTTON_1_HANDLER(void)
5 {
6     btn_pressed++;
7 }

```



## intexc\_Device.S Template File

The file exists for each supported toolchain and is the only toolchain specific NMSIS file.

Normally this file needn't adapt for different device. If CPU CSR registers have done some changes you may need some adaption.

Here we provided `intexc_Device.S` template file as below:

```

1  /*
2   * Copyright (c) 2019 Nuclei Limited. All rights reserved.
3   *
4   * SPDX-License-Identifier: Apache-2.0
5   *
6   * Licensed under the Apache License, Version 2.0 (the License); you may
7   * not use this file except in compliance with the License.
8   * You may obtain a copy of the License at
9   *
10  * www.apache.org/licenses/LICENSE-2.0
11  *
12  * Unless required by applicable law or agreed to in writing, software
13  * distributed under the License is distributed on an AS IS BASIS, WITHOUT
14  * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
15  * See the License for the specific language governing permissions and
16  * limitations under the License.
17  */
18  /*****
19   * \file      intexc_<Device>.S
20   * \brief     NMSIS Interrupt and Exception Handling Template File
21   *           for Nuclei N/NX Class Device
22   * \version   V1.10
23   * \date      30. July 2021
24   *
25   *****/
26
27  #include "riscv_encoding.h"
28
29  /**
30   * \brief     Global interrupt disabled
31   * \details
32   * This function disable global interrupt.
33   * \remarks
34   * - All the interrupt requests will be ignored by CPU.
35   */
36  .macro DISABLE_MIE
37      csrc CSR_MSTATUS, MSTATUS_MIE
38  .endm
39
40  /**
41   * \brief     Macro for context save
42   * \details
43   * This macro save ABI defined caller saved registers in the stack.
44   * \remarks
45   * - This Macro could use to save context when you enter to interrupt
46   * or exception
47   */
48  /* Save caller registers */
49  .macro SAVE_CONTEXT

```

(continues on next page)

(continued from previous page)

```

50  /* Allocate stack space for context saving */
51  #ifndef __riscv_32e
52      addi sp, sp, -20*REGBYTES
53  #else
54      addi sp, sp, -14*REGBYTES
55  #endif /* __riscv_32e */
56
57      STORE x1, 0*REGBYTES(sp)
58      STORE x4, 1*REGBYTES(sp)
59      STORE x5, 2*REGBYTES(sp)
60      STORE x6, 3*REGBYTES(sp)
61      STORE x7, 4*REGBYTES(sp)
62      STORE x10, 5*REGBYTES(sp)
63      STORE x11, 6*REGBYTES(sp)
64      STORE x12, 7*REGBYTES(sp)
65      STORE x13, 8*REGBYTES(sp)
66      STORE x14, 9*REGBYTES(sp)
67      STORE x15, 10*REGBYTES(sp)
68  #ifndef __riscv_32e
69      STORE x16, 14*REGBYTES(sp)
70      STORE x17, 15*REGBYTES(sp)
71      STORE x28, 16*REGBYTES(sp)
72      STORE x29, 17*REGBYTES(sp)
73      STORE x30, 18*REGBYTES(sp)
74      STORE x31, 19*REGBYTES(sp)
75  #endif /* __riscv_32e */
76  .endm
77
78  /**
79   * \brief Macro for restore caller registers
80   * \details
81   * This macro restore ABI defined caller saved registers from stack.
82   * \remarks
83   * - You could use this macro to restore context before you want return
84   * from interrupt or exeception
85   */
86  /* Restore caller registers */
87  .macro RESTORE_CONTEXT
88      LOAD x1, 0*REGBYTES(sp)
89      LOAD x4, 1*REGBYTES(sp)
90      LOAD x5, 2*REGBYTES(sp)
91      LOAD x6, 3*REGBYTES(sp)
92      LOAD x7, 4*REGBYTES(sp)
93      LOAD x10, 5*REGBYTES(sp)
94      LOAD x11, 6*REGBYTES(sp)
95      LOAD x12, 7*REGBYTES(sp)
96      LOAD x13, 8*REGBYTES(sp)
97      LOAD x14, 9*REGBYTES(sp)
98      LOAD x15, 10*REGBYTES(sp)
99  #ifndef __riscv_32e
100      LOAD x16, 14*REGBYTES(sp)
101      LOAD x17, 15*REGBYTES(sp)
102      LOAD x28, 16*REGBYTES(sp)
103      LOAD x29, 17*REGBYTES(sp)
104      LOAD x30, 18*REGBYTES(sp)
105      LOAD x31, 19*REGBYTES(sp)
106

```

(continues on next page)

(continued from previous page)

```

107     /* De-allocate the stack space */
108     addi sp, sp, 20*REGBYTES
109 #else
110     /* De-allocate the stack space */
111     addi sp, sp, 14*REGBYTES
112 #endif /* __riscv_32e */
113
114 .endm
115
116 /**
117  * \brief Macro for save necessary CSRs to stack
118  * \details
119  * This macro store MCAUSE, MEPC, MSUBM to stack.
120  */
121 .macro SAVE_CSR_CONTEXT
122     /* Store CSR mcause to stack using pushmcause */
123     csrrwi x0, CSR_PUSHMCAUSE, 11
124     /* Store CSR mepc to stack using pushmepc */
125     csrrwi x0, CSR_PUSHMEPC, 12
126     /* Store CSR msub to stack using pushmsub */
127     csrrwi x0, CSR_PUSHMSUBM, 13
128 .endm
129
130 /**
131  * \brief Macro for restore necessary CSRs from stack
132  * \details
133  * This macro restore MSUBM, MEPC, MCAUSE from stack.
134  */
135 .macro RESTORE_CSR_CONTEXT
136     LOAD x5, 13*REGBYTES(sp)
137     csrwr CSR_MSUBM, x5
138     LOAD x5, 12*REGBYTES(sp)
139     csrwr CSR_MEPC, x5
140     LOAD x5, 11*REGBYTES(sp)
141     csrwr CSR_MCAUSE, x5
142 .endm
143
144 /**
145  * \brief Exception/NMI Entry
146  * \details
147  * This function provide common entry functions for exception/nmi.
148  * \remarks
149  * This function provide a default exception/nmi entry.
150  * ABI defined caller save register and some CSR registers
151  * to be saved before enter interrupt handler and be restored before return.
152  */
153 .section .text.trap
154 /* In CLIC mode, the exeception entry must be 64bytes aligned */
155 .align 6
156 .global exc_entry
157 .weak exc_entry
158 exc_entry:
159     /* Save the caller saving registers (context) */
160     SAVE_CONTEXT
161     /* Save the necessary CSR registers */
162     SAVE_CSR_CONTEXT
163

```

(continues on next page)

(continued from previous page)

```

164  /*
165  * Set the exception handler function arguments
166  * argument 1: mcause value
167  * argument 2: current stack point (SP) value
168  */
169  csrr a0, mcause
170  mv a1, sp
171  /*
172  * TODO: Call the exception handler function
173  * By default, the function template is provided in
174  * system_Device.c, you can adjust it as you want
175  */
176  call core_exception_handler
177
178  /* Restore the necessary CSR registers */
179  RESTORE_CSR_CONTEXT
180  /* Restore the caller saving registers (context) */
181  RESTORE_CONTEXT
182
183  /* Return to regular code */
184  mret
185
186  /**
187  * \brief Non-Vector Interrupt Entry
188  * \details
189  * This function provide common entry functions for handling
190  * non-vector interrupts
191  * \remarks
192  * This function provide a default non-vector interrupt entry.
193  * ABI defined caller save register and some CSR registers need
194  * to be saved before enter interrupt handler and be restored before return.
195  */
196  .section .text.irq
197  /* In CLIC mode, the interrupt entry must be 4bytes aligned */
198  .align 2
199  .global irq_entry
200  .weak irq_entry
201  /* This label will be set to MTVT2 register */
202  irq_entry:
203  /* Save the caller saving registers (context) */
204  SAVE_CONTEXT
205  /* Save the necessary CSR registers */
206  SAVE_CSR_CONTEXT
207
208  /* This special CSR read/write operation, which is actually
209  * claim the CLIC to find its pending highest ID, if the ID
210  * is not 0, then automatically enable the mstatus.MIE, and
211  * jump to its vector-entry-label, and update the link register
212  */
213  csrrw ra, CSR_JALMNXTI, ra
214
215  /* Critical section with interrupts disabled */
216  DISABLE_MIE
217
218  /* Restore the necessary CSR registers */
219  RESTORE_CSR_CONTEXT
220  /* Restore the caller saving registers (context) */

```

(continues on next page)

(continued from previous page)

```

221     RESTORE_CONTEXT
222
223     /* Return to regular code */
224     mret
225
226     /* Default Handler for Exceptions / Interrupts */
227     .global default_intexc_handler
228     .weak default_intexc_handler
229     Undef_Handler:
230     default_intexc_handler:
231     1:
232         j 1b

```

### Device Linker Script: gcc\_<device>.ld

The Linker Script File gcc\_<device>.ld contains:

- Memory base address and size.
- Code, data section, vector table etc. location.
- Stack & heap location and size.

The file exists for each supported toolchain and is the only toolchain specific NMSIS file.

To adapt the file to a new device only when you need change the memory base address, size, data and code location etc.

### gcc\_Device.ld Template File

Here we provided gcc\_Device.ld template file as below:

```

1  /*
2  * Copyright (c) 2019 Nuclei Limited. All rights reserved.
3  *
4  * SPDX-License-Identifier: Apache-2.0
5  *
6  * Licensed under the Apache License, Version 2.0 (the License); you may
7  * not use this file except in compliance with the License.
8  * You may obtain a copy of the License at
9  *
10 * www.apache.org/licenses/LICENSE-2.0
11 *
12 * Unless required by applicable law or agreed to in writing, software
13 * distributed under the License is distributed on an AS IS BASIS, WITHOUT
14 * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
15 * See the License for the specific language governing permissions and
16 * limitations under the License.
17 */
18 /*****
19 * @file      gcc_<Device>.ld
20 * @brief     GNU Linker Script for Nuclei N/NX based device
21 * @version   V1.1.0
22 * @date      30. July 2021
23 * *****/

```

(continues on next page)

(continued from previous page)

```

24
25 /***** Use Configuration Wizard in Context Menu *****/
26
27 OUTPUT_ARCH( "riscv" )
28 /***** Flash Configuration *****/
29 * <h> Flash Configuration
30 * <o0> Flash Base Address <0x0-0xFFFFFFFF:8>
31 * <o1> Flash Size (in Bytes) <0x0-0xFFFFFFFF:8>
32 * </h>
33 */
34 __ROM_BASE = 0x20000000;
35 __ROM_SIZE = 0x00400000;
36
37 /*----- ILM RAM Configuration -----*/
38 * <h> ILM RAM Configuration
39 * <o0> ILM RAM Base Address <0x0-0xFFFFFFFF:8>
40 * <o1> ILM RAM Size (in Bytes) <0x0-0xFFFFFFFF:8>
41 * </h>
42 */
43 __ILM_RAM_BASE = 0x80000000;
44 __ILM_RAM_SIZE = 0x00010000;
45
46 /*----- Embedded RAM Configuration -----*/
47 * <h> RAM Configuration
48 * <o0> RAM Base Address <0x0-0xFFFFFFFF:8>
49 * <o1> RAM Size (in Bytes) <0x0-0xFFFFFFFF:8>
50 * </h>
51 */
52 __RAM_BASE = 0x90000000;
53 __RAM_SIZE = 0x00010000;
54
55 /***** Stack / Heap Configuration *****/
56 * <h> Stack / Heap Configuration
57 * <o0> Stack Size (in Bytes) <0x0-0xFFFFFFFF:8>
58 * <o1> Heap Size (in Bytes) <0x0-0xFFFFFFFF:8>
59 * </h>
60 */
61 __STACK_SIZE = 0x00000800;
62 __HEAP_SIZE = 0x00000800;
63
64 /***** end of configuration section *****/
65
66 /* Define base address and length of flash and ram */
67 MEMORY
68 {
69     flash (rxai!w) : ORIGIN = __ROM_BASE, LENGTH = __ROM_SIZE
70     ram (wxa!ri) : ORIGIN = __RAM_BASE, LENGTH = __RAM_SIZE
71 }
72 /* Linker script to place sections and symbol values. Should be used together
73 * with other linker script that defines memory regions FLASH,ILM and RAM.
74 * It references following symbols, which must be defined in code:
75 * _start : Entry of reset handler
76 *
77 * It defines following symbols, which code can use without definition:
78 * _ilm_lma
79 * _ilm
80 * __etext

```

(continues on next page)

(continued from previous page)

```

81  *  _etext
82  *  etext
83  *  _eilm
84  *  __preinit_array_start
85  *  __preinit_array_end
86  *  __init_array_start
87  *  __init_array_end
88  *  __fini_array_start
89  *  __fini_array_end
90  *  _data_lma
91  *  _edata
92  *  edata
93  *  __data_end__
94  *  __bss_start
95  *  __fbss
96  *  _end
97  *  end
98  *  __heap_end
99  *  __StackLimit
100 *  __StackTop
101 *  __STACK_SIZE
102 */
103 /* Define entry label of program */
104 ENTRY(_start)
105 SECTIONS
106 {
107     __STACK_SIZE = DEFINED(__STACK_SIZE) ? __STACK_SIZE : 2K;
108
109     .init          :
110     {
111         /* vector table locate at flash */
112         *(.vtable)
113         KEEP (*(SORT_NONE(.init)))
114     } >flash AT>flash
115
116     .ilalign       :
117     {
118         . = ALIGN(4);
119         /* Create a section label as _ilm_lma which located at flash */
120         PROVIDE( _ilm_lma = . );
121     } >flash AT>flash
122
123     .ialign        :
124     {
125         /* Create a section label as _ilm which located at flash */
126         PROVIDE( _ilm = . );
127     } >flash AT>flash
128
129     /* Code section located at flash */
130     .text          :
131     {
132         *(.text.unlikely .text.unlikely.*)
133         *(.text.startup .text.startup.*)
134         *(.text .text.*)
135         *(.gnu.linkonce.t.*)
136     } >flash AT>flash
137

```

(continues on next page)

(continued from previous page)

```

138 .rodata : ALIGN(4)
139 {
140     . = ALIGN(4);
141     *(.rodata)
142     *(.rodata .rodata.*)
143     *(.gnu.linkonce.r.*)
144     . = ALIGN(8);
145     *(.srodata.cst16)
146     *(.srodata.cst8)
147     *(.srodata.cst4)
148     *(.srodata.cst2)
149     *(.srodata .srodata.*)
150 } >flash AT>flash
151
152 .fini      :
153 {
154     KEEP (*(SORT_NONE(.fini)))
155 } >flash AT>flash
156
157 . = ALIGN(4);
158
159 PROVIDE (__etext = .);
160 PROVIDE (_etext = .);
161 PROVIDE (etext = .);
162 PROVIDE( _eilm = . );
163
164
165 .preinit_array :
166 {
167     PROVIDE_HIDDEN (__preinit_array_start = .);
168     KEEP (*(preinit_array))
169     PROVIDE_HIDDEN (__preinit_array_end = .);
170 } >flash AT>flash
171
172 .init_array   :
173 {
174     PROVIDE_HIDDEN (__init_array_start = .);
175     KEEP (*(SORT_BY_INIT_PRIORITY(.init_array.*) SORT_BY_INIT_PRIORITY(.ctors.*)))
176     KEEP (*(init_array EXCLUDE_FILE (*crtbegin.o *crtbegin?.o *crtend.o *crtend?.o )
177 ↪.ctors))
178     PROVIDE_HIDDEN (__init_array_end = .);
179 } >flash AT>flash
180
181 .fini_array   :
182 {
183     PROVIDE_HIDDEN (__fini_array_start = .);
184     KEEP (*(SORT_BY_INIT_PRIORITY(.fini_array.*) SORT_BY_INIT_PRIORITY(.dtors.*)))
185     KEEP (*(fini_array EXCLUDE_FILE (*crtbegin.o *crtbegin?.o *crtend.o *crtend?.o )
186 ↪.dtors))
187     PROVIDE_HIDDEN (__fini_array_end = .);
188 } >flash AT>flash
189
190 .ctors        :
191 {
192     /* gcc uses crtbegin.o to find the start of
193      * the constructors, so we make sure it is
194      * first.  Because this is a wildcard, it

```

(continues on next page)



(continued from previous page)

```

193  * doesn't matter if the user does not
194  * actually link against crtbegin.o; the
195  * linker won't look for a file to match a
196  * wildcard. The wildcard also means that it
197  * doesn't matter which directory crtbegin.o
198  * is in.
199  */
200  KEEP (*crtbegin.o(.ctors))
201  KEEP (*crtbegin?.o(.ctors))
202  /* We don't want to include the .ctor section from
203  * the crtend.o file until after the sorted ctors.
204  * The .ctor section from the crtend file contains the
205  * end of ctors marker and it must be last
206  */
207  KEEP (*(EXCLUDE_FILE (*crtend.o *crtend?.o ) .ctors))
208  KEEP (*(SORT(.ctors.*)))
209  KEEP (*(.ctors))
210 } >flash AT>flash
211
212 .dtors          :
213 {
214     KEEP (*crtbegin.o(.dtors))
215     KEEP (*crtbegin?.o(.dtors))
216     KEEP (*(EXCLUDE_FILE (*crtend.o *crtend?.o ) .dtors))
217     KEEP (*(SORT(.dtors.*)))
218     KEEP (*(.dtors))
219 } >flash AT>flash
220
221 .lalign         :
222 {
223     . = ALIGN(4);
224     PROVIDE( _data_lma = . );
225 } >flash AT>flash
226
227 .dalign         :
228 {
229     . = ALIGN(4);
230     PROVIDE( _data = . );
231 } >ram AT>flash
232
233 /* Define data section virtual address is ram and physical address is flash */
234 .data           :
235 {
236     *(.data .data.*)
237     *(.gnu.linkonce.d.*)
238     . = ALIGN(8);
239     PROVIDE( __global_pointer$ = . + 0x800 );
240     *(.sdata .sdata.* .sdata*)
241     *(.gnu.linkonce.s.*)
242 } >ram AT>flash
243
244 . = ALIGN(4);
245 PROVIDE( _edata = . );
246 PROVIDE( edata = . );
247
248 PROVIDE( _fbss = . );
249 PROVIDE( __bss_start = . );

```

(continues on next page)

(continued from previous page)

```

250  .bss          :
251  {
252      *(.sbss*)
253      *(.gnu.linkonce.sb.*)
254      *(.bss .bss.*)
255      *(.gnu.linkonce.b.*)
256      *(COMMON)
257      . = ALIGN(4);
258  } >ram AT>ram
259
260  . = ALIGN(8);
261  PROVIDE( _end = . );
262  PROVIDE( end = . );
263  /* Define stack and head location at ram */
264  .stack ORIGIN(ram) + LENGTH(ram) - __STACK_SIZE :
265  {
266      PROVIDE( _heap_end = . );
267      __StackLimit = .;
268      . = __STACK_SIZE;
269      __StackTop = .;
270      PROVIDE( _sp = . );
271  } >ram AT>ram
272  }

```

## System Configuration Files `system_<device>.c` and `system_<device>.h`

The **System Configuration Files** `system_<device>.c` and `system_<device>.h` provides as a minimum the functions described under *System Device Configuration* (page 405).

These functions are device specific and need adaptations. In addition, the file might have configuration settings for the device such as XTAL frequency or PLL prescaler settings, necessary system initialization, vendor customized interrupt, exception and nmi handling code, refer to *System Device Configuration* (page 405) for more details.

For devices with external memory BUS the `system_<device>.c` also configures the BUS system.

The silicon vendor might expose other functions (i.e. for power configuration) in the `system_<device>.c` file. In case of additional features the function prototypes need to be added to the `system_<device>.h` header file.

## system\_Device.c Template File

Here we provided `system_Device.c` template file as below:

```

1  /*
2  * Copyright (c) 2009-2018 Arm Limited. All rights reserved.
3  * Copyright (c) 2019 Nuclei Limited. All rights reserved.
4  *
5  * SPDX-License-Identifier: Apache-2.0
6  *
7  * Licensed under the Apache License, Version 2.0 (the License); you may
8  * not use this file except in compliance with the License.
9  * You may obtain a copy of the License at
10 *
11 * www.apache.org/licenses/LICENSE-2.0
12 *

```

(continues on next page)

(continued from previous page)

```

13  * Unless required by applicable law or agreed to in writing, software
14  * distributed under the License is distributed on an AS IS BASIS, WITHOUT
15  * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
16  * See the License for the specific language governing permissions and
17  * limitations under the License.
18  */
19  /*****
20  * @file      system_<Device>.c
21  * @brief     NMSIS Nuclei N/NX Device Peripheral Access Layer Source File for
22  *           Device <Device>
23  * @version   V1.10
24  * @date      30. July 2021
25  *****/
26
27  #include <stdint.h>
28  #include "<Device>.h"
29
30  /-----
31  Define clocks
32  *-----*/
33  /* TODO: add here your necessary defines for device initialization
34     following is an example for different system frequencies */
35  #define XTAL          (12000000U)          /* Oscillator frequency */
36
37  #define SYSTEM_CLOCK  (5 * XTAL)
38
39  /**
40   * \defgroup NMSIS_Core_SystemConfig      System Device Configuration
41   * \brief Functions for system init, clock setup and interrupt/exception/nmi_
42   *        functions available in system_<device>.c.
43   * \details
44   * Nuclei provides a template file **system_Device.c** that must be adapted by
45   * the silicon vendor to match their actual device. As a <b>minimum requirement</b>,
46   * this file must provide:
47   * - A device-specific system configuration function, \ref SystemInit.
48   * - A global variable that contains the system frequency, \ref SystemCoreClock.
49   * - A global eclic configuration initialization, \ref ECLIC_Init.
50   * - Global \ref _premain_init and \ref _postmain_fini functions called right_
51   *   before calling main function.
52   * - Vendor customized interrupt, exception and nmi handling code, see \ref NMSIS_
53   *   Core_IntExcNMI_Handling
54   *
55   * The file configures the device and, typically, initializes the oscillator (PLL)_
56   * that is part
57   * of the microcontroller device. This file might export other functions or variables_
58   * that provide
59   * a more flexible configuration of the microcontroller system.
60   *
61   * And this file also provided common interrupt, exception and NMI exception handling_
62   * framework template,
63   * Silicon vendor can customize these template code as they want.
64   *
65   * \note Please pay special attention to the static variable \c SystemCoreClock. This_
66   *   variable might be
67   *   used throughout the whole system initialization and runtime to calculate frequency/_
68   *   time related values.
69   *   Thus one must assure that the variable always reflects the actual system clock_
70   *   speed.

```

(continues on next page)

(continued from previous page)

```

62  *
63  * \note \ref _init function and \ref _fini function are now deprecated, but still
64  * need to provide a empty implementation
65  * here, due to newlibc still need it. Please put your pre- and post- main steps in
66  * \ref _premain_init and \ref _postmain_fini
67  *
68  * \attention
69  * Be aware that a value stored to \c SystemCoreClock during low level initialization
70  * (i.e. \c SystemInit()) might get
71  * overwritten by C library startup code and/or .bss section initialization.
72  * Thus its highly recommended to call \ref SystemCoreClockUpdate at the beginning of
73  * the user \c main() routine.
74  *
75  * @{
76  */
77  /-----
78  System Core Clock Variable
79  *-----*/
80  /* TODO: initialize SystemCoreClock with the system core clock frequency value
81  achieved after system initialization.
82  This means system core clock frequency after call to SystemInit() */
83  /**
84  * \brief      Variable to hold the system core clock value
85  * \details
86  * Holds the system core clock, which is the system clock frequency supplied to the
87  * SysTick
88  * timer and the processor core clock. This variable can be used by debuggers to
89  * query the
90  * frequency of the debug timer or to configure the trace clock speed.
91  *
92  * \attention
93  * Compilers must be configured to avoid removing this variable in case the
94  * application
95  * program is not using it. Debugging systems require the variable to be physically
96  * present in memory so that it can be examined to configure the debugger.
97  */
98  uint32_t SystemCoreClock = SYSTEM_CLOCK; /* System Clock Frequency (Core Clock) */
99
100 /-----
101 Clock functions
102 *-----*/
103
104 /**
105 * \brief      Function to update the variable \ref SystemCoreClock
106 * \details
107 * Updates the variable \ref SystemCoreClock and must be called whenever the core
108 * clock is changed
109 * during program execution. The function evaluates the clock register settings and
110 * calculates
111 * the current core clock.
112 */
113 void SystemCoreClockUpdate (void) /* Get Core Clock Frequency */
114 {
115     /* TODO: add code to calculate the system frequency based upon the current
116      * register settings.

```

(continues on next page)

(continued from previous page)

```

110     * Note: This function can be used to retrieve the system core clock frequeny
111     *       after user changed register settings.
112     */
113     SystemCoreClock = SYSTEM_CLOCK;
114 }
115
116 /**
117  * \brief      Function to Initialize the system.
118  * \details
119  * Initializes the microcontroller system. Typically, this function configures the
120  * oscillator (PLL) that is part of the microcontroller device. For systems
121  * with a variable clock speed, it updates the variable \ref SystemCoreClock.
122  * SystemInit is called from the file <b>startup<i>_device</i></b>.
123  */
124 void SystemInit (void)
125 {
126     /* TODO: add code to initialize the system
127     * Warn: do not use global variables because this function is called before
128     * reaching pre-main. RW section maybe overwritten afterwards.
129     */
130     SystemCoreClock = SYSTEM_CLOCK;
131 }
132
133 /**
134  * \defgroup NMSIS_Core_IntExcNMI_Handling Interrupt and Exception and NMI Handling
135  * \brief Functions for interrupt, exception and nmi handle available in system_
136  * <device>.c.
137  * \details
138  * Nuclei provide a template for interrupt, exception and NMI handling. Silicon_
139  * Vendor could adapat according
140  * to their requirement. Silicon vendor could implement interface for different_
141  * exception code and
142  * replace current implementation.
143  *
144  * @{
145  */
146 /** \brief Max exception handler number, don't include the NMI(0xFFFF) one */
147 #define MAX_SYSTEM_EXCEPTION_NUM 12
148 /**
149  * \brief      Store the exception handlers for each exception ID
150  * \note
151  * - This SystemExceptionHandlers are used to store all the handlers for all
152  * the exception codes Nuclei N/NX core provided.
153  * - Exception code 0 - 11, totally 12 exceptions are mapped to_
154  * SystemExceptionHandlers[0:11]
155  * - Exception for NMI is also re-routed to exception handling(exception code 0xFFFF)_
156  * in startup code configuration, the handler itself is mapped to_
157  * SystemExceptionHandlers[MAX_SYSTEM_EXCEPTION_NUM]
158  */
159 static unsigned long SystemExceptionHandlers[MAX_SYSTEM_EXCEPTION_NUM+1];
160
161 /**
162  * \brief      Exception Handler Function Typedef
163  * \note
164  * This typedef is only used internal in this system_<Device>.c file.
165  * It is used to do type conversion for registered exception handler before calling_
166  * it.

```

(continues on next page)

(continued from previous page)

```

160  */
161  typedef void (*EXC_HANDLER) (unsigned long mcause, unsigned long sp);
162
163  /**
164   * \brief      System Default Exception Handler
165   * \details
166   * This function provided a default exception and NMI handling code for all exception_
167   * By default, It will just print some information for debug, Vendor can customize it_
168   * according to its requirements.
169   */
170  static void system_default_exception_handler(unsigned long mcause, unsigned long sp)
171  {
172      /* TODO: Uncomment this if you have implement printf function.
173       * Or you can implement your own version as you like */
174      //printf("MCAUSE: 0x%x\r\n", mcause);
175      //printf("MEPC : 0x%x\r\n", __RV_CSR_READ(CSR_MEPC));
176      //printf("MTVAL : 0x%x\r\n", __RV_CSR_READ(CSR_MBADADDR));
177      Exception_DumpFrame(sp);
178      while(1);
179  }
180
181  /**
182   * \brief      Initialize all the default core exception handlers
183   * \details
184   * The core exception handler for each exception id will be initialized to \ref_
185   * system_default_exception_handler.
186   * \note
187   * Called in \ref _init function, used to initialize default exception handlers for_
188   * all exception IDs
189   */
190  static void Exception_Init(void)
191  {
192      for (int i = 0; i < MAX_SYSTEM_EXCEPTION_NUM+1; i++) {
193          SystemExceptionHandlers[i] = (unsigned long)system_default_exception_handler;
194      }
195  }
196
197  /**
198   * \brief      Dump Exception Frame
199   * \details
200   * This function provided feature to dump exception frame stored in stack.
201   */
202  void Exception_DumpFrame(unsigned long sp)
203  {
204      EXC_Frame_Type *exc_frame = (EXC_Frame_Type *)sp;
205
206      #ifndef __riscv_32e
207          printf("ra: 0x%x, tp: 0x%x, t0: 0x%x, t1: 0x%x, t2: 0x%x, t3: 0x%x, t4: 0x%x, t5: 0x%x, t6: 0x%x\n" \
208              "a0: 0x%x, a1: 0x%x, a2: 0x%x, a3: 0x%x, a4: 0x%x, a5: 0x%x, a6: 0x%x, a7: 0x%x\n" \
209              "mcause: 0x%x, mepc: 0x%x, msubm: 0x%x\n", exc_frame->ra, exc_frame->tp,
210              exc_frame->t0, \
211              exc_frame->t1, exc_frame->t2, exc_frame->t3, exc_frame->t4, exc_frame->t5,
212              exc_frame->t6, \
213              exc_frame->a0, exc_frame->a1, exc_frame->a2, exc_frame->a3, exc_frame->a4,
214              exc_frame->a5, \

```

(continues on next page)

(continued from previous page)

```

209     exc_frame->a6, exc_frame->a7, exc_frame->mcause, exc_frame->mepc, exc_
↳frame->msubm);
210 #else
211     printf("ra: 0x%x, tp: 0x%x, t0: 0x%x, t1: 0x%x, t2: 0x%x\n" \
212           "a0: 0x%x, a1: 0x%x, a2: 0x%x, a3: 0x%x, a4: 0x%x, a5: 0x%x\n" \
213           "mcause: 0x%x, mepc: 0x%x, msubm: 0x%x\n", exc_frame->ra, exc_frame->tp,
↳exc_frame->t0, \
214           exc_frame->t1, exc_frame->t2, exc_frame->a0, exc_frame->a1, exc_frame->a2,
↳exc_frame->a3, \
215           exc_frame->a4, exc_frame->a5, exc_frame->mcause, exc_frame->mepc, exc_
↳frame->msubm);
216 #endif
217 }
218
219 /**
220  * \brief      Register an exception handler for exception code EXCn
221  * \details
222  * * For EXCn < \ref MAX_SYSTEM_EXCEPTION_NUM, it will be registered into
↳SystemExceptionHandlers[EXCn-1].
223  * * For EXCn == NMI_EXCn, it will be registered into SystemExceptionHandlers[MAX_
↳SYSTEM_EXCEPTION_NUM].
224  * \param     EXCn      See \ref EXCn_Type
225  * \param     exc_handler  The exception handler for this exception code EXCn
226  */
227 void Exception_Register_EXC(uint32_t EXCn, unsigned long exc_handler)
228 {
229     if ((EXCn < MAX_SYSTEM_EXCEPTION_NUM) && (EXCn >= 0)) {
230         SystemExceptionHandlers[EXCn] = exc_handler;
231     } else if (EXCn == NMI_EXCn) {
232         SystemExceptionHandlers[MAX_SYSTEM_EXCEPTION_NUM] = exc_handler;
233     }
234 }
235
236 /**
237  * \brief      Get current exception handler for exception code EXCn
238  * \details
239  * * For EXCn < \ref MAX_SYSTEM_EXCEPTION_NUM, it will return
↳SystemExceptionHandlers[EXCn-1].
240  * * For EXCn == NMI_EXCn, it will return SystemExceptionHandlers[MAX_SYSTEM_
↳EXCEPTION_NUM].
241  * \param     EXCn      See \ref EXCn_Type
242  * \return     Current exception handler for exception code EXCn, if not found, return 0.
243  */
244 unsigned long Exception_Get_EXC(uint32_t EXCn)
245 {
246     if ((EXCn < MAX_SYSTEM_EXCEPTION_NUM) && (EXCn >= 0)) {
247         return SystemExceptionHandlers[EXCn];
248     } else if (EXCn == NMI_EXCn) {
249         return SystemExceptionHandlers[MAX_SYSTEM_EXCEPTION_NUM];
250     } else {
251         return 0;
252     }
253 }
254
255 /**
256  * \brief      Common NMI and Exception handler entry
257  * \details

```

(continues on next page)

(continued from previous page)

```

258  * This function provided a command entry for NMI and exception. Silicon Vendor could
↪modify
259  * this template implementation according to requirement.
260  * \remarks
261  * - RISCv provided common entry for all types of exception. This is proposed code
↪template
262  * for exception entry function, Silicon Vendor could modify the implementation.
263  * - For the core_exception_handler template, we provided exception register function
↪\ref Exception_Register_EXCn
264  * which can help developer to register your exception handler for specific
↪exception number.
265  */
266  uint32_t core_exception_handler(unsigned long mcause, unsigned long sp)
267  {
268      uint32_t EXCn = (uint32_t)(mcause & 0X00000fff);
269      EXC_HANDLER exc_handler;
270
271      if ((EXCn < MAX_SYSTEM_EXCEPTION_NUM) && (EXCn >= 0)) {
272          exc_handler = (EXC_HANDLER)SystemExceptionHandler[EXCn];
273      } else if (EXCn == NMI_EXCn) {
274          exc_handler = (EXC_HANDLER)SystemExceptionHandler[MAX_SYSTEM_EXCEPTION_NUM];
275      } else {
276          exc_handler = (EXC_HANDLER)system_default_exception_handler;
277      }
278      if (exc_handler != NULL) {
279          exc_handler(mcause, sp);
280      }
281      return 0;
282  }
283
284  /**
285   * \brief Initialize Global ECLIC Config
286   * \details
287   * ECLIC needs be initialized after boot up,
288   * Vendor could also change the initialization
289   * configuration.
290   */
291  void ECLIC_Init(void)
292  {
293      /* Global Configuration about MTH and NLbits.
294       * TODO: Please adapt it according to your system requirement.
295       * This function is called in _init function */
296      /* Set CSR MTH to zero */
297      ECLIC_SetMth(0);
298      /* Set Nlbits to the CLICINTCTLBITS, all the bits are level bits */
299      ECLIC_SetCfgNlbits(__ECLIC_INTCTLBITS);
300  }
301
302  /**
303   * \brief Initialize a specific IRQ and register the handler
304   * \details
305   * This function set vector mode, trigger mode and polarity, interrupt level and
↪priority,
306   * assign handler for specific IRQn.
307   * \param [in] IRQn NMI interrupt handler address
308   * \param [in] shv \ref ECLIC_NON_VECTOR_INTERRUPT means non-vector mode,
↪and \ref ECLIC_VECTOR_INTERRUPT is vector mode

```

(continues on next page)



(continued from previous page)

```

309  * \param [in]  trig_mode    see \ref ECLIC_TRIGGER_Type
310  * \param [in]  lvl          interupt level
311  * \param [in]  priority     interrupt priority
312  * \param [in]  handler      interrupt handler, if NULL, handler will not be installed
313  *
314  * \return       -1 means invalid input parameter. 0 means successful.
315  * \remarks
316  * - This function use to configure specific eclic interrupt and register its_
  ↪ interrupt handler and enable its interrupt.
317  * - If the vector table is placed in read-only section(FLASHXIP mode), handler could_
  ↪ not be installed
318  */
319  int32_t ECLIC_Register_IRQ(IRQn_Type IRQn, uint8_t shv, ECLIC_TRIGGER_Type trig_mode,
  ↪ uint8_t lvl, uint8_t priority, void *handler)
320  {
321      if ((IRQn > SOC_INT_MAX) || (shv > ECLIC_VECTOR_INTERRUPT) \
322          || (trig_mode > ECLIC_NEGTIVE_EDGE_TRIGGER )) {
323          return -1;
324      }
325
326      /* set interrupt vector mode */
327      ECLIC_SetShvIRQ(IRQn, shv);
328      /* set interrupt trigger mode and polarity */
329      ECLIC_SetTrigIRQ(IRQn, trig_mode);
330      /* set interrupt level */
331      ECLIC_SetLevelIRQ(IRQn, lvl);
332      /* set interrupt priority */
333      ECLIC_SetPriorityIRQ(IRQn, priority);
334      if (handler != NULL) {
335          /* set interrupt handler entry to vector table */
336          ECLIC_SetVector(IRQn, (rv_csr_t)handler);
337      }
338      /* enable interrupt */
339      ECLIC_EnableIRQ(IRQn);
340      return 0;
341  }
342  /** @} */ /* End of Doxygen Group NMSIS_Core_ExceptionAndNMI */
343
344  /**
345   * \brief early init function before main
346   * \details
347   * This function is executed right before main function.
348   * For RISC-V gnu toolchain, _init function might not be called
349   * by __libc_init_array function, so we defined a new function
350   * to do initialization
351   */
352  void _premain_init(void)
353  {
354      /* TODO: Add your own initialization code here, called before main */
355      /* __ICACHE_PRESENT and __DCACHE_PRESENT are defined in <Device>.h */
356      #if defined(__ICACHE_PRESENT) && __ICACHE_PRESENT == 1
357          EnableICache();
358      #endif
359      #if defined(__DCACHE_PRESENT) && __DCACHE_PRESENT == 1
360          EnabledDCache();
361      #endif
362      // TODO: Add code to set the system clock frequency value SystemCoreClock

```

(continues on next page)

(continued from previous page)

```

363
364 // TODO: Add code to initialize necessary gpio and basic uart for debug print
365
366 /* Initialize exception default handlers */
367 Exception_Init();
368 /* ECLIC initialization, mainly MTH and NLBIT settings */
369 ECLIC_Init();
370 }
371
372 /**
373  * \brief finish function after main
374  * \param [in] status      status code return from main
375  * \details
376  * This function is executed right after main function.
377  * For RISC-V gnu toolchain, _fini function might not be called
378  * by __libc_fini_array function, so we defined a new function
379  * to do initialization
380  */
381 void _postmain_fini(int status)
382 {
383     /* TODO: Add your own finishing code here, called after main */
384 }
385
386 /**
387  * \brief _init function called in __libc_init_array()
388  * \details
389  * This `__libc_init_array()` function is called during startup code,
390  * user need to implement this function, otherwise when link it will
391  * error init.c:(.text.__libc_init_array+0x26): undefined reference to `_init'
392  * \note
393  * Please use \ref _premain_init function now
394  */
395 void _init(void)
396 {
397     /* Don't put any code here, please use _premain_init now */
398 }
399
400 /**
401  * \brief _fini function called in __libc_fini_array()
402  * \details
403  * This `__libc_fini_array()` function is called when exit main.
404  * user need to implement this function, otherwise when link it will
405  * error fini.c:(.text.__libc_fini_array+0x28): undefined reference to `_fini'
406  * \note
407  * Please use \ref _postmain_fini function now
408  */
409 void _fini(void)
410 {
411     /* Don't put any code here, please use _postmain_fini now */
412 }
413
414 /** @} */ /* End of Doxygen Group NMSIS_Core_SystemAndClock */

```

## system\_Device.h Template File

Here we provided system\_Device.h template file as below:

```

1  /*
2  * Copyright (c) 2009-2018 Arm Limited. All rights reserved.
3  * Copyright (c) 2019 Nuclei Limited. All rights reserved.
4  *
5  * SPDX-License-Identifier: Apache-2.0
6  *
7  * Licensed under the Apache License, Version 2.0 (the License); you may
8  * not use this file except in compliance with the License.
9  * You may obtain a copy of the License at
10 *
11 * www.apache.org/licenses/LICENSE-2.0
12 *
13 * Unless required by applicable law or agreed to in writing, software
14 * distributed under the License is distributed on an AS IS BASIS, WITHOUT
15 * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
16 * See the License for the specific language governing permissions and
17 * limitations under the License.
18 */
19 /*****
20 * @file      system_<Device>.h
21 * @brief     NMSIS Nuclei N/NX Device Peripheral Access Layer Header File for
22 *            Device <Device>
23 * @version   V1.10
24 * @date      30. July 2021
25 *****/
26
27 #ifndef __SYSTEM_<Device>_H__ /* TODO: replace '<Device>' with your device name */
28 #define __SYSTEM_<Device>_H__
29
30 #ifdef __cplusplus
31 extern "C" {
32 #endif
33
34 #include <stdint.h>
35
36 extern uint32_t SystemCoreClock; /*!< System Clock Frequency (Core Clock) */
37
38 /** \brief Exception frame structure store in stack */
39 typedef struct EXC_Frame {
40     unsigned long ra; /* ra: x1, return address for jump */
41     unsigned long tp; /* tp: x4, thread pointer */
42     unsigned long t0; /* t0: x5, temporary register 0 */
43     unsigned long t1; /* t1: x6, temporary register 1 */
44     unsigned long t2; /* t2: x7, temporary register 2 */
45     unsigned long a0; /* a0: x10, return value or function argument 0 */
46     ↪ */
47     unsigned long a1; /* a1: x11, return value or function argument 1 */
48     ↪ */
49     unsigned long a2; /* a2: x12, function argument 2 */
50     unsigned long a3; /* a3: x13, function argument 3 */
51     unsigned long a4; /* a4: x14, function argument 4 */
52     unsigned long a5; /* a5: x15, function argument 5 */
53     unsigned long mcause; /* mcause: machine cause csr register */
54     unsigned long mepc; /* mepc: machine exception program counter csr */
55     ↪ register */
56     unsigned long msubm; /* msubm: machine sub-mode csr register, nuclei */
57     ↪ customized */

```

(continues on next page)

(continued from previous page)

```

54 #ifndef __riscv_32e
55     unsigned long a6;           /* a6: x16, function argument 6 */
56     unsigned long a7;           /* a7: x17, function argument 7 */
57     unsigned long t3;           /* t3: x28, temporary register 3 */
58     unsigned long t4;           /* t4: x29, temporary register 4 */
59     unsigned long t5;           /* t5: x30, temporary register 5 */
60     unsigned long t6;           /* t6: x31, temporary register 6 */
61 #endif
62 } EXC_Frame_Type;
63
64 /**
65  * \brief Setup the microcontroller system.
66  * \details
67  * Initialize the System and update the SystemCoreClock variable.
68  */
69 extern void SystemInit(void);
70
71 /**
72  * \brief Update SystemCoreClock variable.
73  * \details
74  * Updates the SystemCoreClock with current core Clock retrieved from cpu registers.
75  */
76 extern void SystemCoreClockUpdate(void);
77
78 /**
79  * \brief Dump Exception Frame
80  */
81 void Exception_DumpFrame(unsigned long sp);
82
83 /**
84  * \brief Register an exception handler for exception code EXCn
85  */
86 extern void Exception_Register_EXC(uint32_t EXCn, unsigned long exc_handler);
87
88 /**
89  * \brief Get current exception handler for exception code EXCn
90  */
91 extern unsigned long Exception_Get_EXC(uint32_t EXCn);
92
93 /**
94  * \brief Initialize eclic config
95  */
96 extern void ECLIC_Init(void);
97
98 /**
99  * \brief initialize a specific IRQ and register the handler
100  * \details
101  * This function set vector mode, trigger mode and polarity, interrupt level and
102  * ↪priority,
103  * assign handler for specific IRQn.
104  * \param [in] IRQn      NMI interrupt handler address
105  * \param [in] shv        \ref ECLIC_NON_VECTOR_INTERRUPT means non-vector mode,
106  * ↪and \ref ECLIC_VECTOR_INTERRUPT is vector mode
107  * \param [in] trig_mode  see \ref ECLIC_TRIGGER_Type
108  * \param [in] lvl        interrupt level
109  * \param [in] priority    interrupt priority
110  * \param [in] handler     interrupt handler

```

(continues on next page)

(continued from previous page)

```

109  * return      -1 means invalid input parameter. 0 means successful.
110  * \remarks
111  * - This function use to configure specific eclic interrupt and register its_
  ↪ interrupt handler and enable its interrupt.
112  */
113  extern int32_t ECLIC_Register_IRQ(IRQn_Type IRQn, uint8_t shv, ECLIC_TRIGGER_Type_
  ↪ trig_mode, uint8_t lvl, uint8_t priority, void *handler);
114
115  #ifdef __cplusplus
116  }
117  #endif
118
119  #endif /* __SYSTEM_<Device>_H__ */

```

## Device Header File <device.h>

The *Device Header File <device.h>* (page 47) contains the following sections that are device specific:

- *Interrupt Number Definition* (page 47) provides interrupt numbers (IRQn) for all exceptions and interrupts of the device.
- *Configuration of the Processor and Core Peripherals* (page 48) reflect the features of the device.
- *Device Peripheral Access Layer* (page 50) provides definitions for the *Peripheral Access* (page 356) to all device peripherals. It contains all data structures and the address mapping for device-specific peripherals.
- **Access Functions for Peripherals (optional)** provide additional helper functions for peripherals that are useful for programming of these peripherals. Access Functions may be provided as inline functions or can be extern references to a device-specific library provided by the silicon vendor.

*NMSIS Core API* (page 57) describes the standard features and functions of the *Device Header File <device.h>* (page 47) in detail.

## Interrupt Number Definition

*Device Header File <device.h>* (page 47) contains the enumeration `IRQn_Type` that defines all exceptions and interrupts of the

- Negative IRQn values represent processor core exceptions (internal interrupts).
- Positive IRQn values represent device-specific exceptions (external interrupts). The first device-specific interrupt has the IRQn value 0. The IRQn values needs extension to reflect the device-specific interrupt vector table in the *Startup File startup\_<device>.S* (page 14).

The following example shows the extension of the interrupt vector table for the GD32VF103 device family.

```

1  typedef enum IRQn {
2      /***** N200 Processor Exceptions Numbers_
  ↪ *****/
3      Reserved0_IRQn      = 0,      /*!< Internal reserved
  ↪ */
4      Reserved1_IRQn      = 1,      /*!< Internal reserved
  ↪ */
5      Reserved2_IRQn      = 2,      /*!< Internal reserved
  ↪ */
6      SysTimerSW_IRQn     = 3,      /*!< System Timer SW interrupt
  ↪ */

```

(continues on next page)

(continued from previous page)

```

7   Reserved3_IRQn      = 4,      /*!< Internal reserved
↳   */
8   Reserved4_IRQn      = 5,      /*!< Internal reserved
↳   */
9   Reserved5_IRQn      = 6,      /*!< Internal reserved
↳   */
10  SysTimer_IRQn       = 7,      /*!< System Timer Interrupt
↳   */
11  Reserved6_IRQn      = 8,      /*!< Internal reserved
↳   */
12  Reserved7_IRQn      = 9,      /*!< Internal reserved
↳   */
13  Reserved8_IRQn      = 10,     /*!< Internal reserved
↳   */
14  Reserved9_IRQn      = 11,     /*!< Internal reserved
↳   */
15  Reserved10_IRQn     = 12,     /*!< Internal reserved
↳   */
16  Reserved11_IRQn     = 13,     /*!< Internal reserved
↳   */
17  Reserved12_IRQn     = 14,     /*!< Internal reserved
↳   */
18  Reserved13_IRQn     = 15,     /*!< Internal reserved
↳   */
19  Reserved14_IRQn     = 16,     /*!< Internal reserved
↳   */
20  HardFault_IRQn      = 17,     /*!< Hard Fault, storage access error
↳   */
21  Reserved15_IRQn     = 18,     /*!< Internal reserved
↳   */
22
23  /***** GD32VF103 Specific Interrupt Numbers
↳  *****/
24  WWDGT_IRQn          = 19,     /*!< window watchDog timer interrupt
↳   */
25  LVD_IRQn            = 20,     /*!< LVD through EXTI line detect
↳ interrupt */
26  TAMPER_IRQn         = 21,     /*!< tamper through EXTI line detect
↳   */
27  :                   :
28  :                   :
29  CAN1_EWMC_IRQn      = 85,     /*!< CAN1 EWMC interrupt
↳   */
30  USBFS_IRQn          = 86,     /*!< USBFS global interrupt
↳   */
31  SOC_INT_MAX,        /*!< Number of total Interrupts
↳   */
32 } IRQn_Type;

```

## Configuration of the Processor and Core Peripherals

The *Device Header File* <device.h> (page 47) configures the Nuclei N/NX Class Processors and the core peripherals with #define that are set prior to including the file *nmsis\_core.h*.

The following tables list the #define along with the possible values for N200, N300, N600, NX600. If these #define are missing default values are used.

## nmsis\_core.h

Table 6: Macros used in nmsis\_core.h

#define	Value Range	Default	Description
__NUCLEI_N_REV OR __NUCLEI_NX_REV	0x0100   0x0104	0x0100	<ul style="list-style-type: none"> <li>For Nuclei N class device, define __NUCLEI_N_REV, for NX class device, define __NUCLEI_NX_REV.</li> <li>Core revision number ([15:8] revision number, [7:0] patch number), 0x0100 -&gt; 1.0, 0x0104 -&gt; 1.4</li> </ul>
__SYSTIMER_PRESENT	0 .. 1	1	Define whether Private System Timer is present or not. This SysTimer is a Memory Mapped Unit.
__SYSTIMER_BASEADDR	.	0x02000000	Base address of the System Timer Unit.
__ECLIC_PRESENT	0 .. 1	1	Define whether Enhanced Core Local Interrupt Controller (ECLIC) Unit is present or not
__ECLIC_BASEADDR	.	0x0C000000	Base address of the ECLIC unit.
__ECLIC_INTCTLBITS	1 .. 8	1	Define the number of hardware bits are actually implemented in the clicintctl registers.
__ECLIC_INTNUM	1 .. 1024	1	Define the total interrupt number(including the internal core interrupts) of ECLIC Unit
__PMP_PRESENT	0 .. 1	0	Define whether Physical Memory Protection (PMP) Unit is present or not.
__PMP_ENTRY_NUM	8 or 16	8	Define the numbers of PMP entries.
__FPU_PRESENT	0 .. 2	0	Define whether Floating Point Unit (FPU) is present or not. <ul style="list-style-type: none"> <li>0: Not present</li> <li>1: Single precision FPU present</li> <li>2: Double precision FPU present</li> </ul>
__DSP_PRESENT	0 .. 1	0	Define whether Digital Signal Processing Unit (DSP) is present or not.
__ICACHE_PRESENT	0 .. 1	0	Define whether I-Cache Unit is present or not.
__DCACHE_PRESENT	0 .. 1	0	Define whether D-Cache Unit is present or not.
__Vendor_SysTickConfig	0 .. 1	0	If __SYSTIMER_PRESENT is 1, then the __Vendor_SysTickConfig can be set to 0, otherwise it can only set to 1. If this define is set to 1, then the default SysTick_Config and SysTick_Reload function is excluded. In this case, the file Device.h must contain a vendor specific implementation of this function.

## NMSIS Version and Processor Information

The following shows the defines in the *nmsis\_core.h* file that may be used in the *NMSIS-Core Device Templates* (page 12) to verify a minimum version or ensure that the right Nuclei N/NX class is used.

## Device Peripheral Access Layer

The *Device Header File* `<device.h>` (page 47) contains for each peripheral:

- Register Layout Typedef
- Base Address
- Access Definitions

The section *Peripheral Access* (page 356) shows examples for peripheral definitions.

## Device.h Template File

Here we provided `Device.h` template file as below:

```

1  /*
2   * Copyright (c) 2009-2019 Arm Limited. All rights reserved.
3   * Copyright (c) 2019 Nuclei Limited. All rights reserved.
4   *
5   * SPDX-License-Identifier: Apache-2.0
6   *
7   * Licensed under the Apache License, Version 2.0 (the License); you may
8   * not use this file except in compliance with the License.
9   * You may obtain a copy of the License at
10  *
11  * www.apache.org/licenses/LICENSE-2.0
12  *
13  * Unless required by applicable law or agreed to in writing, software
14  * distributed under the License is distributed on an AS IS BASIS, WITHOUT
15  * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
16  * See the License for the specific language governing permissions and
17  * limitations under the License.
18  */
19  /*****
20   * @file      <Device>.h
21   * @brief     NMSIS Nuclei N/NX Core Peripheral Access Layer Header File for
22   *            Device <Device>
23   * @version   V1.10
24   * @date      30. July 2021
25   *****/
26
27  #ifndef __<Device>_H__          /* TODO: replace '<Device>' with your device name */
28  #define __<Device>_H__
29
30  #ifdef __cplusplus
31  extern "C" {
32  #endif
33
34  /* TODO: replace '<Vendor>' with vendor name; add your doxygen comment */
35  /** @addtogroup <Vendor>
36   * @{
37   */
38
39
40  /* TODO: replace '<Device>' with device name; add your doxygen comment */
41  /** @addtogroup <Device>

```

(continues on next page)



(continued from previous page)

```

42  * @{
43  */
44
45
46  /** @addtogroup Configuration_of_NMSIS
47  * @{
48  */
49
50  /** \brief SoC Download mode definition */
51  /* TODO: device vendor can extend more download modes */
52  typedef enum {
53      DOWNLOAD_MODE_FLASHXIP = 0,          /*!< Flashxip download mode */
54      DOWNLOAD_MODE_FLASH = 1,             /*!< Flash download mode */
55      DOWNLOAD_MODE_ILM = 2,               /*!< ilm download mode */
56      DOWNLOAD_MODE_DDR = 3,               /*!< ddr download mode */
57      DOWNLOAD_MODE_MAX,
58  } DownloadMode_Type;
59
60  /*
61  ↳ =====
62  ↳ *
63  ↳ =====
64  ↳ *
65  ↳ =====
66
67  typedef enum IRQn {
68  /* ===== Nuclei N/NX Specific Interrupt Numbers
69  ↳ ===== */
70
71  /* TODO: use this N/NX interrupt numbers if your device is a Nuclei N/NX device */
72      Reserved0_IRQn      = 0,             /*!< Internal reserved */
73      Reserved1_IRQn      = 1,             /*!< Internal reserved */
74      Reserved2_IRQn      = 2,             /*!< Internal reserved */
75      SysTimerSW_IRQn     = 3,             /*!< System Timer SW interrupt */
76      Reserved3_IRQn      = 4,             /*!< Internal reserved */
77      Reserved4_IRQn      = 5,             /*!< Internal reserved */
78      Reserved5_IRQn      = 6,             /*!< Internal reserved */
79      SysTimer_IRQn       = 7,             /*!< System Timer Interrupt */
80      Reserved6_IRQn      = 8,             /*!< Internal reserved */
81      Reserved7_IRQn      = 9,             /*!< Internal reserved */
82      Reserved8_IRQn      = 10,            /*!< Internal reserved */
83      Reserved9_IRQn      = 11,            /*!< Internal reserved */
84      Reserved10_IRQn     = 12,            /*!< Internal reserved */
85      Reserved11_IRQn     = 13,            /*!< Internal reserved */
86      Reserved12_IRQn     = 14,            /*!< Internal reserved */
87      Reserved13_IRQn     = 15,            /*!< Internal reserved */
88      Reserved14_IRQn     = 16,            /*!< Internal reserved */
89      Reserved15_IRQn     = 17,            /*!< Internal reserved */
90      Reserved16_IRQn     = 18,            /*!< Internal reserved */
91
92  /* ===== <Device> Specific Interrupt Numbers
93  ↳ ===== */
94
95  /* TODO: add here your device specific external interrupt numbers. 19~1023 is
96  ↳ reserved number for user. Maximum interrupt supported
97  ↳ could get from clicinfo.NUM_INTERRUPT. According the interrupt handlers
98  ↳ defined in startup_Device.s

```

(continues on next page)

(continued from previous page)

```

91     eg.: Interrupt for Timer#1      eclic_tim0_handler -> TIM0_IRQn */
92     <DeviceInterrupt>_IRQn        = 19,          /*!< Device Interrupt */
93
94     SOC_INT_MAX,                    /* Max SoC interrupt Number */
95 } IRQn_Type;
96
97 /*
98  ↳ =====
99  ↳ */
100
101 /* ===== Exception Code Definition
102  ↳ ===== */
103
104 /* ===== Nuclei N/NX Specific Exception Code
105  ↳ ===== */
106
107 typedef enum EXCn {
108     InsUnalign_EXCn                = 0,          /*!< Instruction address misaligned
109     ↳ */
110     InsAccFault_EXCn               = 1,          /*!< Instruction access fault */
111     IlleIns_EXCn                   = 2,          /*!< Illegal instruction */
112     Break_EXCn                     = 3,          /*!< Breakpoint */
113     LdAddrUnalign_EXCn             = 4,          /*!< Load address misaligned */
114     LdFault_EXCn                   = 5,          /*!< Load access fault */
115     StAddrUnalign_EXCn             = 6,          /*!< Store or AMO address
116     ↳ misaligned */
117     StAccessFault_EXCn             = 7,          /*!< Store or AMO access fault */
118     UmodeEcall_EXCn               = 8,          /*!< Environment call from User
119     ↳ mode */
120     MmodeEcall_EXCn               = 11,         /*!< Environment call from Machine
121     ↳ mode */
122     NMI_EXCn                       = 0xffff,     /*!< NMI interrupt*/
123 } EXCn_Type;
124
125 /*
126  ↳ =====
127  ↳ */
128
129 /* ===== Processor and Core Peripheral Section
130  ↳ ===== */
131
132 /* ===== Configuration of the Nuclei N/NX Processor and Core
133  ↳ ===== */
134
135 /* TODO: set the defines according your Device */
136 /* TODO: define the correct core revision
137  ↳
138  ↳   __NUCLEI_N_REV if your device is a Nuclei-N Class device
139  ↳   __NUCLEI_NX_REV if your device is a Nuclei-NX Class device
140  ↳ */
141
142 #define __NUCLEI_N#_REV            0x0100        /*!< Core Revision rXpY,
143     ↳ version X.Y, change N# to N for Nuclei N class cores, change N# to NX for Nuclei NX
144     ↳ cores */
145
146 /* TODO: define the correct core features for the <Device> */
147 #define __ECLIC_PRESENT            1              /*!< Set to 1 if ECLIC is
148     ↳ present */

```

(continues on next page)

(continued from previous page)

```

129 #define __ECLIC_BASEADDR      0x0C000000UL      /*!< Set to ECLIC baseaddr of
    ↳ your device */
130 #define __ECLIC_INTCTLBITS    8                  /*!< Set to 1 - 8, the number
    ↳ of hardware bits are actually implemented in the clicintctl registers. */
131 #define __ECLIC_INTNUM        51                 /*!< Set to 1 - 1024, total
    ↳ interrupt number of ECLIC Unit */
132 #define __SYSTIMER_PRESENT    1                  /*!< Set to 1 if System Timer
    ↳ is present */
133 #define __SYSTIMER_BASEADDR    0x02000000UL      /*!< Set to SysTimer baseaddr
    ↳ of your device */
134 #define __FPU_PRESENT          1                  /*!< Set to 0, 1, or 2, 0 not
    ↳ present, 1 single floating point unit present, 2 double floating point present
    ↳ */
135 #define __DSP_PRESENT          1                  /*!< Set to 1 if DSP is
    ↳ present */
136 #define __PMP_PRESENT          1                  /*!< Set to 1 if PMP is
    ↳ present */
137 #define __PMP_ENTRY_NUM        16                 /*!< Set to 8 or 16, the
    ↳ number of PMP entries */
138 #define __ICACHE_PRESENT      0                  /*!< Set to 1 if I-Cache is
    ↳ present */
139 #define __DCACHE_PRESENT      0                  /*!< Set to 1 if D-Cache is
    ↳ present */
140 #define __Vendor_SysTickConfig 0                 /*!< Set to 1 if different
    ↳ SysTick Config is used */
141
142 /** @} */ /* End of group Configuration_of_NMSIS */
143
144
145 #include <nmsis_core.h>
146 /* TODO: include your system_<Device>.h file
147    ↳ replace '<Device>' with your device name */
148 #include "system_<Device>.h"                    /*!< <Device> System */
149
150
151 /* ===== Start of section using anonymous unions
    ↳ ===== */
152 #if defined (__GNUC__)
153     /* anonymous unions are enabled by default */
154 #else
155     #warning Not supported compiler type
156 #endif
157
158
159 /*
    ↳ =====
    ↳ */
160 /* ===== Device Specific Peripheral Section
    ↳ ===== */
161 /*
    ↳ =====
    ↳ */
162 /* Macros for memory access operations */
163 #define _REG8P(p, i) ((volatile uint8_t *) ((uintptr_t)(p) +
    ↳ (i)))
164 #define _REG16P(p, i) ((volatile uint16_t *) ((uintptr_t)(p) +
    ↳ (i)))

```

(continues on next page)

(continued from previous page)

```

165 #define _REG32P(p, i) ((volatile uint32_t *) ((uintptr_t)((p) +_
    ↪(i))))
166 #define _REG64P(p, i) ((volatile uint64_t *) ((uintptr_t)((p) +_
    ↪(i))))
167 #define _REG8(p, i) (*(_REG8P(p, i)))
168 #define _REG16(p, i) (*(_REG16P(p, i)))
169 #define _REG32(p, i) (*(_REG32P(p, i)))
170 #define _REG64(p, i) (*(_REG64P(p, i)))
171 #define REG8(addr) _REG8((addr), 0)
172 #define REG16(addr) _REG16((addr), 0)
173 #define REG32(addr) _REG32((addr), 0)
174 #define REG64(addr) _REG64((addr), 0)
175
176 /* Macros for address type convert and access operations */
177 #define ADDR16(addr) ((uint16_t)(uintptr_t)(addr))
178 #define ADDR32(addr) ((uint32_t)(uintptr_t)(addr))
179 #define ADDR64(addr) ((uint64_t)(uintptr_t)(addr))
180 #define ADDR8P(addr) ((uint8_t *) (uintptr_t)(addr))
181 #define ADDR16P(addr) ((uint16_t *) (uintptr_t)(addr))
182 #define ADDR32P(addr) ((uint32_t *) (uintptr_t)(addr))
183 #define ADDR64P(addr) ((uint64_t *) (uintptr_t)(addr))
184
185 /* Macros for Bit Operations */
186 #if __riscv_xlen == 32
187 #define BITMASK_MAX 0xFFFFFFFFUL
188 #define BITOFS_MAX 31
189 #else
190 #define BITMASK_MAX 0xFFFFFFFFFFFFFFFFULL
191 #define BITOFS_MAX 63
192 #endif
193
194 // BIT/BITS only support bit mask for __riscv_xlen
195 // For RISC-V 32 bit, it support mask 32 bit wide
196 // For RISC-V 64 bit, it support mask 64 bit wide
197 #define BIT(ofs) (0x1UL << (ofs))
198 #define BITS(start, end) ((BITMASK_MAX << (start) & (BITMASK_MAX)_
    ↪>> (BITOFS_MAX - (end)))
199 #define GET_BIT(regval, bitofs) (((regval) >> (bitofs)) & 0x1)
200 #define SET_BIT(regval, bitofs) ((regval) |= BIT(bitofs))
201 #define CLR_BIT(regval, bitofs) ((regval) &= (~BIT(bitofs)))
202 #define FLIP_BIT(regval, bitofs) ((regval) ^= BIT(bitofs))
203 #define WRITE_BIT(regval, bitofs, val) CLR_BIT(regval, bitofs); ((regval) |=_
    ↪((val) << bitofs) & BIT(bitofs))
204 #define CHECK_BIT(regval, bitofs) (!((regval) & (0x1UL<<(bitofs))))
205 #define GET_BITS(regval, start, end) (((regval) & BITS((start), (end))) >>_
    ↪(start))
206 #define SET_BITS(regval, start, end) ((regval) |= BITS((start), (end)))
207 #define CLR_BITS(regval, start, end) ((regval) &= (~BITS((start), (end))))
208 #define FLIP_BITS(regval, start, end) ((regval) ^= BITS((start), (end)))
209 #define WRITE_BITS(regval, start, end, val) CLR_BITS(regval, start, end); ((regval)_
    ↪|= ((val) << start) & BITS((start), (end)))
210 #define CHECK_BITS_ALL(regval, start, end) (!((~(regval)) & BITS((start), (end))))
211 #define CHECK_BITS_ANY(regval, start, end) ((regval) & BITS((start), (end)))
212
213 #define BITMASK_SET(regval, mask) ((regval) |= (mask))
214 #define BITMASK_CLR(regval, mask) ((regval) &= (~(mask)))
215 #define BITMASK_FLIP(regval, mask) ((regval) ^= (mask))

```

(continues on next page)

(continued from previous page)

```

216 #define BITMASK_CHECK_ALL(regval, mask)    (!((~(regval)) & (mask)))
217 #define BITMASK_CHECK_ANY(regval, mask)    ((regval) & (mask))
218
219 /** @addtogroup Device_Peripheral_peripherals
220     * @{
221     */
222
223 /* TODO: add here your device specific peripheral access structure typedefs
224     following is an example for UART */
225
226 /*_
227     _
228     _
229     _
230     _
231     _
232     _
233     _
234     _
235     _
236     _
237     _
238     _
239     _
240     _
241     _
242     _
243     _
244     _
245     _
246     _
247     _
248     _
249     _
250     _
251     _
252     _
253     _
254     _
255     _
256     _
257     _
258     _
259     _
260     _
261     _
262     _
263     _
264     _
265     _
266     _
267     _
268     _
269     _
270     _
271     _
272     _
273     _
274     _
275     _
276     _
277     _
278     _
279     _
280     _
281     _
282     _
283     _
284     _
285     _
286     _
287     _
288     _
289     _
290     _
291     _
292     _
293     _
294     _
295     _
296     _
297     _
298     _
299     _
300     _
301     _
302     _
303     _
304     _
305     _
306     _
307     _
308     _
309     _
310     _
311     _
312     _
313     _
314     _
315     _
316     _
317     _
318     _
319     _
320     _
321     _
322     _
323     _
324     _
325     _
326     _
327     _
328     _
329     _
330     _
331     _
332     _
333     _
334     _
335     _
336     _
337     _
338     _
339     _
340     _
341     _
342     _
343     _
344     _
345     _
346     _
347     _
348     _
349     _
350     _
351     _
352     _
353     _
354     _
355     _
356     _
357     _
358     _
359     _
360     _
361     _
362     _
363     _
364     _
365     _
366     _
367     _
368     _
369     _
370     _
371     _
372     _
373     _
374     _
375     _
376     _
377     _
378     _
379     _
380     _
381     _
382     _
383     _
384     _
385     _
386     _
387     _
388     _
389     _
390     _
391     _
392     _
393     _
394     _
395     _
396     _
397     _
398     _
399     _
400     _
401     _
402     _
403     _
404     _
405     _
406     _
407     _
408     _
409     _
410     _
411     _
412     _
413     _
414     _
415     _
416     _
417     _
418     _
419     _
420     _
421     _
422     _
423     _
424     _
425     _
426     _
427     _
428     _
429     _
430     _
431     _
432     _
433     _
434     _
435     _
436     _
437     _
438     _
439     _
440     _
441     _
442     _
443     _
444     _
445     _
446     _
447     _
448     _
449     _
450     _
451     _
452     _
453     _
454     _
455     _
456     _
457     _
458     _
459     _
460     _
461     _
462     _
463     _
464     _
465     _
466     _
467     _
468     _
469     _
470     _
471     _
472     _
473     _
474     _
475     _
476     _
477     _
478     _
479     _
480     _
481     _
482     _
483     _
484     _
485     _
486     _
487     _
488     _
489     _
490     _
491     _
492     _
493     _
494     _
495     _
496     _
497     _
498     _
499     _
500     _
501     _
502     _
503     _
504     _
505     _
506     _
507     _
508     _
509     _
510     _
511     _
512     _
513     _
514     _
515     _
516     _
517     _
518     _
519     _
520     _
521     _
522     _
523     _
524     _
525     _
526     _
527     _
528     _
529     _
530     _
531     _
532     _
533     _
534     _
535     _
536     _
537     _
538     _
539     _
540     _
541     _
542     _
543     _
544     _
545     _
546     _
547     _
548     _
549     _
550     _
551     _
552     _
553     _
554     _
555     _
556     _
557     _
558     _
559     _
560     _
561     _
562     _
563     _
564     _
565     _
566     _
567     _
568     _
569     _
570     _
571     _
572     _
573     _
574     _
575     _
576     _
577     _
578     _
579     _
580     _
581     _
582     _
583     _
584     _
585     _
586     _
587     _
588     _
589     _
590     _
591     _
592     _
593     _
594     _
595     _
596     _
597     _
598     _
599     _
600     _
601     _
602     _
603     _
604     _
605     _
606     _
607     _
608     _
609     _
610     _
611     _
612     _
613     _
614     _
615     _
616     _
617     _
618     _
619     _
620     _
621     _
622     _
623     _
624     _
625     _
626     _
627     _
628     _
629     _
630     _
631     _
632     _
633     _
634     _
635     _
636     _
637     _
638     _
639     _
640     _
641     _
642     _
643     _
644     _
645     _
646     _
647     _
648     _
649     _
650     _
651     _
652     _
653     _
654     _
655     _
656     _
657     _
658     _
659     _
660     _
661     _
662     _
663     _
664     _
665     _
666     _
667     _
668     _
669     _
670     _
671     _
672     _
673     _
674     _
675     _
676     _
677     _
678     _
679     _
680     _
681     _
682     _
683     _
684     _
685     _
686     _
687     _
688     _
689     _
690     _
691     _
692     _
693     _
694     _
695     _
696     _
697     _
698     _
699     _
700     _
701     _
702     _
703     _
704     _
705     _
706     _
707     _
708     _
709     _
710     _
711     _
712     _
713     _
714     _
715     _
716     _
717     _
718     _
719     _
720     _
721     _
722     _
723     _
724     _
725     _
726     _
727     _
728     _
729     _
730     _
731     _
732     _
733     _
734     _
735     _
736     _
737     _
738     _
739     _
740     _
741     _
742     _
743     _
744     _
745     _
746     _
747     _
748     _
749     _
750     _
751     _
752     _
753     _
754     _
755     _
756     _
757     _
758     _
759     _
760     _
761     _
762     _
763     _
764     _
765     _
766     _
767     _
768     _
769     _
770     _
771     _
772     _
773     _
774     _
775     _
776     _
777     _
778     _
779     _
780     _
781     _
782     _
783     _
784     _
785     _
786     _
787     _
788     _
789     _
790     _
791     _
792     _
793     _
794     _
795     _
796     _
797     _
798     _
799     _
800     _
801     _
802     _
803     _
804     _
805     _
806     _
807     _
808     _
809     _
810     _
811     _
812     _
813     _
814     _
815     _
816     _
817     _
818     _
819     _
820     _
821     _
822     _
823     _
824     _
825     _
826     _
827     _
828     _
829     _
830     _
831     _
832     _
833     _
834     _
835     _
836     _
837     _
838     _
839     _
840     _
841     _
842     _
843     _
844     _
845     _
846     _
847     _
848     _
849     _
850     _
851     _
852     _
853     _
854     _
855     _
856     _
857     _
858     _
859     _
860     _
861     _
862     _
863     _
864     _
865     _
866     _
867     _
868     _
869     _
870     _
871     _
872     _
873     _
874     _
875     _
876     _
877     _
878     _
879     _
880     _
881     _
882     _
883     _
884     _
885     _
886     _
887     _
888     _
889     _
890     _
891     _
892     _
893     _
894     _
895     _
896     _
897     _
898     _
899     _
900     _
901     _
902     _
903     _
904     _
905     _
906     _
907     _
908     _
909     _
910     _
911     _
912     _
913     _
914     _
915     _
916     _
917     _
918     _
919     _
920     _
921     _
922     _
923     _
924     _
925     _
926     _
927     _
928     _
929     _
930     _
931     _
932     _
933     _
934     _
935     _
936     _
937     _
938     _
939     _
940     _
941     _
942     _
943     _
944     _
945     _
946     _
947     _
948     _
949     _
950     _
951     _
952     _
953     _
954     _
955     _
956     _
957     _
958     _
959     _
960     _
961     _
962     _
963     _
964     _
965     _
966     _
967     _
968     _
969     _
970     _
971     _
972     _
973     _
974     _
975     _
976     _
977     _
978     _
979     _
980     _
981     _
982     _
983     _
984     _
985     _
986     _
987     _
988     _
989     _
990     _
991     _
992     _
993     _
994     _
995     _
996     _
997     _
998     _
999     _
1000    _

```

(continues on next page)

(continued from previous page)

```

256  /*
257  ↪ =====
258  ↪ */
259  /* TODO: add here your device peripherals base addresses
260      following is an example for timer */
261  /** @addtogroup Device_Peripheral_peripheralAddr
262      * @{
263      */
264
265  /* Peripheral and SRAM base address */
266  #define <DeviceAbbreviation>_FLASH_BASE          (0x00000000UL)
267  ↪      /*!< (FLASH      ) Base Address */
268  #define <DeviceAbbreviation>_SRAM_BASE            (0x20000000UL)
269  ↪      /*!< (SRAM      ) Base Address */
270  #define <DeviceAbbreviation>_PERIPH_BASE          (0x40000000UL)
271  ↪      /*!< (Peripheral) Base Address */
272
273  /* Peripheral memory map */
274  #define <DeviceAbbreviation>_UART0_BASE          (<DeviceAbbreviation>_PERIPH_BASE)
275  ↪      /*!< (UART 0  ) Base Address */
276  #define <DeviceAbbreviation>_I2C_BASE             (<DeviceAbbreviation>_PERIPH_BASE +
277  ↪      0x0800) /*!< (I2C      ) Base Address */
278  #define <DeviceAbbreviation>_GPIO_BASE            (<DeviceAbbreviation>_PERIPH_BASE +
279  ↪      0x1000) /*!< (GPIO      ) Base Address */
280
281  /** @} */ /* End of group Device_Peripheral_peripheralAddr */
282
283  /*
284  ↪ =====
285  ↪ */
286  /* ===== Peripheral declaration ===== */
287  /*
288  ↪ =====
289  ↪ */
290
291  /* TODO: add here your device peripherals pointer definitions
292      following is an example for uart0 */
293  /** @addtogroup Device_Peripheral_declaration
294      * @{
295      */
296  #define <DeviceAbbreviation>_UART0                ((<DeviceAbbreviation>_TMR_TypeDef *)
297  ↪      <DeviceAbbreviation>_UART0_BASE)
298
299  /** @} */ /* End of group <Device> */
300
301  /** @} */ /* End of group <Vendor> */
302
303  #ifdef __cplusplus
304  }
305  #endif

```

(continues on next page)

(continued from previous page)

```
299 #endif /* __<Device>_H__ */
```

## 2.4 Register Mapping

The table below associates some common register names used in NMSIS to the register names used in [Nuclei ISA Spec](#)<sup>12</sup>.

Table 7: Register names used in NMSIS related with the register names in ISA

NMSIS Register Name	N200, N300, N600, NX600	Register Description
<b>Enhanced Core Local Interrupt Controller(ECLIC)</b>		
ECLIC->CFG	cliccfg	ECLIC Global Configuration Register
ECLIC->INFO	clicinfo	ECLIC Global Information Register
ECLIC->MTH	mtm	ECLIC Global Machine Mode Threshold Register
ECLIC->CTRL[i].INTIP	clicintip[i]	ECLIC Interrupt Pending Register
ECLIC->CTRL[i].INTIE	clicintie[i]	ECLIC Interrupt Enable Register
ECLIC->CTRL[i].INTATTR	clicintattr[i]	ECLIC Interrupt Attribute Register
ECLIC->CTRL[i].INTCTRL	clicintctl[i]	ECLIC Interrupt Input Control Register
<b>System Timer Unit(SysTimer)</b>		
SysTimer->MTIMER	mtime_hi<<32 + mtime_lo	System Timer current value 64bits Register
SysTimer->MTIMERCMP	mtimecmp_hi<<32 + mtimecmp_lo	System Timer compare value 64bits Register
SysTimer->MSTOP	mstop	System Timer Stop Register
SysTimer->MSIP	msip	System Timer SW interrupt Register

## 2.5 NMSIS Core API

If you want to access doxygen generated NMSIS Core API, please click [NMSIS Core Doxygen API Documentation](#).

### 2.5.1 Version Control

**group NMSIS\_Core\_VersionControl**

Version #define symbols for NMSIS release specific C/C++ source code.

We followed the [semantic versioning 2.0.0](#)<sup>13</sup> to control NMSIS version. The version format is **MAJOR.MINOR.PATCH**, increment the:

1. MAJOR version when you make incompatible API changes,
2. MINOR version when you add functionality in a backwards compatible manner, and
3. PATCH version when you make backwards compatible bug fixes.

<sup>12</sup> [https://doc.nucleisys.com/nuclei\\_spec/](https://doc.nucleisys.com/nuclei_spec/)

<sup>13</sup> <https://semver.org/>

The header file `nmsis_version.h` is included by each core header so that these definitions are available.

**Example Usage for NMSIS Version Check:**

```
#if defined(__NMSIS_VERSION) && (__NMSIS_VERSION >= 0x00010105)
    #warning "Yes, we have NMSIS 1.1.5 or later"
#else
    #error "We need NMSIS 1.1.5 or later!"
#endif
```

**Unnamed Group****\_\_NUCLEI\_N\_REV (0x0104)**

Nuclei N class core revision number.

Reversion number format: [15:8] revision number, [7:0] patch number

**Attention** This define is exclusive with `__NUCLEI_NX_REV` (page 58)

**\_\_NUCLEI\_NX\_REV (0x0100)**

Nuclei NX class core revision number.

Reversion number format: [15:8] revision number, [7:0] patch number

**Attention** This define is exclusive with `__NUCLEI_N_REV` (page 58)

**Defines****\_\_NMSIS\_VERSION\_MAJOR (1U)**

Represent the NMSIS major version.

The NMSIS major version can be used to differentiate between NMSIS major releases.

**\_\_NMSIS\_VERSION\_MINOR (0U)**

Represent the NMSIS minor version.

The NMSIS minor version can be used to query a NMSIS release update including new features.

**\_\_NMSIS\_VERSION\_PATCH (1U)**

Represent the NMSIS patch version.

The NMSIS patch version can be used to show bug fixes in this package.

**\_\_NMSIS\_VERSION ((\_\_NMSIS\_VERSION\_MAJOR (page 58) << 16U) | (\_\_NMSIS\_VERSION\_MINOR << 8) | \_\_NMSIS\_VERSION\_PATCH)**

Represent the NMSIS Version.

NMSIS Version format: **MAJOR.MINOR.PATCH**

- MAJOR: `__NMSIS_VERSION_MAJOR` (page 58), stored in bits [31:16] of `__NMSIS_VERSION` (page 58)
- MINOR: `__NMSIS_VERSION_MINOR` (page 58), stored in bits [15:8] of `__NMSIS_VERSION` (page 58)
- PATCH: `__NMSIS_VERSION_PATCH` (page 58), stored in bits [7:0] of `__NMSIS_VERSION` (page 58)



## 2.5.2 Compiler Control

### group **NMSIS\_Core\_CompilerControl**

Compiler agnostic #define symbols for generic c/c++ source code.

The NMSIS-Core provides the header file **nmsis\_compiler.h** with consistent #define symbols for generate C or C++ source files that should be compiler agnostic. Each NMSIS compliant compiler should support the functionality described in this section.

The header file **nmsis\_compiler.h** is also included by each Device Header File <device.h> so that these definitions are available.

### Defines

**\_\_has\_builtin** (x) (0)

**\_\_ASM** \_\_asm

Pass information from the compiler to the assembler.

**\_\_INLINE** inline

Recommend that function should be inlined by the compiler.

**\_\_STATIC\_INLINE** static inline

Define a static function that may be inlined by the compiler.

**\_\_STATIC\_FORCEINLINE** \_\_attribute\_\_((always\_inline)) static inline

Define a static function that should be always inlined by the compiler.

**\_\_NO\_RETURN** \_\_attribute\_\_((\_\_noreturn\_\_))

Inform the compiler that a function does not return.

**\_\_USED** \_\_attribute\_\_((used))

Inform that a variable shall be retained in executable image.

**\_\_WEAK** \_\_attribute\_\_((weak))

restrict pointer qualifier to enable additional optimizations.

**\_\_VECTOR\_SIZE** (x) \_\_attribute\_\_((vector\_size(x)))

specified the vector size of the variable, measured in bytes

**\_\_PACKED** \_\_attribute\_\_((packed, aligned(1)))

Request smallest possible alignment.

**\_\_PACKED\_STRUCT** struct \_\_attribute\_\_((packed, aligned(1)))

Request smallest possible alignment for a structure.

**\_\_PACKED\_UNION** union \_\_attribute\_\_((packed, aligned(1)))

Request smallest possible alignment for a union.

**\_\_UNALIGNED\_UINT16\_WRITE** (addr, val) (void)((struct *T\_UINT16\_WRITE* (page 60) \*) (void \*) (addr))->v) = (val))

Pointer for unaligned write of a uint16\_t variable.

**\_\_UNALIGNED\_UINT16\_READ** (addr) (((const struct *T\_UINT16\_READ* (page 60) \*) (const void \*) (addr))->v)

Pointer for unaligned read of a uint16\_t variable.

**\_\_UNALIGNED\_UINT32\_WRITE** (addr, val) (void)((struct *T\_UINT32\_WRITE* (page 60) \*) (void \*) (addr))->v) = (val))

Pointer for unaligned write of a uint32\_t variable.

**\_\_UNALIGNED\_UINT32\_READ** (addr) (((const struct *T\_UINT32\_READ* (page 60) \*) (const void \*) (addr)) -> v)  
Pointer for unaligned read of a uint32\_t variable.

**\_\_ALIGNED** (x) \_\_attribute\_\_((aligned(x)))  
Minimum x bytes alignment for a variable.

**\_\_RESTRICT** \_\_restrict  
restrict pointer qualifier to enable additional optimizations.

**\_\_COMPILER\_BARRIER** () \_\_ASM (page 59) volatile(“”:”:memory”)  
Barrier to prevent compiler from reordering instructions.

**\_\_USUALLY** (exp) \_\_builtin\_expect((exp), 1)  
provide the compiler with branch prediction information, the branch is usually true

**\_\_RARELY** (exp) \_\_builtin\_expect((exp), 0)  
provide the compiler with branch prediction information, the branch is rarely true

**\_\_INTERRUPT** \_\_attribute\_\_((interrupt))  
Use this attribute to indicate that the specified function is an interrupt handler.

## Variables

**\_\_PACKED\_STRUCT T\_UINT16\_WRITE**  
Packed struct for unaligned uint16\_t write access.

**\_\_PACKED\_STRUCT T\_UINT16\_READ**  
Packed struct for unaligned uint16\_t read access.

**\_\_PACKED\_STRUCT T\_UINT32\_WRITE**  
Packed struct for unaligned uint32\_t write access.

**\_\_PACKED\_STRUCT T\_UINT32\_READ**  
Packed struct for unaligned uint32\_t read access.

## 2.5.3 Core CSR Register Access

Click [Nuclei Core CSR<sup>14</sup>](#) to learn about Core CSR in Nuclei ISA Spec.

*group* **NMSIS\_Core\_CSR\_Register\_Access**

Functions to access the Core CSR Registers.

The following functions or macros provide access to Core CSR registers.

- *Core CSR Encodings* (page 71)
- *Core CSR Registers* (page 63)

## Defines

**\_\_RV\_CSR\_SWAP** (csr, val)  
CSR operation Macro for csrrw instruction.  
  
Read the content of csr register to \_\_v, then write content of val into csr register, then return \_\_v  
  
**Return** the CSR register value before written

---

<sup>14</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/core\\_csr.html](https://doc.nucleisys.com/nuclei_spec/isa/core_csr.html)

**Parameters**

- `csr`: CSR macro definition defined in *Core CSR Registers* (page 63), eg. `CSR_MSTATUS` (page 64)
- `val`: value to store into the CSR register

**\_\_RV\_CSR\_READ** (`csr`)

CSR operation Macro for csrr instruction.

Read the content of csr register to `__v` and return it

**Return** the CSR register value

**Parameters**

- `csr`: CSR macro definition defined in *Core CSR Registers* (page 63), eg. `CSR_MSTATUS` (page 64)

**\_\_RV\_CSR\_WRITE** (`csr`, `val`)

CSR operation Macro for csrw instruction.

Write the content of `val` to csr register

**Parameters**

- `csr`: CSR macro definition defined in *Core CSR Registers* (page 63), eg. `CSR_MSTATUS` (page 64)
- `val`: value to store into the CSR register

**\_\_RV\_CSR\_READ\_SET** (`csr`, `val`)

CSR operation Macro for csrrs instruction.

Read the content of csr register to `__v`, then set csr register to be `__v | val`, then return `__v`

**Return** the CSR register value before written

**Parameters**

- `csr`: CSR macro definition defined in *Core CSR Registers* (page 63), eg. `CSR_MSTATUS` (page 64)
- `val`: Mask value to be used with csrrs instruction

**\_\_RV\_CSR\_SET** (`csr`, `val`)

CSR operation Macro for csrs instruction.

Set csr register to be `csr_content | val`

**Parameters**

- `csr`: CSR macro definition defined in *Core CSR Registers* (page 63), eg. `CSR_MSTATUS` (page 64)
- `val`: Mask value to be used with csrs instruction

**\_\_RV\_CSR\_READ\_CLEAR** (`csr`, `val`)

CSR operation Macro for csrrc instruction.

Read the content of csr register to `__v`, then set csr register to be `__v & ~val`, then return `__v`

**Return** the CSR register value before written

**Parameters**

- `csr`: CSR macro definition defined in *Core CSR Registers* (page 63), eg. `CSR_MSTATUS` (page 64)

- `val`: Mask value to be used with `csrrc` instruction

**\_\_RV\_CSR\_CLEAR** (`csr`, `val`)

CSR operation Macro for `csrrc` instruction.

Set `csr` register to be `csr_content` & `~val`

**Parameters**

- `csr`: CSR macro definition defined in *Core CSR Registers* (page 63), eg. `CSR_MSTATUS` (page 64)
- `val`: Mask value to be used with `csrrc` instruction

## Functions

**\_\_STATIC\_FORCEINLINE void \_\_enable\_irq(void)**

Enable IRQ Interrupts.

Enables IRQ interrupts by setting the MIE-bit in the MSTATUS Register.

**Remark** Can only be executed in Privileged modes.

**\_\_STATIC\_FORCEINLINE void \_\_disable\_irq(void)**

Disable IRQ Interrupts.

Disables IRQ interrupts by clearing the MIE-bit in the MSTATUS Register.

**Remark** Can only be executed in Privileged modes.

**\_\_STATIC\_FORCEINLINE uint64\_t \_\_get\_rv\_cycle(void)**

Read whole 64 bits value of `mcycle` counter.

This function will read the whole 64 bits of `MCYCLE` register

**Return** The whole 64 bits value of `MCYCLE`

**Remark** It will work for both RV32 and RV64 to get full 64bits value of `MCYCLE`

**\_\_STATIC\_FORCEINLINE uint64\_t \_\_get\_rv\_instret(void)**

Read whole 64 bits value of machine instruction-retired counter.

This function will read the whole 64 bits of `MINSTRET` register

**Return** The whole 64 bits value of `MINSTRET`

**Remark** It will work for both RV32 and RV64 to get full 64bits value of `MINSTRET`

**\_\_STATIC\_FORCEINLINE uint64\_t \_\_get\_rv\_time(void)**

Read whole 64 bits value of real-time clock.

This function will read the whole 64 bits of `TIME` register

**Return** The whole 64 bits value of `TIME` CSR

**Remark** It will work for both RV32 and RV64 to get full 64bits value of `TIME`

**Attention** only available when user mode available

## 2.5.4 Core CSR Encoding

Click [Nuclei Core CSR<sup>15</sup>](#) to learn about Core CSR in Nuclei ISA Spec.

---

<sup>15</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/core\\_csr.html](https://doc.nucleisys.com/nuclei_spec/isa/core_csr.html)

## Core CSR Register Definitions

### *group* **NMSIS\_Core\_CSR\_Registers**

NMSIS Core CSR Register Definitions.

The following macros are used for CSR Register Definitions.

### **Defines**

```

CSR_USTATUS 0x0
CSR_FFLAGS 0x1
CSR_FRM 0x2
CSR_FCSR 0x3
CSR_CYCLE 0xc00
CSR_TIME 0xc01
CSR_INSTRET 0xc02
CSR_HPMCOUNTER3 0xc03
CSR_HPMCOUNTER4 0xc04
CSR_HPMCOUNTER5 0xc05
CSR_HPMCOUNTER6 0xc06
CSR_HPMCOUNTER7 0xc07
CSR_HPMCOUNTER8 0xc08
CSR_HPMCOUNTER9 0xc09
CSR_HPMCOUNTER10 0xc0a
CSR_HPMCOUNTER11 0xc0b
CSR_HPMCOUNTER12 0xc0c
CSR_HPMCOUNTER13 0xc0d
CSR_HPMCOUNTER14 0xc0e
CSR_HPMCOUNTER15 0xc0f
CSR_HPMCOUNTER16 0xc10
CSR_HPMCOUNTER17 0xc11
CSR_HPMCOUNTER18 0xc12
CSR_HPMCOUNTER19 0xc13
CSR_HPMCOUNTER20 0xc14
CSR_HPMCOUNTER21 0xc15
CSR_HPMCOUNTER22 0xc16
CSR_HPMCOUNTER23 0xc17
CSR_HPMCOUNTER24 0xc18
CSR_HPMCOUNTER25 0xc19

```

CSR\_HPMCOUNTER26 0xc1a  
CSR\_HPMCOUNTER27 0xc1b  
CSR\_HPMCOUNTER28 0xc1c  
CSR\_HPMCOUNTER29 0xc1d  
CSR\_HPMCOUNTER30 0xc1e  
CSR\_HPMCOUNTER31 0xc1f  
CSR\_SSTATUS 0x100  
CSR\_SIE 0x104  
CSR\_STVEC 0x105  
CSR\_SSCRATCH 0x140  
CSR\_SEPC 0x141  
CSR\_SCAUSE 0x142  
CSR\_SBADADDR 0x143  
CSR\_SIP 0x144  
CSR\_SPTBR 0x180  
CSR\_MSTATUS 0x300  
CSR\_MISA 0x301  
CSR\_MEDELEG 0x302  
CSR\_MIDELEG 0x303  
CSR\_MIE 0x304  
CSR\_MTVEC 0x305  
CSR\_MCOUNTEREN 0x306  
CSR\_MSCRATCH 0x340  
CSR\_MEPC 0x341  
CSR\_MCAUSE 0x342  
CSR\_MBADADDR 0x343  
CSR\_MTVAL 0x343  
CSR\_MIP 0x344  
CSR\_PMPCFG0 0x3a0  
CSR\_PMPCFG1 0x3a1  
CSR\_PMPCFG2 0x3a2  
CSR\_PMPCFG3 0x3a3  
CSR\_PMPADDR0 0x3b0  
CSR\_PMPADDR1 0x3b1  
CSR\_PMPADDR2 0x3b2  
CSR\_PMPADDR3 0x3b3

CSR\_PMPADDR4 0x3b4  
CSR\_PMPADDR5 0x3b5  
CSR\_PMPADDR6 0x3b6  
CSR\_PMPADDR7 0x3b7  
CSR\_PMPADDR8 0x3b8  
CSR\_PMPADDR9 0x3b9  
CSR\_PMPADDR10 0x3ba  
CSR\_PMPADDR11 0x3bb  
CSR\_PMPADDR12 0x3bc  
CSR\_PMPADDR13 0x3bd  
CSR\_PMPADDR14 0x3be  
CSR\_PMPADDR15 0x3bf  
CSR\_TSELECT 0x7a0  
CSR\_TDATA1 0x7a1  
CSR\_TDATA2 0x7a2  
CSR\_TDATA3 0x7a3  
CSR\_DCSR 0x7b0  
CSR\_DPC 0x7b1  
CSR\_DSCRATCH 0x7b2  
CSR\_MCYCLE 0xb00  
CSR\_MINSTRET 0xb02  
CSR\_MHPMCOUNTER3 0xb03  
CSR\_MHPMCOUNTER4 0xb04  
CSR\_MHPMCOUNTER5 0xb05  
CSR\_MHPMCOUNTER6 0xb06  
CSR\_MHPMCOUNTER7 0xb07  
CSR\_MHPMCOUNTER8 0xb08  
CSR\_MHPMCOUNTER9 0xb09  
CSR\_MHPMCOUNTER10 0xb0a  
CSR\_MHPMCOUNTER11 0xb0b  
CSR\_MHPMCOUNTER12 0xb0c  
CSR\_MHPMCOUNTER13 0xb0d  
CSR\_MHPMCOUNTER14 0xb0e  
CSR\_MHPMCOUNTER15 0xb0f  
CSR\_MHPMCOUNTER16 0xb10  
CSR\_MHPMCOUNTER17 0xb11

CSR\_MHPMCOUNTER18 0xb12  
CSR\_MHPMCOUNTER19 0xb13  
CSR\_MHPMCOUNTER20 0xb14  
CSR\_MHPMCOUNTER21 0xb15  
CSR\_MHPMCOUNTER22 0xb16  
CSR\_MHPMCOUNTER23 0xb17  
CSR\_MHPMCOUNTER24 0xb18  
CSR\_MHPMCOUNTER25 0xb19  
CSR\_MHPMCOUNTER26 0xb1a  
CSR\_MHPMCOUNTER27 0xb1b  
CSR\_MHPMCOUNTER28 0xb1c  
CSR\_MHPMCOUNTER29 0xb1d  
CSR\_MHPMCOUNTER30 0xb1e  
CSR\_MHPMCOUNTER31 0xb1f  
CSR\_MUCOUNTEREN 0x320  
CSR\_MSCOUNTEREN 0x321  
CSR\_MHPMEVENT3 0x323  
CSR\_MHPMEVENT4 0x324  
CSR\_MHPMEVENT5 0x325  
CSR\_MHPMEVENT6 0x326  
CSR\_MHPMEVENT7 0x327  
CSR\_MHPMEVENT8 0x328  
CSR\_MHPMEVENT9 0x329  
CSR\_MHPMEVENT10 0x32a  
CSR\_MHPMEVENT11 0x32b  
CSR\_MHPMEVENT12 0x32c  
CSR\_MHPMEVENT13 0x32d  
CSR\_MHPMEVENT14 0x32e  
CSR\_MHPMEVENT15 0x32f  
CSR\_MHPMEVENT16 0x330  
CSR\_MHPMEVENT17 0x331  
CSR\_MHPMEVENT18 0x332  
CSR\_MHPMEVENT19 0x333  
CSR\_MHPMEVENT20 0x334  
CSR\_MHPMEVENT21 0x335  
CSR\_MHPMEVENT22 0x336



CSR\_MHPMEVENT23 0x337  
CSR\_MHPMEVENT24 0x338  
CSR\_MHPMEVENT25 0x339  
CSR\_MHPMEVENT26 0x33a  
CSR\_MHPMEVENT27 0x33b  
CSR\_MHPMEVENT28 0x33c  
CSR\_MHPMEVENT29 0x33d  
CSR\_MHPMEVENT30 0x33e  
CSR\_MHPMEVENT31 0x33f  
CSR\_MVENDORID 0xf11  
CSR\_MARCHID 0xf12  
CSR\_MIMPID 0xf13  
CSR\_MHARTID 0xf14  
CSR\_CYCLEH 0xc80  
CSR\_TIMEH 0xc81  
CSR\_INSTRETH 0xc82  
CSR\_HPMCOUNTER3H 0xc83  
CSR\_HPMCOUNTER4H 0xc84  
CSR\_HPMCOUNTER5H 0xc85  
CSR\_HPMCOUNTER6H 0xc86  
CSR\_HPMCOUNTER7H 0xc87  
CSR\_HPMCOUNTER8H 0xc88  
CSR\_HPMCOUNTER9H 0xc89  
CSR\_HPMCOUNTER10H 0xc8a  
CSR\_HPMCOUNTER11H 0xc8b  
CSR\_HPMCOUNTER12H 0xc8c  
CSR\_HPMCOUNTER13H 0xc8d  
CSR\_HPMCOUNTER14H 0xc8e  
CSR\_HPMCOUNTER15H 0xc8f  
CSR\_HPMCOUNTER16H 0xc90  
CSR\_HPMCOUNTER17H 0xc91  
CSR\_HPMCOUNTER18H 0xc92  
CSR\_HPMCOUNTER19H 0xc93  
CSR\_HPMCOUNTER20H 0xc94  
CSR\_HPMCOUNTER21H 0xc95  
CSR\_HPMCOUNTER22H 0xc96

CSR\_HPMCounter23H 0xc97  
CSR\_HPMCounter24H 0xc98  
CSR\_HPMCounter25H 0xc99  
CSR\_HPMCounter26H 0xc9a  
CSR\_HPMCounter27H 0xc9b  
CSR\_HPMCounter28H 0xc9c  
CSR\_HPMCounter29H 0xc9d  
CSR\_HPMCounter30H 0xc9e  
CSR\_HPMCounter31H 0xc9f  
CSR\_MCYCLEH 0xb80  
CSR\_MINSTRETH 0xb82  
CSR\_MHPMCounter3H 0xb83  
CSR\_MHPMCounter4H 0xb84  
CSR\_MHPMCounter5H 0xb85  
CSR\_MHPMCounter6H 0xb86  
CSR\_MHPMCounter7H 0xb87  
CSR\_MHPMCounter8H 0xb88  
CSR\_MHPMCounter9H 0xb89  
CSR\_MHPMCounter10H 0xb8a  
CSR\_MHPMCounter11H 0xb8b  
CSR\_MHPMCounter12H 0xb8c  
CSR\_MHPMCounter13H 0xb8d  
CSR\_MHPMCounter14H 0xb8e  
CSR\_MHPMCounter15H 0xb8f  
CSR\_MHPMCounter16H 0xb90  
CSR\_MHPMCounter17H 0xb91  
CSR\_MHPMCounter18H 0xb92  
CSR\_MHPMCounter19H 0xb93  
CSR\_MHPMCounter20H 0xb94  
CSR\_MHPMCounter21H 0xb95  
CSR\_MHPMCounter22H 0xb96  
CSR\_MHPMCounter23H 0xb97  
CSR\_MHPMCounter24H 0xb98  
CSR\_MHPMCounter25H 0xb99  
CSR\_MHPMCounter26H 0xb9a  
CSR\_MHPMCounter27H 0xb9b

CSR\_MHPMCOUNTER28H 0xb9c  
CSR\_MHPMCOUNTER29H 0xb9d  
CSR\_MHPMCOUNTER30H 0xb9e  
CSR\_MHPMCOUNTER31H 0xb9f  
CSR\_SPMPCFG0 0x1A0  
CSR\_SPMPCFG1 0x1A1  
CSR\_SPMPCFG2 0x1A2  
CSR\_SPMPCFG3 0x1A3  
CSR\_SPMPADDR0 0x1B0  
CSR\_SPMPADDR1 0x1B1  
CSR\_SPMPADDR2 0x1B2  
CSR\_SPMPADDR3 0x1B3  
CSR\_SPMPADDR4 0x1B4  
CSR\_SPMPADDR5 0x1B5  
CSR\_SPMPADDR6 0x1B6  
CSR\_SPMPADDR7 0x1B7  
CSR\_SPMPADDR8 0x1B8  
CSR\_SPMPADDR9 0x1B9  
CSR\_SPMPADDR10 0x1BA  
CSR\_SPMPADDR11 0x1BB  
CSR\_SPMPADDR12 0x1BC  
CSR\_SPMPADDR13 0x1BD  
CSR\_SPMPADDR14 0x1BE  
CSR\_SPMPADDR15 0x1BF  
CSR\_JALSNXTI 0x947  
CSR\_STVT2 0x948  
CSR\_PUSHSCAUSE 0x949  
CSR\_PUSHSEPC 0x94A  
CSR\_MTVT 0x307  
CSR\_MNXTI 0x345  
CSR\_MINTSTATUS 0x346  
CSR\_MSCRATCHCSW 0x348  
CSR\_MSCRATCHCSWL 0x349  
CSR\_MCLICBASE 0x350  
CSR\_UCODE 0x801  
CSR\_MCOUNTINHIBIT 0x320

CSR\_MILM\_CTL 0x7C0  
CSR\_MDLM\_CTL 0x7C1  
CSR\_MECC\_CODE 0x7C2  
CSR\_MNVEC 0x7C3  
CSR\_MSUBM 0x7C4  
CSR\_MDCAUSE 0x7C9  
CSR\_MCACHE\_CTL 0x7CA  
CSR\_MMISC\_CTL 0x7D0  
CSR\_MSAVESTATUS 0x7D6  
CSR\_MSAVEEPC1 0x7D7  
CSR\_MSAVECAUSE1 0x7D8  
CSR\_MSAVEEPC2 0x7D9  
CSR\_MSAVECAUSE2 0x7DA  
CSR\_MSAVEDCAUSE1 0x7DB  
CSR\_MSAVEDCAUSE2 0x7DC  
CSR\_MTLB\_CTL 0x7DD  
CSR\_MECC\_LOCK 0x7DE  
CSR\_MFP16MODE 0x7E2  
CSR\_LSTEPFORC 0x7E9  
CSR\_PUSHMSUBM 0x7EB  
CSR\_MTVT2 0x7EC  
CSR\_JALMNXTI 0x7ED  
CSR\_PUSHMCAUSE 0x7EE  
CSR\_PUSHMEPC 0x7EF  
CSR\_MPPICFG\_INFO 0x7F0  
CSR\_MFIOCFG\_INFO 0x7F1  
CSR\_MSMPCFG\_INFO 0x7F7  
CSR\_SLEEPVALUE 0x811  
CSR\_TXEVT 0x812  
CSR\_WFE 0x810  
CSR\_MICFG\_INFO 0xFC0  
CSR\_MDCFG\_INFO 0xFC1  
CSR\_MCFG\_INFO 0xFC2  
CSR\_MTLBCFG\_INFO 0xFC3  
CSR\_CCM\_MBEGINADDR 0x7CB  
CSR\_CCM\_MCOMMAND 0x7CC

```

CSR_CCM_MDATA 0x7CD
CSR_CCM_SUEN 0x7CE
CSR_CCM_SBEGINADDR 0x5CB
CSR_CCM_SCOMMAND 0x5CC
CSR_CCM_SDATA 0x5CD
CSR_CCM_UBEGINADDR 0x4CB
CSR_CCM_UCOMMAND 0x4CC
CSR_CCM_UDATA 0x4CD
CSR_CCM_FPIPE 0x4CF

```

## Other Core Related Macros

*group* **NMSIS\_Core\_CSR\_Encoding**  
NMSIS Core CSR Encodings.

The following macros are used for CSR encodings

### Defines

```

MSTATUS_UIE 0x00000001
MSTATUS_SIE 0x00000002
MSTATUS_HIE 0x00000004
MSTATUS_MIE 0x00000008
MSTATUS_UPIE 0x00000010
MSTATUS_SPIE 0x00000020
MSTATUS_HPIE 0x00000040
MSTATUS_MPIE 0x00000080
MSTATUS_SPP 0x00000100
MSTATUS_MPP 0x00001800
MSTATUS_FS 0x00006000
MSTATUS_XS 0x00018000
MSTATUS_MPRV 0x00020000
MSTATUS_PUM 0x00040000
MSTATUS_MXR 0x00080000
MSTATUS_VM 0x1F000000
MSTATUS32_SD 0x80000000
MSTATUS64_SD 0x8000000000000000
MSTATUS_FS_INITIAL 0x00002000
MSTATUS_FS_CLEAN 0x00004000

```

**MSTATUS\_FS\_DIRTY** 0x00006000  
**SSTATUS\_UIE** 0x00000001  
**SSTATUS\_SIE** 0x00000002  
**SSTATUS\_UPIE** 0x00000010  
**SSTATUS\_SPIE** 0x00000020  
**SSTATUS\_SPP** 0x00000100  
**SSTATUS\_FS** 0x00006000  
**SSTATUS\_XS** 0x00018000  
**SSTATUS\_PUM** 0x00040000  
**SSTATUS32\_SD** 0x80000000  
**SSTATUS64\_SD** 0x8000000000000000  
**CSR\_MCACHE\_CTL\_IE** 0x00000001  
**CSR\_MCACHE\_CTL\_DE** 0x00010000  
**DCSR\_XDEBUGVER** (3U<<30)  
**DCSR\_NDRESET** (1<<29)  
**DCSR\_FULLRESET** (1<<28)  
**DCSR\_EBREAKM** (1<<15)  
**DCSR\_EBREAKH** (1<<14)  
**DCSR\_EBREAKS** (1<<13)  
**DCSR\_EBREAKU** (1<<12)  
**DCSR\_STOPCYCLE** (1<<10)  
**DCSR\_STOPTIME** (1<<9)  
**DCSR\_CAUSE** (7<<6)  
**DCSR\_DEBUGINT** (1<<5)  
**DCSR\_HALT** (1<<3)  
**DCSR\_STEP** (1<<2)  
**DCSR\_PRV** (3<<0)  
**DCSR\_CAUSE\_NONE** 0  
**DCSR\_CAUSE\_SWBP** 1  
**DCSR\_CAUSE\_HWBP** 2  
**DCSR\_CAUSE\_DEBUGINT** 3  
**DCSR\_CAUSE\_STEP** 4  
**DCSR\_CAUSE\_HALT** 5  
**MCONTROL\_TYPE** (xlen) (0xfULL<<((xlen)-4))  
**MCONTROL\_DMODE** (xlen) (1ULL<<((xlen)-5))  
**MCONTROL\_MASKMAX** (xlen) (0x3fULL<<((xlen)-11))

**MCONTROL\_SELECT** (1<<19)  
**MCONTROL\_TIMING** (1<<18)  
**MCONTROL\_ACTION** (0x3f<<12)  
**MCONTROL\_CHAIN** (1<<11)  
**MCONTROL\_MATCH** (0xf<<7)  
**MCONTROL\_M** (1<<6)  
**MCONTROL\_H** (1<<5)  
**MCONTROL\_S** (1<<4)  
**MCONTROL\_U** (1<<3)  
**MCONTROL\_EXECUTE** (1<<2)  
**MCONTROL\_STORE** (1<<1)  
**MCONTROL\_LOAD** (1<<0)  
**MCONTROL\_TYPE\_NONE** 0  
**MCONTROL\_TYPE\_MATCH** 2  
**MCONTROL\_ACTION\_DEBUG\_EXCEPTION** 0  
**MCONTROL\_ACTION\_DEBUG\_MODE** 1  
**MCONTROL\_ACTION\_TRACE\_START** 2  
**MCONTROL\_ACTION\_TRACE\_STOP** 3  
**MCONTROL\_ACTION\_TRACE\_EMIT** 4  
**MCONTROL\_MATCH\_EQUAL** 0  
**MCONTROL\_MATCH\_NAPOT** 1  
**MCONTROL\_MATCH\_GE** 2  
**MCONTROL\_MATCH\_LT** 3  
**MCONTROL\_MATCH\_MASK\_LOW** 4  
**MCONTROL\_MATCH\_MASK\_HIGH** 5  
**MIP\_SSIP** (1 << *IRQ\_S\_SOFT* (page 76))  
**MIP\_HSIP** (1 << *IRQ\_H\_SOFT* (page 76))  
**MIP\_MSIP** (1 << *IRQ\_M\_SOFT* (page 76))  
**MIP\_STIP** (1 << *IRQ\_S\_TIMER* (page 76))  
**MIP\_HTIP** (1 << *IRQ\_H\_TIMER* (page 76))  
**MIP\_MTIP** (1 << *IRQ\_M\_TIMER* (page 76))  
**MIP\_SEIP** (1 << *IRQ\_S\_EXT* (page 76))  
**MIP\_HEIP** (1 << *IRQ\_H\_EXT* (page 76))  
**MIP\_MEIP** (1 << *IRQ\_M\_EXT* (page 76))  
**MIE\_SSIE** *MIP\_SSIP* (page 73)  
**MIE\_HSIE** *MIP\_HSIP* (page 73)

**MIE\_MSIE** *MIP\_MSIP* (page 73)  
**MIE\_STIE** *MIP\_STIP* (page 73)  
**MIE\_HTIE** *MIP\_HTIP* (page 73)  
**MIE\_MTIE** *MIP\_MTIP* (page 73)  
**MIE\_SEIE** *MIP\_SEIP* (page 73)  
**MIE\_HEIE** *MIP\_HEIP* (page 73)  
**MIE\_MEIE** *MIP\_MEIP* (page 73)  
**UCODE\_OV** (0x1)  
**WFE\_WFE** (0x1)  
**TXEVT\_TXEVT** (0x1)  
**SLEEPVALUE\_SLEEPVALUE** (0x1)  
**MCOUNTINHIBIT\_IR** (1<<2)  
**MCOUNTINHIBIT\_CY** (1<<0)  
**MILM\_CTL\_ILM\_BPA** (((1ULL<<((\_\_riscv\_xlen)-10))-1)<<10)  
**MILM\_CTL\_ILM\_RWECC** (1<<3)  
**MILM\_CTL\_ILM\_ECC\_EXCP\_EN** (1<<2)  
**MILM\_CTL\_ILM\_ECC\_EN** (1<<1)  
**MILM\_CTL\_ILM\_EN** (1<<0)  
**MDLM\_CTL\_DLM\_BPA** (((1ULL<<((\_\_riscv\_xlen)-10))-1)<<10)  
**MDLM\_CTL\_DLM\_RWECC** (1<<3)  
**MDLM\_CTL\_DLM\_ECC\_EXCP\_EN** (1<<2)  
**MDLM\_CTL\_DLM\_ECC\_EN** (1<<1)  
**MDLM\_CTL\_DLM\_EN** (1<<0)  
**MSUBM\_PTyp** (0x3<<8)  
**MSUBM\_Typ** (0x3<<6)  
**MDCAUSE\_MDCAUSE** (0x3)  
**MMISC\_CTL\_NMI\_CAUSE\_FFF** (1<<9)  
**MMISC\_CTL\_MISALIGN** (1<<6)  
**MMISC\_CTL\_BPU** (1<<3)  
**MCACHE\_CTL\_IC\_EN** (1<<0)  
**MCACHE\_CTL\_IC\_SCPD\_MOD** (1<<1)  
**MCACHE\_CTL\_IC\_ECC\_EN** (1<<2)  
**MCACHE\_CTL\_IC\_ECC\_EXCP\_EN** (1<<3)  
**MCACHE\_CTL\_IC\_RWTECC** (1<<4)  
**MCACHE\_CTL\_IC\_RWDECC** (1<<5)  
**MCACHE\_CTL\_DC\_EN** (1<<16)



**MCACHE\_CTL\_DC\_ECC\_EN** (1<<17)  
**MCACHE\_CTL\_DC\_ECC\_EXCP\_EN** (1<<18)  
**MCACHE\_CTL\_DC\_RWTECC** (1<<19)  
**MCACHE\_CTL\_DC\_RWDECC** (1<<20)  
**MTVT2\_MTVT2EN** (1<<0)  
**MTVT2\_COMMON\_CODE\_ENTRY** (((1ULL<<((\_\_riscv\_xlen)-2))-1)<<2)  
**MCFG\_INFO\_TEE** (1<<0)  
**MCFG\_INFO\_ECC** (1<<1)  
**MCFG\_INFO\_CLIC** (1<<2)  
**MCFG\_INFO\_PLIC** (1<<3)  
**MCFG\_INFO\_FIO** (1<<4)  
**MCFG\_INFO\_PPI** (1<<5)  
**MCFG\_INFO\_NICE** (1<<6)  
**MCFG\_INFO\_ILM** (1<<7)  
**MCFG\_INFO\_DLM** (1<<8)  
**MCFG\_INFO\_ICACHE** (1<<9)  
**MCFG\_INFO\_DCACHE** (1<<10)  
**MICFG\_IC\_SET** (0xF<<0)  
**MICFG\_IC\_WAY** (0x7<<4)  
**MICFG\_IC\_LSIZE** (0x7<<7)  
**MICFG\_IC\_ECC** (0x1<<10)  
**MICFG\_ILM\_SIZE** (0x1F<<16)  
**MICFG\_ILM\_XONLY** (0x1<<21)  
**MICFG\_ILM\_ECC** (0x1<<22)  
**MDCFG\_DC\_SET** (0xF<<0)  
**MDCFG\_DC\_WAY** (0x7<<4)  
**MDCFG\_DC\_LSIZE** (0x7<<7)  
**MDCFG\_DC\_ECC** (0x1<<10)  
**MDCFG\_DLM\_SIZE** (0x1F<<16)  
**MDCFG\_DLM\_ECC** (0x1<<21)  
**MPPICFG\_INFO\_PPI\_SIZE** (0x1F<<1)  
**MPPICFG\_INFO\_PPI\_BPA** (((1ULL<<((\_\_riscv\_xlen)-10))-1)<<10)  
**MFIOCFG\_INFO\_FIO\_SIZE** (0x1F<<1)  
**MFIOCFG\_INFO\_FIO\_BPA** (((1ULL<<((\_\_riscv\_xlen)-10))-1)<<10)  
**MECC\_LOCK\_ECC\_LOCK** (0x1)  
**MECC\_CODE\_CODE** (0x1FF)

**MECC\_CODE\_RAMID** (0x1F<<16)  
**MECC\_CODE\_SRAMID** (0x1F<<24)  
**CCM\_SUEN\_SUEN** (0x1<<0)  
**CCM\_DATA\_DATA** (0x7<<0)  
**CCM\_COMMAND\_COMMAND** (0x1F<<0)  
**SIP\_SSIP** *MIP\_SSIP* (page 73)  
**SIP\_STIP** *MIP\_STIP* (page 73)  
**PRV\_U** 0  
**PRV\_S** 1  
**PRV\_H** 2  
**PRV\_M** 3  
**VM\_MBARE** 0  
**VM\_MBB** 1  
**VM\_MBBID** 2  
**VM\_SV32** 8  
**VM\_SV39** 9  
**VM\_SV48** 10  
**IRQ\_S\_SOFT** 1  
**IRQ\_H\_SOFT** 2  
**IRQ\_M\_SOFT** 3  
**IRQ\_S\_TIMER** 5  
**IRQ\_H\_TIMER** 6  
**IRQ\_M\_TIMER** 7  
**IRQ\_S\_EXT** 9  
**IRQ\_H\_EXT** 10  
**IRQ\_M\_EXT** 11  
**IRQ\_COP** 12  
**IRQ\_HOST** 13  
**FRM\_RNDMODE\_RNE** 0x0  
FPU Round to Nearest, ties to Even.  
**FRM\_RNDMODE\_RTZ** 0x1  
FPU Round Towards Zero.  
**FRM\_RNDMODE\_RDN** 0x2  
FPU Round Down (towards -inf)  
**FRM\_RNDMODE\_RUP** 0x3  
FPU Round Up (towards +inf)  
**FRM\_RNDMODE\_RMM** 0x4  
FPU Round to nearest, ties to Max Magnitude.

**FRM\_RNDMODE\_DYN** 0x7

In instruction's rm, selects dynamic rounding mode.

In Rounding Mode register, Invalid

**FFLAGS\_AE\_NX** (1<<0)

FPU Inexact.

**FFLAGS\_AE\_UF** (1<<1)

FPU Underflow.

**FFLAGS\_AE\_OF** (1<<2)

FPU Overflow.

**FFLAGS\_AE\_DZ** (1<<3)

FPU Divide by Zero.

**FFLAGS\_AE\_NV** (1<<4)

FPU Invalid Operation.

**FREG** (idx) f##idx

Floating Point Register f0-f31, eg.

f0 -> *FREG(0)* (page 77)

**PMP\_R** 0x01

**PMP\_W** 0x02

**PMP\_X** 0x04

**PMP\_A** 0x18

**PMP\_A\_TOR** 0x08

**PMP\_A\_NA4** 0x10

**PMP\_A\_NAPOT** 0x18

**PMP\_L** 0x80

**PMP\_SHIFT** 2

**PMP\_COUNT** 16

**PTE\_V** 0x001

**PTE\_R** 0x002

**PTE\_W** 0x004

**PTE\_X** 0x008

**PTE\_U** 0x010

**PTE\_G** 0x020

**PTE\_A** 0x040

**PTE\_D** 0x080

**PTE\_SOFT** 0x300

**PTE\_PPN\_SHIFT** 10

**PTE\_TABLE** (PTE) (((PTE) & (*PTE\_V* (page 77) | *PTE\_R* (page 77) | *PTE\_W* (page 77) | *PTE\_X* (page 77))) == *PTE\_V* (page 77))

**CAUSE\_MISALIGNED\_FETCH** 0x0  
End of Doxygen Group NMSIS\_Core\_CSR\_Registers.

**CAUSE\_FAULT\_FETCH** 0x1

**CAUSE\_ILLEGAL\_INSTRUCTION** 0x2

**CAUSE\_BREAKPOINT** 0x3

**CAUSE\_MISALIGNED\_LOAD** 0x4

**CAUSE\_FAULT\_LOAD** 0x5

**CAUSE\_MISALIGNED\_STORE** 0x6

**CAUSE\_FAULT\_STORE** 0x7

**CAUSE\_USER\_ECALL** 0x8

**CAUSE\_SUPERVISOR\_ECALL** 0x9

**CAUSE\_HYPERVISOR\_ECALL** 0xa

**CAUSE\_MACHINE\_ECALL** 0xb

**DCAUSE\_FAULT\_FETCH\_PMP** 0x1

**DCAUSE\_FAULT\_FETCH\_INST** 0x2

**DCAUSE\_FAULT\_LOAD\_PMP** 0x1

**DCAUSE\_FAULT\_LOAD\_INST** 0x2

**DCAUSE\_FAULT\_LOAD\_NICE** 0x3

**DCAUSE\_FAULT\_STORE\_PMP** 0x1

**DCAUSE\_FAULT\_STORE\_INST** 0x2

## 2.5.5 Register Define and Type Definitions

*group* **NMSIS\_Core\_Registers**  
Type definitions and defines for core registers.

### Defines

**\_\_RISCV\_XLEN** 32  
Refer to the width of an integer register in bits(either 32 or 64)

### Typedefs

**typedef** uint32\_t **rv\_csr\_t**  
Type of Control and Status Register(CSR), depends on the XLEN defined in RISC-V.

### Core

*group* **NMSIS\_Core\_Base\_Registers**  
Type definitions and defines for base core registers.

## Typedefs

**typedef** *CSR\_MMISCTRL\_Type* (page 83) **CSR\_MMISCTRL\_Type**

**union** **CSR\_MISA\_Type**

*#include <core\_feature\_base.h>* Union type to access MISA CSR register.

### Public Members

*rv\_csr\_t* (page 78) **a**

bit: 0 Atomic extension

*rv\_csr\_t* (page 78) **b**

bit: 1 Tentatively reserved for Bit-Manipulation extension

*rv\_csr\_t* (page 78) **c**

bit: 2 Compressed extension

*rv\_csr\_t* (page 78) **d**

bit: 3 Double-precision floating-point extension

Type used for csr data access.

*rv\_csr\_t* (page 78) **e**

bit: 4 RV32E base ISA

*rv\_csr\_t* (page 78) **f**

bit: 5 Single-precision floating-point extension

*rv\_csr\_t* (page 78) **g**

bit: 6 Additional standard extensions present

*rv\_csr\_t* (page 78) **h**

bit: 7 Hypervisor extension

*rv\_csr\_t* (page 78) **i**

bit: 8 RV32I/64I/128I base ISA

*rv\_csr\_t* (page 78) **j**

bit: 9 Tentatively reserved for Dynamically Translated Languages extension

*rv\_csr\_t* (page 78) **\_reserved1**

bit: 10 Reserved

*rv\_csr\_t* (page 78) **l**

bit: 11 Tentatively reserved for Decimal Floating-Point extension

*rv\_csr\_t* (page 78) **m**

bit: 12 Integer Multiply/Divide extension

*rv\_csr\_t* (page 78) **n**

bit: 13 User-level interrupts supported

*rv\_csr\_t* (page 78) **\_reserved2**

bit: 14 Reserved

*rv\_csr\_t* (page 78) **p**

bit: 15 Tentatively reserved for Packed-SIMD extension

*rv\_csr\_t* (page 78) **q**

bit: 16 Quad-precision floating-point extension

*rv\_csr\_t* (page 78) **\_resreved3**  
bit: 17 Reserved

*rv\_csr\_t* (page 78) **s**  
bit: 18 Supervisor mode implemented

*rv\_csr\_t* (page 78) **t**  
bit: 19 Tentatively reserved for Transactional Memory extension

*rv\_csr\_t* (page 78) **u**  
bit: 20 User mode implemented

*rv\_csr\_t* (page 78) **v**  
bit: 21 Tentatively reserved for Vector extension

*rv\_csr\_t* (page 78) **\_reserved4**  
bit: 22 Reserved

*rv\_csr\_t* (page 78) **x**  
bit: 23 Non-standard extensions present

*rv\_csr\_t* (page 78) **\_reserved5**  
bit: 24..29 Reserved

*rv\_csr\_t* (page 78) **mx1**  
bit: 30..31 Machine XLEN

**struct** *CSR\_MISA\_Type* (page 79)::[anonymous] **b**  
Structure used for bit access.

**union** *CSR\_MSTATUS\_Type*  
*#include <core\_feature\_base.h>* Union type to access MSTATUS CSR register.

### Public Members

*rv\_csr\_t* (page 78) **\_reserved0**  
bit: 0 Reserved

*rv\_csr\_t* (page 78) **sie**  
bit: 1 supervisor interrupt enable flag

*rv\_csr\_t* (page 78) **\_reserved1**  
bit: 2 Reserved

*rv\_csr\_t* (page 78) **mie**  
bit: 3 Machine mode interrupt enable flag

*rv\_csr\_t* (page 78) **\_reserved2**  
bit: 4 Reserved

*rv\_csr\_t* (page 78) **spie**  
bit: 3 Supervisor Privileded mode interrupt enable flag

*rv\_csr\_t* (page 78) **\_reserved3**  
bit: Reserved

*rv\_csr\_t* (page 78) **mpie**  
bit: mirror of MIE flag

*rv\_csr\_t* (page 78) **\_reserved4**  
bit: Reserved

*rv\_csr\_t* (page 78) **mpp**  
bit: mirror of Privilege Mode

*rv\_csr\_t* (page 78) **fs**  
bit: FS status flag

*rv\_csr\_t* (page 78) **xs**  
bit: XS status flag

*rv\_csr\_t* (page 78) **mprv**  
bit: Machine mode PMP

*rv\_csr\_t* (page 78) **sum**  
bit: Supervisor Mode load and store protection

*rv\_csr\_t* (page 78) **\_reserved6**  
bit: 19..30 Reserved

*rv\_csr\_t* (page 78) **sd**  
bit: Dirty status for XS or FS

**struct** *CSR\_MSTATUS\_Type* (page 80)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** *CSR\_MTVEC\_Type*

*#include <core\_feature\_base.h>* Union type to access MTVEC CSR register.

### Public Members

*rv\_csr\_t* (page 78) **mode**  
bit: 0..5 interrupt mode control

*rv\_csr\_t* (page 78) **addr**  
bit: 6..31 mtvec address

**struct** *CSR\_MTVEC\_Type* (page 81)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** *CSR\_MCAUSE\_Type*

*#include <core\_feature\_base.h>* Union type to access MCAUSE CSR register.

### Public Members

*rv\_csr\_t* (page 78) **exccode**  
bit: 11..0 exception or interrupt code

*rv\_csr\_t* (page 78) **\_reserved0**  
bit: 15..12 Reserved

*rv\_csr\_t* (page 78) **mpil**  
bit: 23..16 Previous interrupt level

*rv\_csr\_t* (page 78) **\_reserved1**  
bit: 26..24 Reserved

*rv\_csr\_t* (page 78) **mpie**  
bit: 27 Interrupt enable flag before enter interrupt

*rv\_csr\_t* (page 78) **mpp**  
bit: 29..28 Priviledede mode flag before enter interrupt

*rv\_csr\_t* (page 78) **minhv**  
bit: 30 Machine interrupt vector table

*rv\_csr\_t* (page 78) **interrupt**  
bit: 31 trap type.  
0 means exception and 1 means interrupt

**struct** *CSR\_MCAUSE\_Type* (page 81)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** *CSR\_MCOUNTINHIBIT\_Type*  
*#include <core\_feature\_base.h>* Union type to access MCOUNTINHIBIT CSR register.

### Public Members

*rv\_csr\_t* (page 78) **cy**  
bit: 0 1 means disable mcycle counter

*rv\_csr\_t* (page 78) **\_reserved0**  
bit: 1 Reserved

*rv\_csr\_t* (page 78) **ir**  
bit: 2 1 means disable minstret counter

*rv\_csr\_t* (page 78) **\_reserved1**  
bit: 3..31 Reserved

**struct** *CSR\_MCOUNTINHIBIT\_Type* (page 82)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** *CSR\_MSUBM\_Type*  
*#include <core\_feature\_base.h>* Union type to access MSUBM CSR register.

### Public Members

*rv\_csr\_t* (page 78) **\_reserved0**  
bit: 0..5 Reserved

*rv\_csr\_t* (page 78) **typ**  
bit: 6..7 current trap type

*rv\_csr\_t* (page 78) **ptyp**  
bit: 8..9 previous trap type

*rv\_csr\_t* (page 78) **\_reserved1**  
bit: 10..31 Reserved



**struct** *CSR\_MSUBM\_Type* (page 82)::[anonymous] **b**  
 Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
 Type used for csr data access.

**union** *CSR\_MDCAUSE\_Type*  
*#include <core\_feature\_base.h>* Union type to access MDCAUSE CSR register.

### Public Members

*rv\_csr\_t* (page 78) **mdcause**  
 bit: 0..1 More detailed exception information as MCAUSE supplement

*rv\_csr\_t* (page 78) **\_reserved0**  
 bit: 2..XLEN-1 Reserved

**struct** *CSR\_MDCAUSE\_Type* (page 83)::[anonymous] **b**  
 Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
 Type used for csr data access.

**union** *CSR\_MMISCCTRL\_Type*  
*#include <core\_feature\_base.h>* Union type to access MMISC\_CTRL CSR register.

### Public Members

*rv\_csr\_t* (page 78) **\_reserved0**  
 bit: 0..2 Reserved

*rv\_csr\_t* (page 78) **bpu**  
 bit: 3 dynamic prediction enable flag

*rv\_csr\_t* (page 78) **\_reserved1**  
 bit: 4..5 Reserved

*rv\_csr\_t* (page 78) **misalign**  
 bit: 6 misaligned access support flag

*rv\_csr\_t* (page 78) **\_reserved2**  
 bit: 7..8 Reserved

*rv\_csr\_t* (page 78) **nmi\_cause**  
 bit: 9 mnvec control and nmi mcase exccode

*rv\_csr\_t* (page 78) **\_reserved3**  
 bit: 10..31 Reserved

**struct** *CSR\_MMISCCTRL\_Type* (page 83)::[anonymous] **b**  
 Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
 Type used for csr data access.

**union** *CSR\_MCACHECTL\_Type*  
*#include <core\_feature\_base.h>* Union type to access MCACHE\_CTL CSR register.

### Public Members

*rv\_csr\_t* (page 78) **ic\_en**  
I-Cache enable.

*rv\_csr\_t* (page 78) **ic\_scpd\_mod**  
Scratchpad mode, 0: Scratchpad as ICache Data RAM, 1: Scratchpad as ILM SRAM.

*rv\_csr\_t* (page 78) **ic\_ecc\_en**  
I-Cache ECC enable.

*rv\_csr\_t* (page 78) **ic\_ecc\_excp\_en**  
I-Cache 2bit ECC error exception enable.

*rv\_csr\_t* (page 78) **ic\_rwtecc**  
Control I-Cache Tag Ram ECC code injection.

*rv\_csr\_t* (page 78) **ic\_rwdecc**  
Control I-Cache Data Ram ECC code injection.

*rv\_csr\_t* (page 78) **\_reserved0**

*rv\_csr\_t* (page 78) **dc\_en**  
DCache enable.

*rv\_csr\_t* (page 78) **dc\_ecc\_en**  
D-Cache ECC enable.

*rv\_csr\_t* (page 78) **dc\_ecc\_excp\_en**  
D-Cache 2bit ECC error exception enable.

*rv\_csr\_t* (page 78) **dc\_rwtecc**  
Control D-Cache Tag Ram ECC code injection.

*rv\_csr\_t* (page 78) **dc\_rwdecc**  
Control D-Cache Data Ram ECC code injection.

*rv\_csr\_t* (page 78) **\_reserved1**

**struct** *CSR\_MCACHECTL\_Type* (page 83)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** *CSR\_MSAVESTATUS\_Type*  
*#include <core\_feature\_base.h>* Union type to access MSAVESTATUS CSR register.

### Public Members

*rv\_csr\_t* (page 78) **mpie1**  
bit: 0 interrupt enable flag of first level NMI/exception nesting

*rv\_csr\_t* (page 78) **mpp1**  
bit: 1..2 privileged mode of first level NMI/exception nesting

*rv\_csr\_t* (page 78) **\_reserved0**  
bit: 3..5 Reserved

*rv\_csr\_t* (page 78) **ptyp1**  
bit: 6..7 NMI/exception type of before first nesting

*rv\_csr\_t* (page 78) **mpie2**  
bit: 8 interrupt enable flag of second level NMI/exception nestting

*rv\_csr\_t* (page 78) **mpp2**  
bit: 9..10 priviled mode of second level NMI/exception nestting

*rv\_csr\_t* (page 78) **\_reserved1**  
bit: 11..13 Reserved

*rv\_csr\_t* (page 78) **ptyp2**  
bit: 14..15 NMI/exception type of before second nestting

*rv\_csr\_t* (page 78) **\_reserved2**  
bit: 16..31 Reserved

**struct** *CSR\_MSAVESTATUS\_Type* (page 84)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **w**  
Type used for csr data access.

**union** *CSR\_MILMCTL\_Type*  
*#include <core\_feature\_base.h>* Union type to access MILM\_CTL CSR register.

### Public Members

*rv\_csr\_t* (page 78) **ilm\_en**  
ILM enable.

*rv\_csr\_t* (page 78) **ilm\_ecc\_en**  
ILM ECC eanble.

*rv\_csr\_t* (page 78) **ilm\_ecc\_excp\_en**  
ILM ECC exception enable.

*rv\_csr\_t* (page 78) **ilm\_rwecc**  
Control mecc\_code write to ilm, simulate error injection.

*rv\_csr\_t* (page 78) **\_reserved0**  
Reserved.

*rv\_csr\_t* (page 78) **ilm\_bpa**  
ILM base address.

**struct** *CSR\_MILMCTL\_Type* (page 85)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** *CSR\_MDLMCTL\_Type*  
*#include <core\_feature\_base.h>* Union type to access MDLM\_CTL CSR register.

### Public Members

*rv\_csr\_t* (page 78) **d1m\_en**  
DLM enable.

*rv\_csr\_t* (page 78) **d1m\_ecc\_en**  
DLM ECC eanble.

*rv\_csr\_t* (page 78) **d1m\_ecc\_excp\_en**  
DLM ECC exception enable.

*rv\_csr\_t* (page 78) **d1m\_rwecc**  
Control mecc\_code write to d1m, simulate error injection.

*rv\_csr\_t* (page 78) **\_reserved0**  
Reserved.

*rv\_csr\_t* (page 78) **d1m\_bpa**  
DLM base address.

**struct** *CSR\_MDLMCTL\_Type* (page 85)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** *CSR\_MCFGINFO\_Type*  
*#include <core\_feature\_base.h>* Union type to access MCFG\_INFO CSR register.

### Public Members

*rv\_csr\_t* (page 78) **tee**  
TEE present.

*rv\_csr\_t* (page 78) **ecc**  
ECC present.

*rv\_csr\_t* (page 78) **clic**  
CLIC present.

*rv\_csr\_t* (page 78) **plic**  
PLIC present.

*rv\_csr\_t* (page 78) **fio**  
FIO present.

*rv\_csr\_t* (page 78) **ppl**  
PPI present.

*rv\_csr\_t* (page 78) **nice**  
NICE present.

*rv\_csr\_t* (page 78) **ilm**  
ILM present.

*rv\_csr\_t* (page 78) **d1m**  
DLM present.

*rv\_csr\_t* (page 78) **icache**  
ICache present.

*rv\_csr\_t* (page 78) **dcache**  
DCache present.

*rv\_csr\_t* (page 78) **\_reserved0**

**struct** *CSR\_MCFGINFO\_Type* (page 86)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union CSR\_MICFGINFO\_Type**

*#include <core\_feature\_base.h>* Union type to access MICFG\_INFO CSR register.

**Public Members**

*rv\_csr\_t* (page 78) **set**

I-Cache sets per way.

*rv\_csr\_t* (page 78) **way**

I-Cache way.

*rv\_csr\_t* (page 78) **lsize**

I-Cache line size.

*rv\_csr\_t* (page 78) **cache\_ecc**

I-Cache ECC present.

*rv\_csr\_t* (page 78) **\_reserved0**

*rv\_csr\_t* (page 78) **lm\_size**

ILM size, need to be  $2^n$  size.

*rv\_csr\_t* (page 78) **lm\_xonly**

ILM Execute only permission.

*rv\_csr\_t* (page 78) **lm\_ecc**

ILM ECC present.

*rv\_csr\_t* (page 78) **\_reserved1**

**struct CSR\_MICFGINFO\_Type** (page 87)::[anonymous] **b**

Structure used for bit access.

*rv\_csr\_t* (page 78) **d**

Type used for csr data access.

**union CSR\_MDCFGINFO\_Type**

*#include <core\_feature\_base.h>* Union type to access MDCFG\_INFO CSR register.

**Public Members**

*rv\_csr\_t* (page 78) **set**

D-Cache sets per way.

*rv\_csr\_t* (page 78) **way**

D-Cache way.

*rv\_csr\_t* (page 78) **lsize**

D-Cache line size.

*rv\_csr\_t* (page 78) **cache\_ecc**

D-Cache ECC present.

*rv\_csr\_t* (page 78) **\_reserved0**

*rv\_csr\_t* (page 78) **lm\_size**

DLM size, need to be  $2^n$  size.

*rv\_csr\_t* (page 78) **lm\_xonly**

DLM Execute only permission.

*rv\_csr\_t* (page 78) **lm\_ecc**  
DLM ECC present.

*rv\_csr\_t* (page 78) **\_reserved1**

**struct** *CSR\_MDCFGINFO\_Type* (page 87)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** **CSR\_MPPICFGINFO\_Type**  
*#include <core\_feature\_base.h>* Union type to access MPPICFG\_INFO CSR register.

### Public Members

*rv\_csr\_t* (page 78) **\_reserved0**  
Reserved.

*rv\_csr\_t* (page 78) **ppi\_size**  
PPI size, need to be 2^n size.

*rv\_csr\_t* (page 78) **\_reserved1**  
Reserved.

*rv\_csr\_t* (page 78) **ppi\_bpa**  
PPI base address.

**struct** *CSR\_MPPICFGINFO\_Type* (page 88)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** **CSR\_MFIOCFGINFO\_Type**  
*#include <core\_feature\_base.h>* Union type to access MFIOCFG\_INFO CSR register.

### Public Members

*rv\_csr\_t* (page 78) **\_reserved0**  
Reserved.

*rv\_csr\_t* (page 78) **fio\_size**  
FIO size, need to be 2^n size.

*rv\_csr\_t* (page 78) **\_reserved1**  
Reserved.

*rv\_csr\_t* (page 78) **fio\_bpa**  
FIO base address.

**struct** *CSR\_MFIOCFGINFO\_Type* (page 88)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union** **CSR\_MECCLOCK\_Type**  
*#include <core\_feature\_base.h>* Union type to access MECC\_LOCK CSR register.

### Public Members

*rv\_csr\_t* (page 78) **ecc\_lock**  
RW permission, ECC Lock configure.

*rv\_csr\_t* (page 78) **\_reserved0**  
Reserved.

**struct CSR\_MECCLOCK\_Type** (page 88)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

**union CSR\_MECCCODE\_Type**  
*#include <core\_feature\_base.h>* Union type to access MECC\_CODE CSR register.

### Public Members

*rv\_csr\_t* (page 78) **code**  
Used to inject ECC check code.

*rv\_csr\_t* (page 78) **\_reserved0**  
Reserved.

*rv\_csr\_t* (page 78) **ramid**  
Indicate 2bit ECC error, software can clear these bits.

*rv\_csr\_t* (page 78) **\_reserved1**  
Reserved.

*rv\_csr\_t* (page 78) **sramid**  
Indicate 1bit ECC error, software can clear these bits.

*rv\_csr\_t* (page 78) **\_reserved2**  
Reserved.

**struct CSR\_MECCCODE\_Type** (page 89)::[anonymous] **b**  
Structure used for bit access.

*rv\_csr\_t* (page 78) **d**  
Type used for csr data access.

## ECLIC

*group* **NMSIS\_Core\_ECLIC\_Registers**  
Type definitions and defines for eclic registers.

### Defines

**CLIC\_CLICCFG\_NLBIT\_Pos** 1U  
CLIC CLICCFG: NLBIT Position.

**CLIC\_CLICCFG\_NLBIT\_Msk** (0xFUL << CLIC\_CLICCFG\_NLBIT\_Pos)  
CLIC CLICCFG: NLBIT Mask.

**CLIC\_CLICINFO\_CTLBIT\_Pos** 21U  
CLIC INTINFO: \_\_ECLIC\_GetInfoCtlbits() Position.

**CLIC\_CLICINFO\_CTLBIT\_Msk** (0xFUL << CLIC\_CLICINFO\_CTLBIT\_Pos)  
CLIC INTINFO: \_\_ECLIC\_GetInfoCtlbits() Mask.

**CLIC\_CLICINFO\_VER\_Pos** 13U  
CLIC CLICINFO: VERSION Position.

**CLIC\_CLICINFO\_VER\_Msk** (0xFFUL << CLIC\_CLICCFG\_NLBIT\_Pos)  
CLIC CLICINFO: VERSION Mask.

**CLIC\_CLICINFO\_NUM\_Pos** 0U  
CLIC CLICINFO: NUM Position.

**CLIC\_CLICINFO\_NUM\_Msk** (0xFFFUL << CLIC\_CLICINFO\_NUM\_Pos)  
CLIC CLICINFO: NUM Mask.

**CLIC\_INTIP\_IP\_Pos** 0U  
CLIC INTIP: IP Position.

**CLIC\_INTIP\_IP\_Msk** (0x1UL << CLIC\_INTIP\_IP\_Pos)  
CLIC INTIP: IP Mask.

**CLIC\_INTIE\_IE\_Pos** 0U  
CLIC INTIE: IE Position.

**CLIC\_INTIE\_IE\_Msk** (0x1UL << CLIC\_INTIE\_IE\_Pos)  
CLIC INTIE: IE Mask.

**CLIC\_INTATTR\_TRIG\_Pos** 1U  
CLIC INTATTR: TRIG Position.

**CLIC\_INTATTR\_TRIG\_Msk** (0x3UL << CLIC\_INTATTR\_TRIG\_Pos)  
CLIC INTATTR: TRIG Mask.

**CLIC\_INTATTR\_SHV\_Pos** 0U  
CLIC INTATTR: SHV Position.

**CLIC\_INTATTR\_SHV\_Msk** (0x1UL << CLIC\_INTATTR\_SHV\_Pos)  
CLIC INTATTR: SHV Mask.

**ECLIC\_MAX\_NLBITS** 8U  
Max nlbit of the CLICINTCTLBITS.

**ECLIC\_MODE\_MTVEC\_Msk** 3U  
ECLIC Mode mask for MTVT CSR Register.

**ECLIC\_NON\_VECTOR\_INTERRUPT** 0x0  
Non-Vector Interrupt Mode of ECLIC.

**ECLIC\_VECTOR\_INTERRUPT** 0x1  
Vector Interrupt Mode of ECLIC.

**ECLIC\_BASE \_\_ECLIC\_BASEADDR**  
ECLIC Base Address.

**ECLIC** ((CLIC\_Type \*) *ECLIC\_BASE* (page 90))  
CLIC configuration struct.

## Enums

**enum ECLIC\_TRIGGER\_Type**  
ECLIC Trigger Enum for different Trigger Type.

*Values:*



**ECLIC\_LEVEL\_TRIGGER** = 0x0

Level Triggerred, trig[0] = 0.

**ECLIC\_POSTIVE\_EDGE\_TRIGGER** = 0x1

Postive/Rising Edge Triggered, trig[0] = 1, trig[1] = 0.

**ECLIC\_NEGTIVE\_EDGE\_TRIGGER** = 0x3

Negative/Falling Edge Triggered, trig[0] = 1, trig[1] = 1.

**ECLIC\_MAX\_TRIGGER** = 0x3

MAX Supported Trigger Mode.

**union CLICCFG\_Type**

*#include <core\_feature\_eclic.h>* Union type to access CLICCFG configure register.

### Public Members

uint8\_t **\_reserved0**

bit: 0 Overflow condition code flag

uint8\_t **nlbits**

bit: 29 Carry condition code flag

uint8\_t **\_reserved1**

bit: 30 Zero condition code flag

uint8\_t **\_reserved2**

bit: 31 Negative condition code flag

**struct CLICCFG\_Type** (page 91)::[anonymous] **b**

Structure used for bit access.

uint8\_t **w**

Type used for byte access.

**union CLICINFO\_Type**

*#include <core\_feature\_eclic.h>* Union type to access CLICINFO information register.

### Public Members

uint32\_t **numint**

bit: 0..12 number of maximum interrupt inputs supported

uint32\_t **version**

bit: 13..20 20:17 for architecture version,16:13 for implementation version

uint32\_t **intctlbits**

bit: 21..24 specifies how many hardware bits are actually implemented in the clicintctl registers

uint32\_t **\_reserved0**

bit: 25..31 Reserved

**struct CLICINFO\_Type** (page 91)::[anonymous] **b**

Structure used for bit access.

uint32\_t **w**

Type used for word access.

**struct CLIC\_CTRL\_Type**

*#include <core\_feature\_eclic.h>* Access to the structure of a vector interrupt controller.

## SysTimer

### group NMSIS\_Core\_SysTimer\_Registers

Type definitions and defines for system timer registers.

#### Defines

**SysTimer\_MTIMECTL\_TIMESTOP\_Pos** 0U

SysTick Timer MTIMECTL: TIMESTOP bit Position.

**SysTimer\_MTIMECTL\_TIMESTOP\_Msk** (1UL << SysTimer\_MTIMECTL\_TIMESTOP\_Pos)

SysTick Timer MTIMECTL: TIMESTOP Mask.

**SysTimer\_MTIMECTL\_CMPCLREN\_Pos** 1U

SysTick Timer MTIMECTL: CMPCLREN bit Position.

**SysTimer\_MTIMECTL\_CMPCLREN\_Msk** (1UL << SysTimer\_MTIMECTL\_CMPCLREN\_Pos)

SysTick Timer MTIMECTL: CMPCLREN Mask.

**SysTimer\_MTIMECTL\_CLKSRC\_Pos** 2U

SysTick Timer MTIMECTL: CLKSRC bit Position.

**SysTimer\_MTIMECTL\_CLKSRC\_Msk** (1UL << SysTimer\_MTIMECTL\_CLKSRC\_Pos)

SysTick Timer MTIMECTL: CLKSRC Mask.

**SysTimer\_MSIP\_MSIP\_Pos** 0U

SysTick Timer MSIP: MSIP bit Position.

**SysTimer\_MSIP\_MSIP\_Msk** (1UL << SysTimer\_MSIP\_MSIP\_Pos)

SysTick Timer MSIP: MSIP Mask.

**SysTimer\_MTIMER\_Msk** (0xFFFFFFFFFFFFFFFFFULL)

SysTick Timer MTIMER value Mask.

**SysTimer\_MTIMERCMP\_Msk** (0xFFFFFFFFFFFFFFFFFULL)

SysTick Timer MTIMERCMP value Mask.

**SysTimer\_MTIMECTL\_Msk** (0xFFFFFFFFFUL)

SysTick Timer MTIMECTL/MSTOP value Mask.

**SysTimer\_MSIP\_Msk** (0xFFFFFFFFFUL)

SysTick Timer MSIP value Mask.

**SysTimer\_MSFRST\_Msk** (0xFFFFFFFFFUL)

SysTick Timer MSFRST value Mask.

**SysTimer\_MSFRST\_KEY** (0x80000A5FUL)

SysTick Timer Software Reset Request Key.

**SysTimer\_BASE\_\_SYSTIMER\_BASEADDR**

SysTick Base Address.

**SysTimer** ((*SysTimer\_Type* (page 92) \*) *SysTimer\_BASE* (page 92))

SysTick configuration struct.

**struct SysTimer\_Type**

*#include <core\_feature\_timer.h>* Structure type to access the System Timer (SysTimer).

Structure definition to access the system timer(SysTimer).

**Remark**

- MSFTRST register is introduced in Nuclei N Core version 1.3([\\_\\_NUCLEI\\_N\\_REV](#) (page 58) >= 0x0103)
- MSTOP register is renamed to MTIMECTL register in Nuclei N Core version 1.4([\\_\\_NUCLEI\\_N\\_REV](#) (page 58) >= 0x0104)
- CMPCLREN and CLKSRC bit in MTIMECTL register is introduced in Nuclei N Core version 1.4([\\_\\_NUCLEI\\_N\\_REV](#) (page 58) >= 0x0104)

## 2.5.6 CPU Intrinsic Functions

**enum** NMSIS\_Core\_CPU\_Intrinsic::WFI\_SleepMode\_Type

*Values:*

WFI\_SHALLOW\_SLEEP = 0

WFI\_DEEP\_SLEEP = 1

**\_\_STATIC\_FORCEINLINE** void \_\_NOP(void)

**\_\_STATIC\_FORCEINLINE** void \_\_WFI(void)

**\_\_STATIC\_FORCEINLINE** void \_\_WFE(void)

**\_\_STATIC\_FORCEINLINE** void \_\_EBREAK(void)

**\_\_STATIC\_FORCEINLINE** void \_\_ECALL(void)

**\_\_STATIC\_FORCEINLINE** void \_\_set\_wfi\_sleepmode(WFI\_SleepMode\_Type mode)

**\_\_STATIC\_FORCEINLINE** void \_\_TXEVT(void)

**\_\_STATIC\_FORCEINLINE** void \_\_enable\_mcycle\_counter(void)

**\_\_STATIC\_FORCEINLINE** void \_\_disable\_mcycle\_counter(void)

**\_\_STATIC\_FORCEINLINE** void \_\_enable\_minstret\_counter(void)

**\_\_STATIC\_FORCEINLINE** void \_\_disable\_minstret\_counter(void)

**\_\_STATIC\_FORCEINLINE** void \_\_enable\_all\_counter(void)

**\_\_STATIC\_FORCEINLINE** void \_\_disable\_all\_counter(void)

**\_\_STATIC\_FORCEINLINE** void \_\_FENCE\_I(void)

**\_\_STATIC\_FORCEINLINE** uint8\_t \_\_LB(volatile void \* addr)

**\_\_STATIC\_FORCEINLINE** uint16\_t \_\_LH(volatile void \* addr)

**\_\_STATIC\_FORCEINLINE** uint32\_t \_\_LW(volatile void \* addr)

**\_\_STATIC\_FORCEINLINE** void \_\_SB(volatile void \* addr, uint8\_t val)

**\_\_STATIC\_FORCEINLINE** void \_\_SH(volatile void \* addr, uint16\_t val)

**\_\_STATIC\_FORCEINLINE** void \_\_SW(volatile void \* addr, uint32\_t val)

**\_\_STATIC\_FORCEINLINE** uint32\_t \_\_CAS\_W(volatile uint32\_t \* addr, uint32\_t oldval, uint32\_t newval)

**\_\_STATIC\_FORCEINLINE** uint32\_t \_\_AMOSWAP\_W(volatile uint32\_t \* addr, uint32\_t newval)

**\_\_STATIC\_FORCEINLINE** int32\_t \_\_AMOADD\_W(volatile int32\_t \* addr, int32\_t value)

**\_\_STATIC\_FORCEINLINE** int32\_t \_\_AMOAND\_W(volatile int32\_t \* addr, int32\_t value)

**\_\_STATIC\_FORCEINLINE** int32\_t \_\_AMOOR\_W(volatile int32\_t \* addr, int32\_t value)

```

__STATIC_FORCEINLINE int32_t __AMOXOR_W(volatile int32_t * addr, int32_t value)
__STATIC_FORCEINLINE uint32_t __AMOMAXU_W(volatile uint32_t * addr, uint32_t value)
__STATIC_FORCEINLINE int32_t __AMOMAX_W(volatile int32_t * addr, int32_t value)
__STATIC_FORCEINLINE uint32_t __AMOMINU_W(volatile uint32_t * addr, uint32_t value)
__STATIC_FORCEINLINE int32_t __AMOMIN_W(volatile int32_t * addr, int32_t value)
__FENCE(p, s) __ASM (page 59) volatile (“fence ” #p “,” #s : : “memory”)
__RWMB () __FENCE(iorw,iorw)
__RMB () __FENCE(ir,ir)
__WMB () __FENCE(ow,ow)
__SMP_RWMB () __FENCE(rw,rw)
__SMP_RMB () __FENCE(r,r)
__SMP_WMB () __FENCE(w,w)
__CPU_RELAX () __ASM (page 59) volatile (“” : : “memory”)

```

#### group NMSIS\_Core\_CPU\_Intrinsic

Functions that generate RISC-V CPU instructions.

The following functions generate specified RISC-V instructions that cannot be directly accessed by compiler.

### Defines

**\_\_FENCE** (p, s) [\\_\\_ASM](#) (page 59) volatile (“fence ” #p “,” #s : : “memory”)  
Execute fence instruction, p -> pred, s -> succ.

the FENCE instruction ensures that all memory accesses from instructions preceding the fence in program order (the predecessor set) appear earlier in the global memory order than memory accesses from instructions appearing after the fence in program order (the successor set). For details, please refer to The RISC-V Instruction Set Manual

#### Parameters

- p: predecessor set, such as iorw, rw, r, w
- s: successor set, such as iorw, rw, r, w

**\_\_RWMB** () **\_\_FENCE**(iorw,iorw)  
Read & Write Memory barrier.

**\_\_RMB** () **\_\_FENCE**(ir,ir)  
Read Memory barrier.

**\_\_WMB** () **\_\_FENCE**(ow,ow)  
Write Memory barrier.

**\_\_SMP\_RWMB** () **\_\_FENCE**(rw,rw)  
SMP Read & Write Memory barrier.

**\_\_SMP\_RMB** () **\_\_FENCE**(r,r)  
SMP Read Memory barrier.

**\_\_SMP\_WMB** () **\_\_FENCE**(w,w)  
SMP Write Memory barrier.

**\_\_CPU\_RELAX()** *\_\_ASM* (page 59) volatile (“” : : : “memory”)  
CPU relax for busy loop.

## Enums

**enum WFI\_SleepMode\_Type**  
WFI Sleep Mode enumeration.

*Values:*

**WFI\_SHALLOW\_SLEEP** = 0  
Shallow sleep mode, the core\_clk will poweroff.

**WFI\_DEEP\_SLEEP** = 1  
Deep sleep mode, the core\_clk and core\_ano\_clk will poweroff.

## Functions

**\_\_STATIC\_FORCEINLINE void \_\_NOP(void)**  
NOP Instruction.

No Operation does nothing. This instruction can be used for code alignment purposes.

**\_\_STATIC\_FORCEINLINE void \_\_WFI(void)**  
Wait For Interrupt.

Wait For Interrupt is executed using CSR\_WFE.WFE=0 and WFI instruction. It will suspends execution until interrupt, NMI or Debug happened. When Core is waked up by interrupt, if

1. mstatus.MIE == 1(interrupt enabled), Core will enter ISR code
2. mstatus.MIE == 0(interrupt disabled), Core will resume previous execution

**\_\_STATIC\_FORCEINLINE void \_\_WFE(void)**  
Wait For Event.

Wait For Event is executed using CSR\_WFE.WFE=1 and WFI instruction. It will suspends execution until event, NMI or Debug happened. When Core is waked up, Core will resume previous execution

**\_\_STATIC\_FORCEINLINE void \_\_EBREAK(void)**  
Breakpoint Instruction.

Causes the processor to enter Debug state. Debug tools can use this to investigate system state when the instruction at a particular address is reached.

**\_\_STATIC\_FORCEINLINE void \_\_ECALL(void)**  
Environment Call Instruction.

The ECALL instruction is used to make a service request to the execution environment.

**\_\_STATIC\_FORCEINLINE void \_\_set\_wfi\_sleepmode(WFI\_SleepMode\_Type mode)**  
Set Sleep mode of WFI.

Set the SLEEPVALUE CSR register to control the WFI Sleep mode.

### Parameters

- [in] mode: The sleep mode to be set

**\_\_STATIC\_FORCEINLINE void \_\_TXEVT(void)**  
Send TX Event.

Set the CSR TXEVT to control send a TX Event. The Core will output signal tx\_evt as output event signal.

**\_\_STATIC\_FORCEINLINE void \_\_enable\_mcycle\_counter(void)**  
Enable MCYCLE counter.  
Clear the CY bit of MCOUNTINHIBIT to 0 to enable MCYCLE Counter

**\_\_STATIC\_FORCEINLINE void \_\_disable\_mcycle\_counter(void)**  
Disable MCYCLE counter.  
Set the CY bit of MCOUNTINHIBIT to 1 to disable MCYCLE Counter

**\_\_STATIC\_FORCEINLINE void \_\_enable\_minstret\_counter(void)**  
Enable MINSTRET counter.  
Clear the IR bit of MCOUNTINHIBIT to 0 to enable MINSTRET Counter

**\_\_STATIC\_FORCEINLINE void \_\_disable\_minstret\_counter(void)**  
Disable MINSTRET counter.  
Set the IR bit of MCOUNTINHIBIT to 1 to disable MINSTRET Counter

**\_\_STATIC\_FORCEINLINE void \_\_enable\_all\_counter(void)**  
Enable MCYCLE & MINSTRET counter.  
Clear the IR and CY bit of MCOUNTINHIBIT to 1 to enable MINSTRET & MCYCLE Counter

**\_\_STATIC\_FORCEINLINE void \_\_disable\_all\_counter(void)**  
Disable MCYCLE & MINSTRET counter.  
Set the IR and CY bit of MCOUNTINHIBIT to 1 to disable MINSTRET & MCYCLE Counter

**\_\_STATIC\_FORCEINLINE void \_\_FENCE\_I(void)**  
Fence.i Instruction.  
The FENCE.I instruction is used to synchronize the instruction and data streams.

**\_\_STATIC\_FORCEINLINE uint8\_t \_\_LB(volatile void \* addr)**  
Load 8bit value from address (8 bit)  
Load 8 bit value.  
**Return** value of type uint8\_t at (\*addr)  
**Parameters**

- [in] addr: Address pointer to data

**\_\_STATIC\_FORCEINLINE uint16\_t \_\_LH(volatile void \* addr)**  
Load 16bit value from address (16 bit)  
Load 16 bit value.  
**Return** value of type uint16\_t at (\*addr)  
**Parameters**

- [in] addr: Address pointer to data

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_LW(volatile void \* addr)**  
Load 32bit value from address (32 bit)  
Load 32 bit value.  
**Return** value of type uint32\_t at (\*addr)  
**Parameters**

- [in] addr: Address pointer to data

**\_\_STATIC\_FORCEINLINE void \_\_SB(volatile void \* addr, uint8\_t val)**

Write 8bit value to address (8 bit)

Write 8 bit value.

**Parameters**

- [in] addr: Address pointer to data
- [in] val: Value to set

**\_\_STATIC\_FORCEINLINE void \_\_SH(volatile void \* addr, uint16\_t val)**

Write 16bit value to address (16 bit)

Write 16 bit value.

**Parameters**

- [in] addr: Address pointer to data
- [in] val: Value to set

**\_\_STATIC\_FORCEINLINE void \_\_SW(volatile void \* addr, uint32\_t val)**

Write 32bit value to address (32 bit)

Write 32 bit value.

**Parameters**

- [in] addr: Address pointer to data
- [in] val: Value to set

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_CAS\_W(volatile uint32\_t \* addr, uint32\_t oldval, uint32\_t newval)**

Compare and Swap 32bit value using LR and SC.

Compare old value with memory, if identical, store new value in memory. Return the initial value in memory. Success is indicated by comparing return value with OLD. memory address, return 0 if successful, otherwise return !0

**Return** return the initial value in memory

**Parameters**

- [in] addr: Address pointer to data, address need to be 4byte aligned
- [in] oldval: Old value of the data in address
- [in] newval: New value to be stored into the address

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_AMOSWAP\_W(volatile uint32\_t \* addr, uint32\_t newval)**

Atomic Swap 32bit value into memory.

Atomically swap new 32bit value into memory using amoswap.d.

**Return** return the original value in memory

**Parameters**

- [in] addr: Address pointer to data, address need to be 4byte aligned
- [in] newval: New value to be stored into the address

**\_\_STATIC\_FORCEINLINE int32\_t \_\_AMOADD\_W(volatile int32\_t \* addr, int32\_t value)**

Atomic Add with 32bit value.

Atomically ADD 32bit value with value in memory using amoadd.d.

**Return** return memory value + add value

**Parameters**

- [in] addr: Address pointer to data, address need to be 4byte aligned
- [in] value: value to be ADDED

**\_\_STATIC\_FORCEINLINE int32\_t \_\_AMOAND\_W(volatile int32\_t \* addr, int32\_t value)**  
Atomic And with 32bit value.

Atomically AND 32bit value with value in memory using amoand.d.

**Return** return memory value & and value

**Parameters**

- [in] addr: Address pointer to data, address need to be 4byte aligned
- [in] value: value to be ANDed

**\_\_STATIC\_FORCEINLINE int32\_t \_\_AMOOR\_W(volatile int32\_t \* addr, int32\_t value)**  
Atomic OR with 32bit value.

Atomically OR 32bit value with value in memory using amoor.d.

**Return** return memory value | and value

**Parameters**

- [in] addr: Address pointer to data, address need to be 4byte aligned
- [in] value: value to be ORed

**\_\_STATIC\_FORCEINLINE int32\_t \_\_AMOXOR\_W(volatile int32\_t \* addr, int32\_t value)**  
Atomic XOR with 32bit value.

Atomically XOR 32bit value with value in memory using amoxor.d.

**Return** return memory value ^ and value

**Parameters**

- [in] addr: Address pointer to data, address need to be 4byte aligned
- [in] value: value to be XORed

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_AMOMAXU\_W(volatile uint32\_t \* addr, uint32\_t value)**  
Atomic unsigned MAX with 32bit value.

Atomically unsigned max compare 32bit value with value in memory using amomaxu.d.

**Return** return the bigger value

**Parameters**

- [in] addr: Address pointer to data, address need to be 4byte aligned
- [in] value: value to be compared

**\_\_STATIC\_FORCEINLINE int32\_t \_\_AMOMAX\_W(volatile int32\_t \* addr, int32\_t value)**  
Atomic signed MAX with 32bit value.

Atomically signed max compare 32bit value with value in memory using amomax.d.

**Return** the bigger value

**Parameters**

- [in] addr: Address pointer to data, address need to be 4byte aligned
- [in] value: value to be compared



`__STATIC_FORCEINLINE uint32_t __AMOMINU_W(volatile uint32_t * addr, uint32_t value)`  
Atomic unsigned MIN with 32bit value.

Atomically unsigned min compare 32bit value with value in memory using amominu.d.

**Return** the smaller value

**Parameters**

- [in] `addr`: Address pointer to data, address need to be 4byte aligned
- [in] `value`: value to be compared

`__STATIC_FORCEINLINE int32_t __AMOMIN_W(volatile int32_t * addr, int32_t value)`  
Atomic signed MIN with 32bit value.

Atomically signed min compare 32bit value with value in memory using amomin.d.

**Return** the smaller value

**Parameters**

- [in] `addr`: Address pointer to data, address need to be 4byte aligned
- [in] `value`: value to be compared

## 2.5.7 Intrinsic Functions for SIMD Instructions

Click [Nuclei DSP Feature<sup>16</sup>](#) to learn about Core DSP in Nuclei ISA Spec.

### SIMD Data Processing Instructions

#### SIMD 16-bit Add/Subtract Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_ADD16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_CRAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_CRSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KADD16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KCRAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KCRSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KSTAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KSTSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KSUB16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RADD16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RCRAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RCRSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RSTAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RSTSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RSUB16(unsigned long a, unsigned long b)
```

<sup>16</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/dsp.html](https://doc.nucleisys.com/nuclei_spec/isa/dsp.html)

```

__STATIC_FORCEINLINE unsigned long __RV_STAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_STSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SUB16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKADD16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKCRAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKCRSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKSTAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKSTSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKSUB16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URADD16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URCRAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URCRSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URSTAS16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URSTSA16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URSUB16(unsigned long a, unsigned long b)

```

group **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_16B\_ADDSUB**

SIMD 16-bit Add/Subtract Instructions.

Based on the combination of the types of the two 16-bit arithmetic operations, the SIMD 16-bit add/subtract instructions can be classified into 6 main categories: Addition (two 16-bit addition), Subtraction (two 16-bit subtraction), Crossed Add & Sub (one addition and one subtraction), and Crossed Sub & Add (one subtraction and one addition), Straight Add & Sub (one addition and one subtraction), and Straight Sub & Add (one subtraction and one addition). Based on the way of how an overflow condition is handled, the SIMD 16-bit add/subtract instructions can be classified into 5 groups: Wrap-around (dropping overflow), Signed Halving (keeping overflow by dropping 1 LSB bit), Unsigned Halving, Signed Saturation (clipping overflow), and Unsigned Saturation. Together, there are 30 SIMD 16-bit add/subtract instructions.

## Functions

```

__STATIC_FORCEINLINE unsigned long __RV_ADD16(unsigned long a, unsigned long b)
ADD16 (SIMD 16-bit Addition)

```

**Type:** SIMD

**Syntax:**

ADD16 Rd, Rs1, Rs2
--------------------

**Purpose:** Do 16-bit integer element additions simultaneously.

**Description:** This instruction adds the 16-bit integer elements in Rs1 with the 16-bit integer elements in Rs2, and then writes the 16-bit element results to Rd.

**Note:** This instruction can be used for either signed or unsigned addition.

**Operations:**

```
Rd.H[x] = Rs1.H[x] + Rs2.H[x];
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CRAS16(unsigned long a, unsigned long b)**  
CRAS16 (SIMD 16-bit Cross Addition & Subtraction)

**Type:** SIMD

**Syntax:**

```
CRAS16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit integer element addition and 16-bit integer element subtraction in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks.

**Description:** This instruction adds the 16-bit integer element in [31:16] of 32-bit chunks in Rs1 with the 16-bit integer element in [15:0] of 32-bit chunks in Rs2, and writes the result to [31:16] of 32-bit chunks in Rd; at the same time, it subtracts the 16-bit integer element in [31:16] of 32-bit chunks in Rs2 from the 16-bit integer element in [15:0] of 32-bit chunks, and writes the result to [15:0] of 32-bit chunks in Rd.

**Note:** This instruction can be used for either signed or unsigned operations.

**Operations:**

```
Rd.W[x][31:16] = Rs1.W[x][31:16] + Rs2.W[x][15:0];
Rd.W[x][15:0] = Rs1.W[x][15:0] - Rs2.W[x][31:16];
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CRSA16(unsigned long a, unsigned long b)**  
CRSA16 (SIMD 16-bit Cross Subtraction & Addition)

**Type:** SIMD

**Syntax:**

```
CRSA16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit integer element subtraction and 16-bit integer element addition in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks.

**Description:** This instruction subtracts the 16-bit integer element in [15:0] of 32-bit chunks in Rs2 from the 16-bit integer element in [31:16] of 32-bit chunks in Rs1, and writes the result to [31:16] of 32-bit chunks in Rd; at the same time, it adds the 16-bit integer element in [31:16] of 32-bit chunks in Rs2 with

the 16-bit integer element in [15:0] of 32-bit chunks in Rs1, and writes the result to [15:0] of 32-bit chunks in Rd.

**Note:** This instruction can be used for either signed or unsigned operations.

**Operations:**

```
Rd.W[x][31:16] = Rs1.W[x][31:16] - Rs2.W[x][15:0];
Rd.W[x][15:0] = Rs1.W[x][15:0] + Rs2.W[x][31:16];
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KADD16(unsigned long a, unsigned long b)**  
KADD16 (SIMD 16-bit Signed Saturating Addition)

**Type:** SIMD

**Syntax:**

```
KADD16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element saturating additions simultaneously.

**Description:** This instruction adds the 16-bit signed integer elements in Rs1 with the 16-bit signed integer elements in Rs2. If any of the results are beyond the Q15 number range ( $-2^{15} \leq Q15 \leq 2^{15}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```
res[x] = Rs1.H[x] + Rs2.H[x];
if (res[x] > 32767) {
    res[x] = 32767;
    OV = 1;
} else if (res[x] < -32768) {
    res[x] = -32768;
    OV = 1;
}
Rd.H[x] = res[x];
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KCRAS16(unsigned long a, unsigned long b)**  
KCRAS16 (SIMD 16-bit Signed Saturating Cross Addition & Subtraction)

**Type:** SIMD

**Syntax:**

```
KCRAS16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element saturating addition and 16-bit signed integer element saturating subtraction in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks.

**Description:** This instruction adds the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs1 with the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs2; at the same time, it subtracts the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs2 from the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs1. If any of the results are beyond the Q15 number range ( $-2^{15} \leq Q15 \leq 2^{15}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [31:16] of 32-bit chunks in Rd for addition and [15:0] of 32-bit chunks in Rd for subtraction.

**Operations:**

```
res1 = Rs1.W[x][31:16] + Rs2.W[x][15:0];
res2 = Rs1.W[x][15:0] - Rs2.W[x][31:16];
for (res in [res1, res2]) {
    if (res > (2^15)-1) {
        res = (2^15)-1;
        OV = 1;
    } else if (res < -2^15) {
        res = -2^15;
        OV = 1;
    }
}
Rd.W[x][31:16] = res1;
Rd.W[x][15:0] = res2;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE unsigned long __RV_KCRSA16(unsigned long a, unsigned long b)
KCRSA16 (SIMD 16-bit Signed Saturating Cross Subtraction & Addition)
```

**Type:** SIMD

**Syntax:**

```
KCRSA16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element saturating subtraction and 16-bit signed integer element saturating addition in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks.

**Description:** This instruction subtracts the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs2 from the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs1; at the same time, it adds the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs2 with the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs1. If any of the results are beyond the Q15 number range ( $-2^{15} \leq Q15 \leq 2^{15}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [31:16] of 32-bit chunks in Rd for subtraction and [15:0] of 32-bit chunks in Rd for addition.

**Operations:**

```

res1 = Rs1.W[x][31:16] - Rs2.W[x][15:0];
res2 = Rs1.W[x][15:0] + Rs2.W[x][31:16];
for (res in [res1, res2]) {
    if (res > (2^15)-1) {
        res = (2^15)-1;
        OV = 1;
    } else if (res < -2^15) {
        res = -2^15;
        OV = 1;
    }
}
Rd.W[x][31:16] = res1;
Rd.W[x][15:0] = res2;
for RV32, x=0
for RV64, x=1...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSTAS16(unsigned long a, unsigned long b)**  
KSTAS16 (SIMD 16-bit Signed Saturating Straight Addition & Subtraction)

**Type:** SIMD

#### Syntax:

```
KSTAS16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element saturating addition and 16-bit signed integer element saturating subtraction in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks.

**Description:** This instruction adds the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs1 with the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs2; at the same time, it subtracts the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs2 from the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs1. If any of the results are beyond the Q15 number range ( $-2^{15} \leq Q15 \leq 2^{15}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [31:16] of 32-bit chunks in Rd for addition and [15:0] of 32-bit chunks in Rd for subtraction.

#### Operations:

```

res1 = Rs1.W[x][31:16] + Rs2.W[x][31:16];
res2 = Rs1.W[x][15:0] - Rs2.W[x][15:0];
for (res in [res1, res2]) {
    if (res > (2^15)-1) {
        res = (2^15)-1;
        OV = 1;
    } else if (res < -2^15) {
        res = -2^15;
        OV = 1;
    }
}
Rd.W[x][31:16] = res1;
Rd.W[x][15:0] = res2;

```

(continues on next page)

(continued from previous page)

```

for RV32, x=0
for RV64, x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KTSA16(unsigned long a, unsigned long b)**  
 K TSA16 (SIMD 16-bit Signed Saturating Straight Subtraction & Addition)

**Type:** SIMD

**Syntax:**

```
K TSA16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element saturating subtraction and 16-bit signed integer element saturating addition in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks.

**Description:** This instruction subtracts the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs2 from the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs1; at the same time, it adds the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs2 with the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs1. If any of the results are beyond the Q15 number range ( $-2^{15} \leq Q15 \leq 2^{15}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [31:16] of 32-bit chunks in Rd for subtraction and [15:0] of 32-bit chunks in Rd for addition.

**Operations:**

```

res1 = Rs1.W[x][31:16] - Rs2.W[x][31:16];
res2 = Rs1.W[x][15:0] + Rs2.W[x][15:0];
for (res in [res1, res2]) {
    if (res > (2^15)-1) {
        res = (2^15)-1;
        OV = 1;
    } else if (res < -2^15) {
        res = -2^15;
        OV = 1;
    }
}
Rd.W[x][31:16] = res1;
Rd.W[x][15:0] = res2;
for RV32, x=0
for RV64, x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSUB16(unsigned long a, unsigned long b)**  
 K SUB16 (SIMD 16-bit Signed Saturating Subtraction)

**Type:** SIMD**Syntax:**

KSUB16 Rd, Rs1, Rs2

**Purpose:** Do 16-bit signed integer elements saturating subtractions simultaneously.**Description:** This instruction subtracts the 16-bit signed integer elements in Rs2 from the 16-bit signed integer elements in Rs1. If any of the results are beyond the Q15 number range ( $-2^{15} \leq Q15 \leq 2^{15}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.**Operations:**

```

res[x] = Rs1.H[x] - Rs2.H[x];
if (res[x] > (2^15)-1) {
    res[x] = (2^15)-1;
    OV = 1;
} else if (res[x] < -2^15) {
    res[x] = -2^15;
    OV = 1;
}
Rd.H[x] = res[x];
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RADD16(unsigned long a, unsigned long b)**  
 RADD16 (SIMD 16-bit Signed Halving Addition)

**Type:** SIMD**Syntax:**

RADD16 Rd, Rs1, Rs2

**Purpose:** Do 16-bit signed integer element additions simultaneously. The results are halved to avoid overflow or saturation.**Description:** This instruction adds the 16-bit signed integer elements in Rs1 with the 16-bit signed integer elements in Rs2. The results are first arithmetically right-shifted by 1 bit and then written to Rd.**Examples:**

```

* Rs1 = 0x7FFF, Rs2 = 0x7FFF, Rd = 0x7FFF
* Rs1 = 0x8000, Rs2 = 0x8000, Rd = 0x8000
* Rs1 = 0x4000, Rs2 = 0x8000, Rd = 0xE000

```

**Operations:**

```

Rd.H[x] = (Rs1.H[x] + Rs2.H[x]) s>> 1; for RV32: x=1...0, for RV64: x=3...0

```

**Return** value stored in unsigned long type



**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RCRAS16(unsigned long a, unsigned long b)**  
RCRAS16 (SIMD 16-bit Signed Halving Cross Addition & Subtraction)

**Type:** SIMD

**Syntax:**

```
RCRAS16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element addition and 16-bit signed integer element subtraction in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs1 with the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs2, and subtracts the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs2 from the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs1. The element results are first arithmetically right-shifted by 1 bit and then written to [31:16] of 32-bit chunks in Rd and [15:0] of 32-bit chunks in Rd.

**Examples:**

```
Please see `RADD16` and `RSUB16` instructions.
```

**Operations:**

```
Rd.W[x][31:16] = (Rs1.W[x][31:16] + Rs2.W[x][15:0]) s>> 1;
Rd.W[x][15:0] = (Rs1.W[x][15:0] - Rs2.W[x][31:16]) s>> 1;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RCRSA16(unsigned long a, unsigned long b)**  
RCRSA16 (SIMD 16-bit Signed Halving Cross Subtraction & Addition)

**Type:** SIMD

**Syntax:**

```
RCRSA16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element subtraction and 16-bit signed integer element addition in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs2 from the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs1, and adds the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs1 with the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs2. The two results are first arithmetically right-shifted by 1 bit and then written to [31:16] of 32-bit chunks in Rd and [15:0] of 32-bit chunks in Rd.

**Examples:**

Please see `RADD16` and `RSUB16` instructions.

**Operations:**

```
Rd.W[x][31:16] = (Rs1.W[x][31:16] - Rs2.W[x][15:0]) s>> 1;
Rd.W[x][15:0] = (Rs1.W[x][15:0] + Rs2.W[x][31:16]) s>> 1;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RSTAS16(unsigned long a, unsigned long b)**  
RSTAS16 (SIMD 16-bit Signed Halving Straight Addition & Subtraction)

**Type:** SIMD

**Syntax:**

RSTAS16 Rd, Rs1, Rs2

**Purpose:** Do 16-bit signed integer element addition and 16-bit signed integer element subtraction in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs1 with the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs2, and subtracts the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs2 from the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs1. The element results are first arithmetically right-shifted by 1 bit and then written to [31:16] of 32-bit chunks in Rd and [15:0] of 32-bit chunks in Rd.

**Examples:**

Please see `RADD16` and `RSUB16` instructions.

**Operations:**

```
Rd.W[x][31:16] = (Rs1.W[x][31:16] + Rs2.W[x][31:16]) s>> 1;
Rd.W[x][15:0] = (Rs1.W[x][15:0] - Rs2.W[x][15:0]) s>> 1;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RSTSA16(unsigned long a, unsigned long b)**  
RSTSA16 (SIMD 16-bit Signed Halving Straight Subtraction & Addition)

**Type:** SIMD

**Syntax:**

```
RSTSA16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element subtraction and 16-bit signed integer element addition in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs2 from the 16-bit signed integer element in [31:16] of 32-bit chunks in Rs1, and adds the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs1 with the 16-bit signed integer element in [15:0] of 32-bit chunks in Rs2. The two results are first arithmetically right-shifted by 1 bit and then written to [31:16] of 32-bit chunks in Rd and [15:0] of 32-bit chunks in Rd.

**Examples:**

```
Please see `RADD16` and `RSUB16` instructions.
```

**Operations:**

```
Rd.W[x][31:16] = (Rs1.W[x][31:16] - Rs2.W[x][31:16]) s>> 1;
Rd.W[x][15:0] = (Rs1.W[x][15:0] + Rs2.W[x][15:0]) s>> 1;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE unsigned long __RV_RSUB16(unsigned long a, unsigned long b)
RSUB16 (SIMD 16-bit Signed Halving Subtraction)
```

**Type:** SIMD

**Syntax:**

```
RSUB16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer element subtractions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 16-bit signed integer elements in Rs2 from the 16-bit signed integer elements in Rs1. The results are first arithmetically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Ra = 0x7FFF, Rb = 0x8000, Rt = 0x7FFF
* Ra = 0x8000, Rb = 0x7FFF, Rt = 0x8000
* Ra = 0x8000, Rb = 0x4000, Rt = 0xA000
```

**Operations:**

```
Rd.H[x] = (Rs1.H[x] - Rs2.H[x]) s>> 1;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_STAS16(unsigned long a, unsigned long b)**  
STAS16 (SIMD 16-bit Straight Addition & Subtraction)

**Type:** SIMD

**Syntax:**

```
STAS16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit integer element addition and 16-bit integer element subtraction in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks.

**Description:** This instruction adds the 16-bit integer element in [31:16] of 32-bit chunks in Rs1 with the 16-bit integer element in [31:16] of 32-bit chunks in Rs2, and writes the result to [31:16] of 32-bit chunks in Rd; at the same time, it subtracts the 16-bit integer element in [15:0] of 32-bit chunks in Rs2 from the 16-bit integer element in [15:0] of 32-bit chunks, and writes the result to [15:0] of 32-bit chunks in Rd.

**Note:** This instruction can be used for either signed or unsigned operations.

**Operations:**

```
Rd.W[x][31:16] = Rs1.W[x][31:16] + Rs2.W[x][31:16];
Rd.W[x][15:0] = Rs1.W[x][15:0] - Rs2.W[x][15:0];
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_STSA16(unsigned long a, unsigned long b)**  
STSA16 (SIMD 16-bit Straight Subtraction & Addition)

**Type:** SIMD

**Syntax:**

```
STSA16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit integer element subtraction and 16-bit integer element addition in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks.

**Description:** This instruction subtracts the 16-bit integer element in [31:16] of 32-bit chunks in Rs2 from the 16-bit integer element in [31:16] of 32-bit chunks in Rs1, and writes the result to [31:16] of 32-bit chunks in Rd; at the same time, it adds the 16-bit integer element in [15:0] of 32-bit chunks in Rs2 with the 16-bit integer element in [15:0] of 32-bit chunks in Rs1, and writes the result to [15:0] of 32-bit chunks in Rd.

**Note:** This instruction can be used for either signed or unsigned operations.

**Operations:**

```

Rd.W[x][31:16] = Rs1.W[x][31:16] - Rs2.W[x][31:16];
Rd.W[x][15:0] = Rs1.W[x][15:0] + Rs2.W[x][15:0];
for RV32, x=0
for RV64, x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SUB16(unsigned long a, unsigned long b)**  
SUB16 (SIMD 16-bit Subtraction)

**Type:** SIMD

**Syntax:**

```
SUB16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit integer element subtractions simultaneously.

**Description:** This instruction subtracts the 16-bit integer elements in Rs2 from the 16-bit integer elements in Rs1, and then writes the result to Rd.

**Note:** This instruction can be used for either signed or unsigned subtraction.

**Operations:**

```

Rd.H[x] = Rs1.H[x] - Rs2.H[x];
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKADD16(unsigned long a, unsigned long b)**  
UKADD16 (SIMD 16-bit Unsigned Saturating Addition)

**Type:** SIMD

**Syntax:**

```
UKADD16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer element saturating additions simultaneously.

**Description:** This instruction adds the 16-bit unsigned integer elements in Rs1 with the 16-bit unsigned integer elements in Rs2. If any of the results are beyond the 16-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{16}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```

res[x] = Rs1.H[x] + Rs2.H[x];
if (res[x] > (2^16)-1) {
    res[x] = (2^16)-1;
    OV = 1;
}
Rd.H[x] = res[x];
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKCRAS16(unsigned long a, unsigned long b)**  
 UKCRAS16 (SIMD 16-bit Unsigned Saturating Cross Addition & Subtraction)

**Type:** SIMD

**Syntax:**

```
UKCRAS16 Rd, Rs1, Rs2
```

**Purpose:** Do one 16-bit unsigned integer element saturating addition and one 16-bit unsigned integer element saturating subtraction in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks.

**Description:** This instruction adds the 16-bit unsigned integer element in [31:16] of 32-bit chunks in Rs1 with the 16-bit unsigned integer element in [15:0] of 32-bit chunks in Rs2; at the same time, it subtracts the 16-bit unsigned integer element in [31:16] of 32-bit chunks in Rs2 from the 16-bit unsigned integer element in [15:0] of 32-bit chunks in Rs1. If any of the results are beyond the 16-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{16}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [31:16] of 32-bit chunks in Rd for addition and [15:0] of 32-bit chunks in Rd for subtraction.

**Operations:**

```

res1 = Rs1.W[x][31:16] + Rs2.W[x][15:0];
res2 = Rs1.W[x][15:0] - Rs2.W[x][31:16];
if (res1 > (2^16)-1) {
    res1 = (2^16)-1;
    OV = 1;
}
if (res2 < 0) {
    res2 = 0;
    OV = 1;
}
Rd.W[x][31:16] = res1;
Rd.W[x][15:0] = res2;
for RV32, x=0
for RV64, x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKCRSA16(unsigned long a, unsigned long b)**  
UKCRSA16 (SIMD 16-bit Unsigned Saturating Cross Subtraction & Addition)

**Type:** SIMD

**Syntax:**

```
UKCRSA16 Rd, Rs1, Rs2
```

**Purpose:** Do one 16-bit unsigned integer element saturating subtraction and one 16-bit unsigned integer element saturating addition in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks.

**Description:** This instruction subtracts the 16-bit unsigned integer element in [15:0] of 32-bit chunks in Rs2 from the 16-bit unsigned integer element in [31:16] of 32-bit chunks in Rs1; at the same time, it adds the 16-bit unsigned integer element in [31:16] of 32-bit chunks in Rs2 with the 16-bit unsigned integer element in [15:0] of 32-bit chunks in Rs1. If any of the results are beyond the 16-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{16}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [31:16] of 32-bit chunks in Rd for subtraction and [15:0] of 32-bit chunks in Rd for addition.

**Operations:**

```
res1 = Rs1.W[x][31:16] - Rs2.W[x][15:0];
res2 = Rs1.W[x][15:0] + Rs2.W[x][31:16];
if (res1 < 0) {
    res1 = 0;
    OV = 1;
} else if (res2 > (2^16)-1) {
    res2 = (2^16)-1;
    OV = 1;
}
Rd.W[x][31:16] = res1;
Rd.W[x][15:0] = res2;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSTAS16(unsigned long a, unsigned long b)**  
UKSTAS16 (SIMD 16-bit Unsigned Saturating Straight Addition & Subtraction)

**Type:** SIMD

**Syntax:**

```
UKSTAS16 Rd, Rs1, Rs2
```

**Purpose:** Do one 16-bit unsigned integer element saturating addition and one 16-bit unsigned integer element saturating subtraction in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks.

**Description:** This instruction adds the 16-bit unsigned integer element in [31:16] of 32-bit chunks in Rs1 with the 16-bit unsigned integer element in [31:16] of 32-bit chunks in Rs2; at the same time, it subtracts the 16-bit unsigned integer element in [15:0] of 32-bit chunks in Rs2 from the 16-bit unsigned integer element in [15:0] of 32-bit chunks in Rs1. If any of the results are beyond the 16-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{16}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [31:16] of 32-bit chunks in Rd for addition and [15:0] of 32-bit chunks in Rd for subtraction.

**Operations:**

```
res1 = Rs1.W[x][31:16] + Rs2.W[x][31:16];
res2 = Rs1.W[x][15:0] - Rs2.W[x][15:0];
if (res1 > (2^16)-1) {
    res1 = (2^16)-1;
    OV = 1;
}
if (res2 < 0) {
    res2 = 0;
    OV = 1;
}
Rd.W[x][31:16] = res1;
Rd.W[x][15:0] = res2;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSTSA16(unsigned long a, unsigned long b)**  
UKSTSA16 (SIMD 16-bit Unsigned Saturating Straight Subtraction & Addition)

**Type:** SIMD

**Syntax:**

```
UKSTSA16 Rd, Rs1, Rs2
```

**Purpose:** Do one 16-bit unsigned integer element saturating subtraction and one 16-bit unsigned integer element saturating addition in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks.

**Description:** This instruction subtracts the 16-bit unsigned integer element in [31:16] of 32-bit chunks in Rs2 from the 16-bit unsigned integer element in [31:16] of 32-bit chunks in Rs1; at the same time, it adds the 16-bit unsigned integer element in [15:0] of 32-bit chunks in Rs2 with the 16-bit unsigned integer element in [15:0] of 32-bit chunks in Rs1. If any of the results are beyond the 16-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{16}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [31:16] of 32-bit chunks in Rd for subtraction and [15:0] of 32-bit chunks in Rd for addition.

**Operations:**

```
res1 = Rs1.W[x][31:16] - Rs2.W[x][31:16];
res2 = Rs1.W[x][15:0] + Rs2.W[x][15:0];
if (res1 < 0) {
    res1 = 0;
}
```

(continues on next page)



(continued from previous page)

```

    OV = 1;
} else if (res2 > (2^16)-1) {
    res2 = (2^16)-1;
    OV = 1;
}
Rd.W[x][31:16] = res1;
Rd.W[x][15:0] = res2;
for RV32, x=0
for RV64, x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSUB16(unsigned long a, unsigned long b)**  
 UKSUB16 (SIMD 16-bit Unsigned Saturating Subtraction)

**Type:** SIMD

**Syntax:**

```
UKSUB16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer elements saturating subtractions simultaneously.

**Description:** This instruction subtracts the 16-bit unsigned integer elements in Rs2 from the 16-bit unsigned integer elements in Rs1. If any of the results are beyond the 16-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{16}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```

res[x] = Rs1.H[x] - Rs2.H[x];
if (res[x] < 0) {
    res[x] = 0;
    OV = 1;
}
Rd.H[x] = res[x];
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URADD16(unsigned long a, unsigned long b)**  
 URADD16 (SIMD 16-bit Unsigned Halving Addition)

**Type:** SIMD

**Syntax:**

```
URADD16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer element additions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 16-bit unsigned integer elements in Rs1 with the 16-bit unsigned integer elements in Rs2. The results are first logically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Ra = 0x7FFF, Rb = 0x7FFF Rt = 0x7FFF
* Ra = 0x8000, Rb = 0x8000 Rt = 0x8000
* Ra = 0x4000, Rb = 0x8000 Rt = 0x6000
```

**Operations:**

```
Rd.H[x] = (Rs1.H[x] + Rs2.H[x]) u>> 1;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE unsigned long __RV_URCRAS16(unsigned long a, unsigned long b)
    URCRAS16 (SIMD 16-bit Unsigned Halving Cross Addition & Subtraction)
```

**Type:** SIMD

**Syntax:**

```
URCRAS16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer element addition and 16-bit unsigned integer element subtraction in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 16-bit unsigned integer in [31:16] of 32-bit chunks in Rs1 with the 16-bit unsigned integer in [15:0] of 32-bit chunks in Rs2, and subtracts the 16-bit unsigned integer in [31:16] of 32-bit chunks in Rs2 from the 16-bit unsigned integer in [15:0] of 32-bit chunks in Rs1. The element results are first logically right-shifted by 1 bit and then written to [31:16] of 32-bit chunks in Rd and [15:0] of 32-bit chunks in Rd.

**Examples:**

```
Please see `URADD16` and `URSUB16` instructions.
```

**Operations:**

```
Rd.W[x][31:16] = (Rs1.W[x][31:16] + Rs2.W[x][15:0]) u>> 1;
Rd.W[x][15:0] = (Rs1.W[x][15:0] - Rs2.W[x][31:16]) u>> 1;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URCRSA16(unsigned long a, unsigned long b)**  
URCRSA16 (SIMD 16-bit Unsigned Halving Cross Subtraction & Addition)

**Type:** SIMD

**Syntax:**

```
URCRSA16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer element subtraction and 16-bit unsigned integer element addition in a 32-bit chunk simultaneously. Operands are from crossed positions in 32-bit chunks. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 16-bit unsigned integer in [15:0] of 32-bit chunks in Rs2 from the 16-bit unsigned integer in [31:16] of 32-bit chunks in Rs1, and adds the 16-bit unsigned integer in [15:0] of 32-bit chunks in Rs1 with the 16-bit unsigned integer in [31:16] of 32-bit chunks in Rs2. The two results are first logically right-shifted by 1 bit and then written to [31:16] of 32-bit chunks in Rd and [15:0] of 32-bit chunks in Rd.

**Examples:**

```
Please see `URADD16` and `URSUB16` instructions.
```

**Operations:**

```
Rd.W[x][31:16] = (Rs1.W[x][31:16] - Rs2.W[x][15:0]) u>> 1;
Rd.W[x][15:0] = (Rs1.W[x][15:0] + Rs2.W[x][31:16]) u>> 1;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URSTAS16(unsigned long a, unsigned long b)**  
URSTAS16 (SIMD 16-bit Unsigned Halving Straight Addition & Subtraction)

**Type:** SIMD

**Syntax:**

```
URSTAS16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer element addition and 16-bit unsigned integer element subtraction in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 16-bit unsigned integer in [31:16] of 32-bit chunks in Rs1 with the 16-bit unsigned integer in [31:16] of 32-bit chunks in Rs2, and subtracts the 16-bit unsigned integer in [15:0] of 32-bit chunks in Rs2 from the 16-bit unsigned integer in [15:0] of 32-bit chunks in Rs1. The element results are first logically right-shifted by 1 bit and then written to [31:16] of 32-bit chunks in Rd and [15:0] of 32-bit chunks in Rd.

**Examples:**

Please see `URADD16` and `URSUB16` instructions.

**Operations:**

```
Rd.W[x][31:16] = (Rs1.W[x][31:16] + Rs2.W[x][31:16]) u>> 1;
Rd.W[x][15:0] = (Rs1.W[x][15:0] - Rs2.W[x][15:0]) u>> 1;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URSTSA16(unsigned long a, unsigned long b)**  
URSTSA16 (SIMD 16-bit Unsigned Halving Straight Subtraction & Addition)

**Type:** SIMD

**Syntax:**

URCRSA16 Rd, Rs1, Rs2

**Purpose:** Do 16-bit unsigned integer element subtraction and 16-bit unsigned integer element addition in a 32-bit chunk simultaneously. Operands are from corresponding positions in 32-bit chunks. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 16-bit unsigned integer in [31:16] of 32-bit chunks in Rs2 from the 16-bit unsigned integer in [31:16] of 32-bit chunks in Rs1, and adds the 16-bit unsigned integer in [15:0] of 32-bit chunks in Rs1 with the 16-bit unsigned integer in [15:0] of 32-bit chunks in Rs2. The two results are first logically right-shifted by 1 bit and then written to [31:16] of 32-bit chunks in Rd and [15:0] of 32-bit chunks in Rd.

**Examples:**

Please see `URADD16` and `URSUB16` instructions.

**Operations:**

```
Rd.W[x][31:16] = (Rs1.W[x][31:16] - Rs2.W[x][31:16]) u>> 1;
Rd.W[x][15:0] = (Rs1.W[x][15:0] + Rs2.W[x][15:0]) u>> 1;
for RV32, x=0
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URSUB16(unsigned long a, unsigned long b)**  
URSUB16 (SIMD 16-bit Unsigned Halving Subtraction)

**Type:** SIMD

**Syntax:**

```
URSUB16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer element subtractions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 16-bit unsigned integer elements in Rs2 from the 16-bit unsigned integer elements in Rs1. The results are first logically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Ra = 0x7FFF, Rb = 0x8000 Rt = 0xFFFF
* Ra = 0x8000, Rb = 0x7FFF Rt = 0x0000
* Ra = 0x8000, Rb = 0x4000 Rt = 0x2000
```

**Operations:**

```
Rd.H[x] = (Rs1.H[x] - Rs2.H[x]) u>> 1;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**SIMD 8-bit Addition & Subtraction Instructions**

```
__STATIC_FORCEINLINE unsigned long __RV_ADD8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KADD8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KSUB8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RADD8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RSUB8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SUB8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKADD8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKSUB8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URADD8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URSUB8(unsigned long a, unsigned long b)
```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_8B\_ADDSUB**

SIMD 8-bit Addition & Subtraction Instructions.

Based on the types of the four 8-bit arithmetic operations, the SIMD 8-bit add/subtract instructions can be classified into 2 main categories: Addition (four 8-bit addition), and Subtraction (four 8-bit subtraction). Based on the way of how an overflow condition is handled for signed or unsigned operation, the SIMD 8-bit add/subtract instructions can be classified into 5 groups: Wrap-around (dropping overflow), Signed Halving (keeping overflow by dropping 1 LSB bit), Unsigned Halving, Signed Saturation (clipping overflow), and Unsigned Saturation. Together, there are 10 SIMD 8-bit add/subtract instructions.

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_ADD8(unsigned long a, unsigned long b)**  
ADD8 (SIMD 8-bit Addition)

**Type:** SIMD

**Syntax:**

```
ADD8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit integer element additions simultaneously.

**Description:** This instruction adds the 8-bit integer elements in Rs1 with the 8-bit integer elements in Rs2, and then writes the 8-bit element results to Rd.

**Note:** This instruction can be used for either signed or unsigned addition.

**Operations:**

```
Rd.B[x] = Rs1.B[x] + Rs2.B[x];
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KADD8(unsigned long a, unsigned long b)**  
KADD8 (SIMD 8-bit Signed Saturating Addition)

**Type:** SIMD

**Syntax:**

```
KADD8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit signed integer element saturating additions simultaneously.

**Description:** This instruction adds the 8-bit signed integer elements in Rs1 with the 8-bit signed integer elements in Rs2. If any of the results are beyond the Q7 number range ( $-2^7 \leq Q7 \leq 2^7-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```
res[x] = Rs1.B[x] + Rs2.B[x];
if (res[x] > 127) {
    res[x] = 127;
    OV = 1;
} else if (res[x] < -128) {
    res[x] = -128;
    OV = 1;
}
Rd.B[x] = res[x];
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSUB8(unsigned long a, unsigned long b)**  
 KSUB8 (SIMD 8-bit Signed Saturating Subtraction)

**Type:** SIMD

**Syntax:**

```
KSUB8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit signed elements saturating subtractions simultaneously.

**Description:** This instruction subtracts the 8-bit signed integer elements in Rs2 from the 8-bit signed integer elements in Rs1. If any of the results are beyond the Q7 number range ( $-2^7 \leq Q7 \leq 2^7 - 1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```
res[x] = Rs1.B[x] - Rs2.B[x];
if (res[x] > (2^7)-1) {
    res[x] = (2^7)-1;
    OV = 1;
} else if (res[x] < -2^7) {
    res[x] = -2^7;
    OV = 1;
}
Rd.B[x] = res[x];
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RADD8(unsigned long a, unsigned long b)**  
 RADD8 (SIMD 8-bit Signed Halving Addition)

**Type:** SIMD

**Syntax:**

```
RADD8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit signed integer element additions simultaneously. The element results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 8-bit signed integer elements in Rs1 with the 8-bit signed integer elements in Rs2. The results are first arithmetically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Rs1 = 0x7F, Rs2 = 0x7F, Rd = 0x7F
* Rs1 = 0x80, Rs2 = 0x80, Rd = 0x80
* Rs1 = 0x40, Rs2 = 0x80, Rd = 0xE0
```

**Operations:**

```
Rd.B[x] = (Rs1.B[x] + Rs2.B[x]) s>> 1; for RV32: x=3...0, for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RSUB8(unsigned long a, unsigned long b)**  
RSUB8 (SIMD 8-bit Signed Halving Subtraction)

**Type:** SIMD

**Syntax:**

```
RSUB8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit signed integer element subtractions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 8-bit signed integer elements in Rs2 from the 8-bit signed integer elements in Rs1. The results are first arithmetically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Rs1 = 0x7F, Rs2 = 0x80, Rd = 0x7F
* Rs1 = 0x80, Rs2 = 0x7F, Rd = 0x80
* Rs1 = 0x80, Rs2 = 0x40, Rd = 0xA0
```

**Operations:**

```
Rd.B[x] = (Rs1.B[x] - Rs2.B[x]) s>> 1;
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SUB8(unsigned long a, unsigned long b)**  
SUB8 (SIMD 8-bit Subtraction)

**Type:** SIMD

**Syntax:**

```
SUB8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit integer element subtractions simultaneously.

**Description:** This instruction subtracts the 8-bit integer elements in Rs2 from the 8-bit integer elements in Rs1, and then writes the result to Rd.

**Note:** This instruction can be used for either signed or unsigned subtraction.



**Operations:**

```
Rd.B[x] = Rs1.B[x] - Rs2.B[x];
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKADD8(unsigned long a, unsigned long b)**  
 UKADD8 (SIMD 8-bit Unsigned Saturating Addition)

**Type:** SIMD**Syntax:**

```
UKADD8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit unsigned integer element saturating additions simultaneously.

**Description:** This instruction adds the 8-bit unsigned integer elements in Rs1 with the 8-bit unsigned integer elements in Rs2. If any of the results are beyond the 8-bit unsigned number range ( $0 \leq \text{RES} \leq 28-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```
res[x] = Rs1.B[x] + Rs2.B[x];
if (res[x] > (2^8)-1) {
    res[x] = (2^8)-1;
    OV = 1;
}
Rd.B[x] = res[x];
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSUB8(unsigned long a, unsigned long b)**  
 UKSUB8 (SIMD 8-bit Unsigned Saturating Subtraction)

**Type:** SIMD**Syntax:**

```
UKSUB8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit unsigned integer elements saturating subtractions simultaneously.

**Description:** This instruction subtracts the 8-bit unsigned integer elements in Rs2 from the 8-bit unsigned integer elements in Rs1. If any of the results are beyond the 8-bit unsigned number range ( $0 \leq \text{RES} \leq 28-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```

res[x] = Rs1.B[x] - Rs2.B[x];
if (res[x] < 0) {
    res[x] = 0;
    OV = 1;
}
Rd.B[x] = res[x];
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URADD8(unsigned long a, unsigned long b)**  
 URADD8 (SIMD 8-bit Unsigned Halving Addition)

**Type:** SIMD**Syntax:**

```
URADD8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit unsigned integer element additions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 8-bit unsigned integer elements in Rs1 with the 8-bit unsigned integer elements in Rs2. The results are first logically right-shifted by 1 bit and then written to Rd.

**Examples:**

```

* Ra = 0x7F, Rb = 0x7F, Rt = 0x7F
* Ra = 0x80, Rb = 0x80, Rt = 0x80
* Ra = 0x40, Rb = 0x80, Rt = 0x60

```

**Operations:**

```

Rd.B[x] = (Rs1.B[x] + Rs2.B[x]) u>> 1;
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URSUB8(unsigned long a, unsigned long b)**  
 URSUB8 (SIMD 8-bit Unsigned Halving Subtraction)

**Type:** SIMD**Syntax:**

```
URSUB8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit unsigned integer element subtractions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 8-bit unsigned integer elements in Rs2 from the 8-bit unsigned integer elements in Rs1. The results are first logically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Ra = 0x7F, Rb = 0x80 Rt = 0xFF
* Ra = 0x80, Rb = 0x7F Rt = 0x00
* Ra = 0x80, Rb = 0x40 Rt = 0x20
```

**Operations:**

```
Rd.B[x] = (Rs1.B[x] - Rs2.B[x]) u>> 1;
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## SIMD 16-bit Shift Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_KSLI16(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_KSLRA16(unsigned long a, int b)
__STATIC_FORCEINLINE unsigned long __RV_KSLRA16_U(unsigned long a, int b)
__STATIC_FORCEINLINE unsigned long __RV_SLLI16(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRAI16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SRAI16_U(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SRLI16(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRLI16_U(unsigned long a, unsigned int b)
__RV_KSLI16(a, b)
__RV_SLLI16(a, b)
__RV_SRAI16(a, b)
__RV_SRAI16_U(a, b)
__RV_SRLI16(a, b)
__RV_SRLI16_U(a, b)
```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_16B\_SHIFT**

SIMD 16-bit Shift Instructions.

there are 14 SIMD 16-bit shift instructions.

## Defines

### RV\_KSLLI16 (a, b)

KSLLI16 (SIMD 16-bit Saturating Shift Left Logical Immediate)

**Type:** SIMD

**Syntax:**

```
KSLLI16 Rd, Rs1, imm4u
```

**Purpose:** Do 16-bit elements logical left shift operations with saturation simultaneously. The shift amount is an immediate value.

**Description:** The 16-bit data elements in Rs1 are left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the imm4u constant. Any shifted value greater than  $2^{15}-1$  is saturated to  $2^{15}-1$ . Any shifted value smaller than  $-2^{15}$  is saturated to  $-2^{15}$ . And the saturated results are written to Rd. If any saturation is performed, set OV bit to 1.

**Operations:**

```
sa = imm4u[3:0];
if (sa != 0) {
    res[(15+sa):0] = Rs1.H[x] << sa;
    if (res > (2^15)-1) {
        res = 0x7fff; OV = 1;
    } else if (res < -2^15) {
        res = 0x8000; OV = 1;
    }
    Rd.H[x] = res[15:0];
} else {
    Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

### RV\_SLLI16 (a, b)

SLLI16 (SIMD 16-bit Shift Left Logical Immediate)

**Type:** SIMD

**Syntax:**

```
SLLI16 Rd, Rs1, imm4[3:0]
```

**Purpose:** Do 16-bit element logical left shift operations simultaneously. The shift amount is an immediate value.

**Description:** The 16-bit elements in Rs1 are left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the imm4[3:0] constant. And the results are written to Rd.

**Operations:**

```

sa = imm4[3:0];
Rd.H[x] = Rs1.H[x] << sa;
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

\_\_\_RV\_SRAI16 (a, b)

SRAI16 (SIMD 16-bit Shift Right Arithmetic Immediate)

**Type:** SIMD

**Syntax:**

```

SRAI16 Rd, Rs1, imm4u
SRAI16.u Rd, Rs1, imm4u

```

**Purpose:** Do 16-bit elements arithmetic right shift operations simultaneously. The shift amount is an immediate value. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 16-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the 16-bit data elements. The shift amount is specified by the imm4u constant. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 16-bit data to calculate the final results. And the results are written to Rd.

**Operations:**

```

sa = imm4u[3:0];
if (sa > 0) {
    if (`.u` form) { // SRAI16.u
        res[15:-1] = SE17(Rs1.H[x][15:sa-1]) + 1;
        Rd.H[x] = res[15:0];
    } else { // SRAI16
        Rd.H[x] = SE16(Rs1.H[x][15:sa]);
    }
} else {
    Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

\_\_\_RV\_SRAI16\_U (a, b)

SRAI16.u (SIMD 16-bit Rounding Shift Right Arithmetic Immediate)

**Type:** SIMD

**Syntax:**

```
SRAI16 Rd, Rs1, imm4u
SRAI16.u Rd, Rs1, imm4u
```

**Purpose:** Do 16-bit elements arithmetic right shift operations simultaneously. The shift amount is an immediate value. The `.u` form performs additional rounding up operations on the shifted results.

**Description:** The 16-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the 16-bit data elements. The shift amount is specified by the imm4u constant. For the rounding operation of the `.u` form, a value of 1 is added to the most significant discarded bit of each 16-bit data to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = imm4u[3:0];
if (sa > 0) {
    if (`.u` form) { // SRAI16.u
        res[15:-1] = SE17(Rs1.H[x][15:sa-1]) + 1;
        Rd.H[x] = res[15:0];
    } else { // SRAI16
        Rd.H[x] = SE16(Rs1.H[x][15:sa]);
    }
} else {
    Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**RV\_SRLI16 (a, b)**

SRLI16 (SIMD 16-bit Shift Right Logical Immediate)

**Type:** SIMD

**Syntax:**

```
SRLI16 Rt, Ra, imm4u
SRLI16.u Rt, Ra, imm4u
```

**Purpose:** Do 16-bit elements logical right shift operations simultaneously. The shift amount is an immediate value. The `.u` form performs additional rounding up operations on the shifted results.

**Description:** The 16-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the imm4u constant. For the rounding operation of the `.u` form, a value of 1 is added to the most significant discarded bit of each 16-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = imm4u;
if (sa > 0) {
    if (`.u` form) { // SRLI16.u
```

(continues on next page)

(continued from previous page)

```

    res[16:0] = ZE17(Rs1.H[x][15:sa-1]) + 1;
    Rd.H[x] = res[16:1];
  } else { // SRLI16
    Rd.H[x] = ZE16(Rs1.H[x][15:sa]);
  }
} else {
  Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_RV\_SRLI16\_U**(a, b)

SRLI16.u (SIMD 16-bit Rounding Shift Right Logical Immediate)

**Type:** SIMD

**Syntax:**

```

SRLI16 Rt, Ra, imm4u
SRLI16.u Rt, Ra, imm4u

```

**Purpose:** Do 16-bit elements logical right shift operations simultaneously. The shift amount is an immediate value. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 16-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the imm4u constant. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 16-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```

sa = imm4u;
if (sa > 0) {
  if (`.u` form) { // SRLI16.u
    res[16:0] = ZE17(Rs1.H[x][15:sa-1]) + 1;
    Rd.H[x] = res[16:1];
  } else { // SRLI16
    Rd.H[x] = ZE16(Rs1.H[x][15:sa]);
  }
} else {
  Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

- [in] b: unsigned int type of value stored in b

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSL16(unsigned long a, unsigned int b)**  
KSL16 (SIMD 16-bit Saturating Shift Left Logical)

**Type:** SIMD

**Syntax:**

```
KSL16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit elements logical left shift operations with saturation simultaneously. The shift amount is a variable from a GPR.

**Description:** The 16-bit data elements in Rs1 are left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the low-order 4-bits of the value in the Rs2 register. Any shifted value greater than  $2^{15}-1$  is saturated to  $2^{15}-1$ . Any shifted value smaller than  $-2^{15}$  is saturated to  $-2^{15}$ . And the saturated results are written to Rd. If any saturation is performed, set OV bit to 1.

**Operations:**

```
sa = Rs2[3:0];
if (sa != 0) {
    res[(15+sa):0] = Rs1.H[x] << sa;
    if (res > (2^15)-1) {
        res = 0x7fff; OV = 1;
    } else if (res < -2^15) {
        res = 0x8000; OV = 1;
    }
    Rd.H[x] = res[15:0];
} else {
    Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSLRA16(unsigned long a, int b)**  
KSLRA16 (SIMD 16-bit Shift Left Logical with Saturation or Shift Right Arithmetic)

**Type:** SIMD

**Syntax:**

```
KSLRA16 Rd, Rs1, Rs2
KSLRA16.u Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit elements logical left (positive) or arithmetic right (negative) shift operation with Q15 saturation for the left shift. The .u form performs additional rounding up operations for the right shift.



**Description:** The 16-bit data elements of Rs1 are left-shifted logically or right-shifted arithmetically based on the value of Rs2[4:0]. Rs2[4:0] is in the signed range of  $[-2^4, 2^4-1]$ . A positive Rs2[4:0] means logical left shift and a negative Rs2[4:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[4:0]. However, the behavior of  $\text{Rs2}[4:0] == -2^4$  (0x10) is defined to be equivalent to the behavior of  $\text{Rs2}[4:0] == -(2^4-1)$  (0x11). The left-shifted results are saturated to the 16-bit signed integer range of  $[-2^{15}, 2^{15}-1]$ . For the .u form of the instruction, the right-shifted results are added a 1 to the most significant discarded bit position for rounding effect. After the shift, saturation, or rounding, the final results are written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:5] will not affect this instruction.

**Operations:**

```

if (Rs2[4:0] < 0) {
    sa = -Rs2[4:0];
    sa = (sa == 16)? 15 : sa;
    if (`.u` form) {
        res[15:-1] = SE17(Rs1.H[x][15:sa-1]) + 1;
        Rd.H[x] = res[15:0];
    } else {
        Rd.H[x] = SE16(Rs1.H[x][15:sa]);
    }
} else {
    sa = Rs2[3:0];
    res[(15+sa):0] = Rs1.H[x] <<(logic) sa;
    if (res > (2^15)-1) {
        res[15:0] = 0x7fff; OV = 1;
    } else if (res < -2^15) {
        res[15:0] = 0x8000; OV = 1;
    }
    d.H[x] = res[15:0];
}
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSLRA16\_U(unsigned long a, int b)**  
KSLRA16.u (SIMD 16-bit Shift Left Logical with Saturation or Rounding Shift Right Arithmetic)

**Type:** SIMD

**Syntax:**

```

KSLRA16 Rd, Rs1, Rs2
KSLRA16.u Rd, Rs1, Rs2

```

**Purpose:** Do 16-bit elements logical left (positive) or arithmetic right (negative) shift operation with Q15 saturation for the left shift. The .u form performs additional rounding up operations for the right shift.

**Description:** The 16-bit data elements of Rs1 are left-shifted logically or right-shifted arithmetically based on the value of Rs2[4:0]. Rs2[4:0] is in the signed range of  $[-2^4, 2^4-1]$ . A positive Rs2[4:0] means logical left shift and a negative Rs2[4:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[4:0]. However, the behavior of  $\text{Rs2}[4:0] == -2^4$  (0x10) is defined to be equivalent to the behavior of  $\text{Rs2}[4:0] == -(2^4-1)$  (0x11). The left-shifted results are saturated to the 16-bit

signed integer range of  $[-2^{15}, 2^{15}-1]$ . For the `.u` form of the instruction, the right-shifted results are added a 1 to the most significant discarded bit position for rounding effect. After the shift, saturation, or rounding, the final results are written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:5] will not affect this instruction.

#### Operations:

```

if (Rs2[4:0] < 0) {
    sa = -Rs2[4:0];
    sa = (sa == 16)? 15 : sa;
    if (`.u` form) {
        res[15:-1] = SE17(Rs1.H[x][15:sa-1]) + 1;
        Rd.H[x] = res[15:0];
    } else {
        Rd.H[x] = SE16(Rs1.H[x][15:sa]);
    }
} else {
    sa = Rs2[3:0];
    res[(15+sa):0] = Rs1.H[x] <<(logic) sa;
    if (res > (2^15)-1) {
        res[15:0] = 0x7fff; OV = 1;
    } else if (res < -2^15) {
        res[15:0] = 0x8000; OV = 1;
    }
    d.H[x] = res[15:0];
}
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SLL16(unsigned long a, unsigned int b)**  
SLL16 (SIMD 16-bit Shift Left Logical)

**Type:** SIMD

#### Syntax:

```
SLL16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit elements logical left shift operations simultaneously. The shift amount is a variable from a GPR.

**Description:** The 16-bit elements in Rs1 are left-shifted logically. And the results are written to Rd. The shifted out bits are filled with zero and the shift amount is specified by the low-order 4-bits of the value in the Rs2 register.

#### Operations:

```

sa = Rs2[3:0];
Rd.H[x] = Rs1.H[x] << sa;
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRA16(unsigned long a, unsigned long b)**  
SRA16 (SIMD 16-bit Shift Right Arithmetic)

**Type:** SIMD

**Syntax:**

```
SRA16 Rd, Rs1, Rs2
SRA16.u Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit element arithmetic right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 16-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the data elements. The shift amount is specified by the low-order 4-bits of the value in the Rs2 register. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 16-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = Rs2[3:0];
if (sa != 0) {
    if (`.u` form) { // SRA16.u
        res[15:-1] = SE17(Rs1.H[x][15:sa-1]) + 1;
        Rd.H[x] = res[15:0];
    } else { // SRA16
        Rd.H[x] = SE16(Rs1.H[x][15:sa])
    }
} else {
    Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRA16\_U(unsigned long a, unsigned long b)**  
SRA16.u (SIMD 16-bit Rounding Shift Right Arithmetic)

**Type:** SIMD

**Syntax:**

```
SRA16 Rd, Rs1, Rs2
SRA16.u Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit element arithmetic right shift operations simultaneously. The shift amount is a variable from a GPR. The `.u` form performs additional rounding up operations on the shifted results.

**Description:** The 16-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the data elements. The shift amount is specified by the low-order 4-bits of the value in the Rs2 register. For the rounding operation of the `.u` form, a value of 1 is added to the most significant discarded bit of each 16-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = Rs2[3:0];
if (sa != 0) {
    if (`.u` form) { // SRA16.u
        res[15:-1] = SE17(Rs1.H[x][15:sa-1]) + 1;
        Rd.H[x] = res[15:0];
    } else { // SRA16
        Rd.H[x] = SE16(Rs1.H[x][15:sa])
    }
} else {
    Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRL16(unsigned long a, unsigned int b)**  
SRL16 (SIMD 16-bit Shift Right Logical)

**Type:** SIMD

**Syntax:**

```
SRL16 Rt, Ra, Rb
SRL16.u Rt, Ra, Rb
```

**Purpose:** Do 16-bit elements logical right shift operations simultaneously. The shift amount is a variable from a GPR. The `.u` form performs additional rounding up operations on the shifted results.

**Description:** The 16-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the low-order 4-bits of the value in the Rs2 register. For the rounding operation of the `.u` form, a value of 1 is added to the most significant discarded bit of each 16-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = Rs2[3:0];
if (sa > 0) {
    if (`.u` form) { // SRL16.u
        res[16:0] = ZE17(Rs1.H[x][15:sa-1]) + 1;
        Rd.H[x] = res[16:1];
    } else { // SRL16
        Rd.H[x] = ZE16(Rs1.H[x][15:sa]);
    }
}
```

(continues on next page)

(continued from previous page)

```

    }
} else {
    Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRL16\_U(unsigned long a, unsigned int b)**  
SRL16.u (SIMD 16-bit Rounding Shift Right Logical)

**Type:** SIMD

**Syntax:**

```

SRL16 Rt, Ra, Rb
SRL16.u Rt, Ra, Rb

```

**Purpose:** Do 16-bit elements logical right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding upoperations on the shifted results.

**Description:** The 16-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the low-order 4-bits of the value in the Rs2 register. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 16-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```

sa = Rs2[3:0];
if (sa > 0) {
    if (` .u` form) { // SRL16.u
        res[16:0] = ZE17(Rs1.H[x][15:sa-1]) + 1;
        Rd.H[x] = res[16:1];
    } else { // SRL16
        Rd.H[x] = ZE16(Rs1.H[x][15:sa]);
    }
} else {
    Rd = Rs1;
}
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

## SIMD 8-bit Shift Instructions

```

__STATIC_FORCEINLINE unsigned long __RV_KSLL8(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_KSLRA8(unsigned long a, int b)
__STATIC_FORCEINLINE unsigned long __RV_KSLRA8_U(unsigned long a, int b)
__STATIC_FORCEINLINE unsigned long __RV_SLL8(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRA8(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRA8_U(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRL8(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRL8_U(unsigned long a, unsigned int b)

__RV_KSLLI8(a, b)
__RV_SLLI8(a, b)
__RV_SRAI8(a, b)
__RV_SRAI8_U(a, b)
__RV_SRLI8(a, b)
__RV_SRLI8_U(a, b)

```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_8B\_SHIFT**

SIMD 8-bit Shift Instructions.

there are 14 SIMD 8-bit shift instructions.

## Defines

**\_\_RV\_KSLLI8**(a, b)  
KSLLI8 (SIMD 8-bit Saturating Shift Left Logical Immediate)

**Type:** SIMD

**Syntax:**

```
KSLLI8 Rd, Rs1, imm3u
```

**Purpose:** Do 8-bit elements logical left shift operations with saturation simultaneously. The shift amount is an immediate value.

**Description:** The 8-bit data elements in Rs1 are left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the imm3u constant. Any shifted value greater than  $2^7-1$  is saturated to  $2^7-1$ . Any shifted value smaller than  $-2^7$  is saturated to  $-2^7$ . And the saturated results are written to Rd. If any saturation is performed, set OV bit to 1.

**Operations:**

```

sa = imm3u[2:0];
if (sa != 0) {
    res[(7+sa):0] = Rs1.B[x] << sa;
    if (res > (2^7)-1) {
        res = 0x7f; OV = 1;
    } else if (res < -2^7) {
        res = 0x80; OV = 1;
    }
}

```

(continues on next page)

(continued from previous page)

```

    }
    Rd.B[x] = res[7:0];
} else {
    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

\_\_RV\_SLLI8 (a, b)

SLLI8 (SIMD 8-bit Shift Left Logical Immediate)

**Type:** SIMD

**Syntax:**

```
SLLI8 Rd, Rs1, imm3u
```

**Purpose:** Do 8-bit elements logical left shift operations simultaneously. The shift amount is an immediate value.

**Description:** The 8-bit elements in Rs1 are left-shifted logically. And the results are written to Rd. The shifted out bits are filled with zero and the shift amount is specified by the imm3u constant.

**Operations:**

```

sa = imm3u[2:0];
Rd.B[x] = Rs1.B[x] << sa;
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

\_\_RV\_SRAI8 (a, b)

SRAI8 (SIMD 8-bit Shift Right Arithmetic Immediate)

**Type:** SIMD

**Syntax:**

```

SRAI8 Rd, Rs1, imm3u
SRAI8.u Rd, Rs1, imm3u

```

**Purpose:** Do 8-bit element arithmetic right shift operations simultaneously. The shift amount is an immediate value. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 8-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the data elements. The shift amount is specified by the imm3u constant. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 8-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = imm3u[2:0];
if (sa > 0) {
    if (`.u` form) { // SRA8.u
        res[7:-1] = SE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[7:0];
    } else { // SRA8
        Rd.B[x] = SE8(Rd.B[x][7:sa])
    }
} else {
    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**RV\_SRAI8\_U** (a, b)

SRAI8.u (SIMD 8-bit Rounding Shift Right Arithmetic Immediate)

**Type:** SIMD

**Syntax:**

```
SRAI8 Rd, Rs1, imm3u
SRAI8.u Rd, Rs1, imm3u
```

**Purpose:** Do 8-bit element arithmetic right shift operations simultaneously. The shift amount is an immediate value. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 8-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the data elements. The shift amount is specified by the imm3u constant. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 8-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = imm3u[2:0];
if (sa > 0) {
    if (`.u` form) { // SRA8.u
        res[7:-1] = SE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[7:0];
    } else { // SRA8
        Rd.B[x] = SE8(Rd.B[x][7:sa])
    }
} else {
    Rd = Rs1;
}
```

(continues on next page)



(continued from previous page)

```

}
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

\_\_\_RV\_SRLI8 (a, b)

SRLI8 (SIMD 8-bit Shift Right Logical Immediate)

**Type:** SIMD

**Syntax:**

```

SRLI8 Rt, Ra, imm3u
SRLI8.u Rt, Ra, imm3u

```

**Purpose:** Do 8-bit elements logical right shift operations simultaneously. The shift amount is an immediate value. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 8-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the imm3u constant. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 8-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```

sa = imm3u[2:0];
if (sa > 0) {
    if (`.u` form) { // SRLI8.u
        res[8:0] = ZE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[8:1];
    } else { // SRLI8
        Rd.B[x] = ZE8(Rs1.B[x][7:sa]);
    }
} else {
    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

\_\_\_RV\_SRLI8\_U (a, b)

SRLI8.u (SIMD 8-bit Rounding Shift Right Logical Immediate)

**Type:** SIMD

**Syntax:**

```
SRLI8 Rt, Ra, imm3u
SRLI8.u Rt, Ra, imm3u
```

**Purpose:** Do 8-bit elements logical right shift operations simultaneously. The shift amount is an immediate value. The `.u` form performs additional rounding up operations on the shifted results.

**Description:** The 8-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the `imm3u` constant. For the rounding operation of the `.u` form, a value of 1 is added to the most significant discarded bit of each 8-bit data element to calculate the final results. And the results are written to Rd.

#### Operations:

```
sa = imm3u[2:0];
if (sa > 0) {
    if (`.u` form) { // SRLI8.u
        res[8:0] = ZE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[8:1];
    } else { // SRLI8
        Rd.B[x] = ZE8(Rs1.B[x][7:sa]);
    }
} else {
    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

## Functions

`__STATIC_FORCEINLINE unsigned long __RV_KSLL8(unsigned long a, unsigned int b)`  
KSLL8 (SIMD 8-bit Saturating Shift Left Logical)

**Type:** SIMD

#### Syntax:

```
KSLL8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit elements logical left shift operations with saturation simultaneously. The shift amount is a variable from a GPR.

**Description:** The 8-bit data elements in Rs1 are left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the low-order 3-bits of the value in the Rs2 register. Any shifted value greater than  $2^7-1$  is saturated to  $2^7-1$ . Any shifted value smaller than  $-2^7$  is saturated to  $-2^7$ . And the saturated results are written to Rd. If any saturation is performed, set OV bit to 1.

#### Operations:

```

sa = Rs2[2:0];
if (sa != 0) {
    res[(7+sa):0] = Rs1.B[x] << sa;
    if (res > (2^7)-1) {
        res = 0x7f; OV = 1;
    } else if (res < -2^7) {
        res = 0x80; OV = 1;
    }
    Rd.B[x] = res[7:0];
} else {
    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSLRA8(unsigned long a, int b)**  
 KSLRA8 (SIMD 8-bit Shift Left Logical with Saturation or Shift Right Arithmetic)

**Type:** SIMD

#### Syntax:

```

KSLRA8 Rd, Rs1, Rs2
KSLRA8.u Rd, Rs1, Rs2

```

**Purpose:** Do 8-bit elements logical left (positive) or arithmetic right (negative) shift operation with Q7 saturation for the left shift. The .u form performs additional rounding up operations for the right shift.

**Description:** The 8-bit data elements of Rs1 are left-shifted logically or right-shifted arithmetically based on the value of Rs2[3:0]. Rs2[3:0] is in the signed range of  $[-2^3, 2^3-1]$ . A positive Rs2[3:0] means logical left shift and a negative Rs2[3:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[3:0]. However, the behavior of  $Rs2[3:0] == -2^3$  (0x8) is defined to be equivalent to the behavior of  $Rs2[3:0] == -(2^3-1)$  (0x9). The left-shifted results are saturated to the 8-bit signed integer range of  $[-2^7, 2^7-1]$ . For the .u form of the instruction, the right-shifted results are added a 1 to the most significant discarded bit position for rounding effect. After the shift, saturation, or rounding, the final results are written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:4] will not affect this instruction.

#### Operations:

```

if (Rs2[3:0] < 0) {
    sa = -Rs2[3:0];
    sa = (sa == 8)? 7 : sa;
    if (`u` form) {
        res[7:-1] = SE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[7:0];
    } else {
        Rd.B[x] = SE8(Rs1.B[x][7:sa]);
    }
} else {
    sa = Rs2[2:0];

```

(continues on next page)

(continued from previous page)

```

    res[(7+sa):0] = Rs1.B[x] <<(logic) sa;
    if (res > (2^7)-1) {
        res[7:0] = 0x7f; OV = 1;
    } else if (res < -2^7) {
        res[7:0] = 0x80; OV = 1;
    }
    Rd.B[x] = res[7:0];
}
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSLRA8\_U(unsigned long a, int b)**  
 KSLRA8.u (SIMD 8-bit Shift Left Logical with Saturation or Rounding Shift Right Arithmetic)

**Type:** SIMD

#### Syntax:

```

KSLRA8 Rd, Rs1, Rs2
KSLRA8.u Rd, Rs1, Rs2

```

**Purpose:** Do 8-bit elements logical left (positive) or arithmetic right (negative) shift operation with Q7 saturation for the left shift. The .u form performs additional rounding up operations for the right shift.

**Description:** The 8-bit data elements of Rs1 are left-shifted logically or right-shifted arithmetically based on the value of Rs2[3:0]. Rs2[3:0] is in the signed range of  $[-2^3, 2^3-1]$ . A positive Rs2[3:0] means logical left shift and a negative Rs2[3:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[3:0]. However, the behavior of  $Rs2[3:0] == -2^3$  (0x8) is defined to be equivalent to the behavior of  $Rs2[3:0] == -(2^3-1)$  (0x9). The left-shifted results are saturated to the 8-bit signed integer range of  $[-2^7, 2^7-1]$ . For the .u form of the instruction, the right-shifted results are added a 1 to the most significant discarded bit position for rounding effect. After the shift, saturation, or rounding, the final results are written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:4] will not affect this instruction.

#### Operations:

```

if (Rs2[3:0] < 0) {
    sa = -Rs2[3:0];
    sa = (sa == 8)? 7 : sa;
    if (`.u` form) {
        res[7:-1] = SE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[7:0];
    } else {
        Rd.B[x] = SE8(Rs1.B[x][7:sa]);
    }
} else {
    sa = Rs2[2:0];
    res[(7+sa):0] = Rs1.B[x] <<(logic) sa;
    if (res > (2^7)-1) {
        res[7:0] = 0x7f; OV = 1;
    }
}

```

(continues on next page)

(continued from previous page)

```

    } else if (res < -2^7) {
        res[7:0] = 0x80; OV = 1;
    }
    Rd.B[x] = res[7:0];
}
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SLL8(unsigned long a, unsigned int b)**  
SLL8 (SIMD 8-bit Shift Left Logical)

**Type:** SIMD

#### Syntax:

```
SLL8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit elements logical left shift operations simultaneously. The shift amount is a variable from a GPR.

**Description:** The 8-bit elements in Rs1 are left-shifted logically. And the results are written to Rd. The shifted out bits are filled with zero and the shift amount is specified by the low-order 3-bits of the value in the Rs2 register.

#### Operations:

```

sa = Rs2[2:0];
Rd.B[x] = Rs1.B[x] << sa;
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRA8(unsigned long a, unsigned int b)**  
SRA8 (SIMD 8-bit Shift Right Arithmetic)

**Type:** SIMD

#### Syntax:

```

SRA8 Rd, Rs1, Rs2
SRA8.u Rd, Rs1, Rs2

```

**Purpose:** Do 8-bit element arithmetic right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 8-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the data elements. The shift amount is specified by the low-order 3-bits of the value in the Rs2 register. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 8-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = Rs2[2:0];
if (sa > 0) {
    if (`.u` form) { // SRA8.u
        res[7:-1] = SE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[7:0];
    } else { // SRA8
        Rd.B[x] = SE8(Rd.B[x][7:sa])
    }
} else {
    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRA8\_U(unsigned long a, unsigned int b)**  
SRA8.u (SIMD 8-bit Rounding Shift Right Arithmetic)

**Type:** SIMD

**Syntax:**

```
SRA8 Rd, Rs1, Rs2
SRA8.u Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit element arithmetic right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 8-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the data elements. The shift amount is specified by the low-order 3-bits of the value in the Rs2 register. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 8-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = Rs2[2:0];
if (sa > 0) {
    if (`.u` form) { // SRA8.u
        res[7:-1] = SE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[7:0];
    } else { // SRA8
        Rd.B[x] = SE8(Rd.B[x][7:sa])
    }
} else {
```

(continues on next page)

(continued from previous page)

```

    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRL8(unsigned long a, unsigned int b)**  
SRL8 (SIMD 8-bit Shift Right Logical)

**Type:** SIMD

**Syntax:**

```

SRL8 Rt, Ra, Rb
SRL8.u Rt, Ra, Rb

```

**Purpose:** Do 8-bit elements logical right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 8-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the low-order 3-bits of the value in the Rs2 register. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 8-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```

sa = Rs2[2:0];
if (sa > 0) {
    if (`.u` form) { // SRL8.u
        res[8:0] = ZE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[8:1];
    } else { // SRL8
        Rd.B[x] = ZE8(Rs1.B[x][7:sa]);
    }
} else {
    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRL8\_U(unsigned long a, unsigned int b)**  
SRL8.u (SIMD 8-bit Rounding Shift Right Logical)

**Type:** SIMD

**Syntax:**

```
SRL8 Rt, Ra, Rb
SRL8.u Rt, Ra, Rb
```

**Purpose:** Do 8-bit elements logical right shift operations simultaneously. The shift amount is a variable from a GPR. The `.u` form performs additional rounding up operations on the shifted results.

**Description:** The 8-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the low-order 3-bits of the value in the Rs2 register. For the rounding operation of the `.u` form, a value of 1 is added to the most significant discarded bit of each 8-bit data element to calculate the final results. And the results are written to Rd.

#### Operations:

```
sa = Rs2[2:0];
if (sa > 0) {
    if (`.u` form) { // SRL8.u
        res[8:0] = ZE9(Rs1.B[x][7:sa-1]) + 1;
        Rd.B[x] = res[8:1];
    } else { // SRL8
        Rd.B[x] = ZE8(Rs1.B[x][7:sa]);
    }
} else {
    Rd = Rs1;
}
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

## SIMD 16-bit Compare Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_CMPEQ16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SCMPL16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SCMPLT16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UCMPL16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UCMPLT16(unsigned long a, unsigned long b)
```

group **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_16B\_CMP**

SIMD 16-bit Compare Instructions.

there are 5 SIMD 16-bit Compare instructions.

#### Functions

```
__STATIC_FORCEINLINE unsigned long __RV_CMPEQ16(unsigned long a, unsigned long b)
    CMPEQ16 (SIMD 16-bit Integer Compare Equal)
```

Type: SIMD



**Syntax:**

```
CMPEQ16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit integer elements equal comparisons simultaneously.

**Description:** This instruction compares the 16-bit integer elements in Rs1 with the 16-bit integer elements in Rs2 to see if they are equal. If they are equal, the result is 0xFFFF; otherwise, the result is 0x0. The 16-bit element comparison results are written to Rt.

**Note:** This instruction can be used for either signed or unsigned numbers.

**Operations:**

```
Rd.H[x] = (Rs1.H[x] == Rs2.H[x])? 0xffff : 0x0;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE unsigned long __RV_SCMPLE16(unsigned long a, unsigned long b)
    SCMPLE16 (SIMD 16-bit Signed Compare Less Than & Equal)
```

**Type:** SIMD

**Syntax:**

```
SCMPLE16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer elements less than & equal comparisons simultaneously.

**Description:** This instruction compares the 16-bit signed integer elements in Rs1 with the 16-bit signed integer elements in Rs2 to see if the one in Rs1 is less than or equal to the one in Rs2. If it is true, the result is 0xFFFF; otherwise, the result is 0x0. The element comparison results are written to Rd.

**Operations:**

```
Rd.H[x] = (Rs1.H[x] {le} Rs2.H[x])? 0xffff : 0x0;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE unsigned long __RV_SCMPLE16(unsigned long a, unsigned long b)
    SCMPLE16 (SIMD 16-bit Signed Compare Less Than)
```

**Type:** SIMD

**Syntax:**

```
SCMPLT16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer elements less than comparisons simultaneously.

**Description:** This instruction compares the 16-bit signed integer elements in Rs1 with the two 16-bit signed integer elements in Rs2 to see if the one in Rs1 is less than the one in Rs2. If it is true, the result is 0xFFFF; otherwise, the result is 0x0. The element comparison results are written to Rd.

**Operations:**

```
Rd.H[x] = (Rs1.H[x] < Rs2.H[x])? 0xffff : 0x0;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UCMPLE16(unsigned long a, unsigned long b)**  
UCMPLE16 (SIMD 16-bit Unsigned Compare Less Than & Equal)

**Type:** SIMD

**Syntax:**

```
UCMPLE16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer elements less than & equal comparisons simultaneously.

**Description:** This instruction compares the 16-bit unsigned integer elements in Rs1 with the 16-bit unsigned integer elements in Rs2 to see if the one in Rs1 is less than or equal to the one in Rs2. If it is true, the result is 0xFFFF; otherwise, the result is 0x0. The element comparison results are written to Rd.

**Operations:**

```
Rd.H[x] = (Rs1.H[x] <=u Rs2.H[x])? 0xffff : 0x0;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UCMPLT16(unsigned long a, unsigned long b)**  
UCMPLT16 (SIMD 16-bit Unsigned Compare Less Than)

**Type:** SIMD

**Syntax:**

```
UCMPLT16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer elements less than comparisons simultaneously.

**Description:** This instruction compares the 16-bit unsigned integer elements in Rs1 with the 16-bit unsigned integer elements in Rs2 to see if the one in Rs1 is less than the one in Rs2. If it is true, the result is 0xFFFF; otherwise, the result is 0x0. The element comparison results are written to Rd.

**Operations:**

```
Rd.H[x] = (Rs1.H[x] <u Rs2.H[x])? 0xffff : 0x0;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## SIMD 8-bit Compare Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_CMPEQ8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SCMPLT8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SCMLT8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UCMPLE8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UCMLT8(unsigned long a, unsigned long b)
```

group NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_8B\_CMP  
SIMD 8-bit Compare Instructions.

there are 5 SIMD 8-bit Compare instructions.

## Functions

```
__STATIC_FORCEINLINE unsigned long __RV_CMPEQ8(unsigned long a, unsigned long b)
CMPEQ8 (SIMD 8-bit Integer Compare Equal)
```

**Type:** SIMD

**Syntax:**

```
CMPEQ8 Rs, Rs1, Rs2
```

**Purpose:** Do 8-bit integer elements equal comparisons simultaneously.

**Description:** This instruction compares the 8-bit integer elements in Rs1 with the 8-bit integer elements in Rs2 to see if they are equal. If they are equal, the result is 0xFF; otherwise, the result is 0x0. The 8-bit element comparison results are written to Rd.

**Note:** This instruction can be used for either signed or unsigned numbers.

**Operations:**

```
Rd.B[x] = (Rs1.B[x] == Rs2.B[x])? 0xff : 0x0;
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SCMPLT8(unsigned long a, unsigned long b)**  
SCMPLT8 (SIMD 8-bit Signed Compare Less Than & Equal)

**Type:** SIMD

**Syntax:**

SCMPLT8 Rd, Rs1, Rs2
----------------------

**Purpose:** Do 8-bit signed integer elements less than & equal comparisons simultaneously.

**Description:** This instruction compares the 8-bit signed integer elements in Rs1 with the 8-bit signed integer elements in Rs2 to see if the one in Rs1 is less than or equal to the one in Rs2. If it is true, the result is 0xFF; otherwise, the result is 0x0. The element comparison results are written to Rd

**Operations:**

Rd.B[x] = (Rs1.B[x] {le} Rs2.B[x])? 0xff : 0x0; for RV32: x=3...0, for RV64: x=7...0
--

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SCMPLT8(unsigned long a, unsigned long b)**  
SCMPLT8 (SIMD 8-bit Signed Compare Less Than)

**Type:** SIMD

**Syntax:**

SCMPLT8 Rd, Rs1, Rs2
----------------------

**Purpose:** Do 8-bit signed integer elements less than comparisons simultaneously.

**Description:** This instruction compares the 8-bit signed integer elements in Rs1 with the 8-bit signed integer elements in Rs2 to see if the one in Rs1 is less than the one in Rs2. If it is true, the result is 0xFF; otherwise, the result is 0x0. The element comparison results are written to Rd.

**Operations:**

Rd.B[x] = (Rs1.B[x] < Rs2.B[x])? 0xff : 0x0; for RV32: x=3...0, for RV64: x=7...0
---

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UCMPLE8(unsigned long a, unsigned long b)**  
UCMPLE8 (SIMD 8-bit Unsigned Compare Less Than & Equal)

**Type:** SIMD

**Syntax:**

```
UCMPLE8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit unsigned integer elements less than & equal comparisons simultaneously.

**Description:** This instruction compares the 8-bit unsigned integer elements in Rs1 with the 8-bit unsigned integer elements in Rs2 to see if the one in Rs1 is less than or equal to the one in Rs2. If it is true, the result is 0xFF; otherwise, the result is 0x0. The four comparison results are written to Rd.

**Operations:**

```
Rd.B[x] = (Rs1.B[x] <=u Rs2.B[x])? 0xff : 0x0;
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UCMPLT8(unsigned long a, unsigned long b)**  
UCMPLT8 (SIMD 8-bit Unsigned Compare Less Than)

**Type:** SIMD

**Syntax:**

```
UCMPLT8 Rd, Rs1, Rs2
```

**Purpose:** Do 8-bit unsigned integer elements less than comparisons simultaneously.

**Description:** This instruction compares the 8-bit unsigned integer elements in Rs1 with the 8-bit unsigned integer elements in Rs2 to see if the one in Rs1 is less than the one in Rs2. If it is true, the result is 0xFF; otherwise, the result is 0x0. The element comparison results are written to Rd.

**Operations:**

```
Rd.B[x] = (Rs1.B[x] <u Rs2.B[x])? 0xff : 0x0;
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## SIMD 16-bit Multiply Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_KHM16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KHMX16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long long __RV_SMUL16(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long long __RV_SMULX16(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long long __RV_UMUL16(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long long __RV_UMULX16(unsigned int a, unsigned int b)
```

group **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_16B\_MULTIPLY**  
SIMD 16-bit Multiply Instructions.

there are 6 SIMD 16-bit Multiply instructions.

### Functions

```
__STATIC_FORCEINLINE unsigned long __RV_KHM16(unsigned long a, unsigned long b)
    KHM16 (SIMD Signed Saturating Q15 Multiply)
```

**Type:** SIMD

**Syntax:**

```
KHM16 Rd, Rs1, Rs2
KHMX16 Rd, Rs1, Rs2
```

**Purpose:** Do Q15xQ15 element multiplications simultaneously. The Q30 results are then reduced to Q15 numbers again.

**Description:** For the KHM16 instruction, multiply the top 16-bit Q15 content of 32-bit chunks in Rs1 with the top 16-bit Q15 content of 32-bit chunks in Rs2. At the same time, multiply the bottom 16-bit Q15 content of 32-bit chunks in Rs1 with the bottom 16-bit Q15 content of 32-bit chunks in Rs2. For the KHMX16 instruction, multiply the top 16-bit Q15 content of 32-bit chunks in Rs1 with the bottom 16-bit Q15 content of 32-bit chunks in Rs2. At the same time, multiply the bottom 16-bit Q15 content of 32-bit chunks in Rs1 with the top 16-bit Q15 content of 32-bit chunks in Rs2. The Q30 results are then right-shifted 15-bits and saturated into Q15 values. The Q15 results are then written into Rd. When both the two Q15 inputs of a multiplication are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

**Operations:**

```
if (is `KHM16`) {
    op1t = Rs1.H[x+1]; op2t = Rs2.H[x+1]; // top
    op1b = Rs1.H[x]; op2b = Rs2.H[x]; // bottom
} else if (is `KHMX16`) {
    op1t = Rs1.H[x+1]; op2t = Rs2.H[x]; // Rs1 top
    op1b = Rs1.H[x]; op2b = Rs2.H[x+1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    if (0x8000 != aop | 0x8000 != bop) {
        res = (aop s* bop) >> 15;
    } else {
        res= 0x7FFF;
        OV = 1;
    }
}
```

(continues on next page)

(continued from previous page)

```

    }
}
Rd.W[x/2] = concat(rest, resb);
for RV32: x=0
for RV64: x=0,2

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KHMX16(unsigned long a, unsigned long b)**  
KHMX16 (SIMD Signed Saturating Crossed Q15 Multiply)

**Type:** SIMD

**Syntax:**

```

KHMX16 Rd, Rs1, Rs2
KHMX16 Rd, Rs1, Rs2

```

**Purpose:** Do Q15xQ15 element multiplications simultaneously. The Q30 results are then reduced to Q15 numbers again.

**Description:** For the KHMX16 instruction, multiply the top 16-bit Q15 content of 32-bit chunks in Rs1 with the top 16-bit Q15 content of 32-bit chunks in Rs2. At the same time, multiply the bottom 16-bit Q15 content of 32-bit chunks in Rs1 with the bottom 16-bit Q15 content of 32-bit chunks in Rs2. For the KHMX16 instruction, multiply the top 16-bit Q15 content of 32-bit chunks in Rs1 with the bottom 16-bit Q15 content of 32-bit chunks in Rs2. At the same time, multiply the bottom 16-bit Q15 content of 32-bit chunks in Rs1 with the top 16-bit Q15 content of 32-bit chunks in Rs2. The Q30 results are then right-shifted 15-bits and saturated into Q15 values. The Q15 results are then written into Rd. When both the two Q15 inputs of a multiplication are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

**Operations:**

```

if (is `KHMX16`) {
    op1t = Rs1.H[x+1]; op2t = Rs2.H[x+1]; // top
    op1b = Rs1.H[x]; op2b = Rs2.H[x]; // bottom
} else if (is `KHMX16`) {
    op1t = Rs1.H[x+1]; op2t = Rs2.H[x]; // Rs1 top
    op1b = Rs1.H[x]; op2b = Rs2.H[x+1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    if (0x8000 != aop | 0x8000 != bop) {
        res = (aop s* bop) >> 15;
    } else {
        res= 0x7FFF;
        OV = 1;
    }
}
Rd.W[x/2] = concat(rest, resb);
for RV32: x=0
for RV64: x=0,2

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_SMUL16(unsigned int a, unsigned int b)**  
SMUL16 (SIMD Signed 16-bit Multiply)

**Type:** SIMD

**Syntax:**

```
SMUL16 Rd, Rs1, Rs2
SMULX16 Rd, Rs1, Rs2
```

**Purpose:** Do signed 16-bit multiplications and generate two 32-bit results simultaneously.

**RV32 Description:** For the SMUL16 instruction, multiply the top 16-bit Q15 content of Rs1 with the top 16-bit Q15 content of Rs2. At the same time, multiply the bottom 16-bit Q15 content of Rs1 with the bottom 16-bit Q15 content of Rs2. For the SMULX16 instruction, multiply the top 16-bit Q15 content of Rs1 with the bottom 16-bit Q15 content of Rs2. At the same time, multiply the bottom 16-bit Q15 content of Rs1 with the top 16-bit Q15 content of Rs2. The two Q30 results are then written into an even/odd pair of registers specified by Rd(4,1). Rd(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the 32-bit result calculated from the top part of Rs1 and the even 2d register of the pair contains the 32-bit result calculated from the bottom part of Rs1.

**RV64 Description:** For the SMUL16 instruction, multiply the top 16-bit Q15 content of the lower 32-bit word in Rs1 with the top 16-bit Q15 content of the lower 32-bit word in Rs2. At the same time, multiply the bottom 16-bit Q15 content of the lower 32-bit word in Rs1 with the bottom 16-bit Q15 content of the lower 32-bit word in Rs2. For the SMULX16 instruction, multiply the top 16-bit Q15 content of the lower 32-bit word in Rs1 with the bottom 16-bit Q15 content of the lower 32-bit word in Rs2. At the same time, multiply the bottom 16-bit Q15 content of the lower 32-bit word in Rs1 with the top 16-bit Q15 content of the lower 32-bit word in Rs2. The two 32-bit Q30 results are then written into Rd. The result calculated from the top 16-bit of the lower 32-bit word in Rs1 is written to Rd.W[1]. And the result calculated from the bottom 16-bit of the lower 32-bit word in Rs1 is written to Rd.W[0]

**Operations:**

```
* RV32:
if (is `SMUL16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[1]; // top
    op1b = Rs1.H[0]; op2b = Rs2.H[0]; // bottom
} else if (is `SMULX16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[0]; // Rs1 top
    op1b = Rs1.H[0]; op2b = Rs2.H[1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    res = aop s* bop;
}
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H] = rest;
R[t_L] = resb;
* RV64:
if (is `SMUL16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[1]; // top
    op1b = Rs1.H[0]; op2b = Rs2.H[0]; // bottom
```

(continues on next page)



(continued from previous page)

```

} else if (is `SMULX16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[0]; // Rs1 top
    op1b = Rs1.H[0]; op2b = Rs2.H[1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    res = aop s* bop;
}
Rd.W[1] = rest;
Rd.W[0] = resb;

```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_SMULX16(unsigned int a, unsigned int b)**  
 SMULX16 (SIMD Signed Crossed 16-bit Multiply)

**Type:** SIMD

#### Syntax:

```

SMUL16 Rd, Rs1, Rs2
SMULX16 Rd, Rs1, Rs2

```

**Purpose:** Do signed 16-bit multiplications and generate two 32-bit results simultaneously.

**RV32 Description:** For the SMUL16 instruction, multiply the top 16-bit Q15 content of Rs1 with the top 16-bit Q15 content of Rs2. At the same time, multiply the bottom 16-bit Q15 content of Rs1 with the bottom 16-bit Q15 content of Rs2. For the SMULX16 instruction, multiply the top 16-bit Q15 content of Rs1 with the bottom 16-bit Q15 content of Rs2. At the same time, multiply the bottom 16-bit Q15 content of Rs1 with the top 16-bit Q15 content of Rs2. The two Q30 results are then written into an even/odd pair of registers specified by Rd(4,1). Rd(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the 32-bit result calculated from the top part of Rs1 and the even 2d register of the pair contains the 32-bit result calculated from the bottom part of Rs1.

**RV64 Description:** For the SMUL16 instruction, multiply the top 16-bit Q15 content of the lower 32-bit word in Rs1 with the top 16-bit Q15 content of the lower 32-bit word in Rs2. At the same time, multiply the bottom 16-bit Q15 content of the lower 32-bit word in Rs1 with the bottom 16-bit Q15 content of the lower 32-bit word in Rs2. For the SMULX16 instruction, multiply the top 16-bit Q15 content of the lower 32-bit word in Rs1 with the bottom 16-bit Q15 content of the lower 32-bit word in Rs2. At the same time, multiply the bottom 16-bit Q15 content of the lower 32-bit word in Rs1 with the top 16-bit Q15 content of the lower 32-bit word in Rs2. The two 32-bit Q30 results are then written into Rd. The result calculated from the top 16-bit of the lower 32-bit word in Rs1 is written to Rd.W[1]. And the result calculated from the bottom 16-bit of the lower 32-bit word in Rs1 is written to Rd.W[0]

#### Operations:

```

* RV32:
if (is `SMUL16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[1]; // top
    op1b = Rs1.H[0]; op2b = Rs2.H[0]; // bottom
} else if (is `SMULX16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[0]; // Rs1 top

```

(continues on next page)

(continued from previous page)

```

    op1b = Rs1.H[0]; op2b = Rs2.H[1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    res = aop s* bop;
}
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H] = rest;
R[t_L] = resb;
* RV64:
if (is `SMUL16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[1]; // top
    op1b = Rs1.H[0]; op2b = Rs2.H[0]; // bottom
} else if (is `SMULX16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[0]; // Rs1 top
    op1b = Rs1.H[0]; op2b = Rs2.H[1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    res = aop s* bop;
}
Rd.W[1] = rest;
Rd.W[0] = resb;

```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UMUL16(unsigned int a, unsigned int b)**  
 UMUL16 (SIMD Unsigned 16-bit Multiply)

**Type:** SIMD

**Syntax:**

```

UMUL16 Rd, Rs1, Rs2
UMULX16 Rd, Rs1, Rs2

```

**Purpose:** Do unsigned 16-bit multiplications and generate two 32-bit results simultaneously.

**RV32 Description:** For the UMUL16 instruction, multiply the top 16-bit U16 content of Rs1 with the top 16-bit U16 content of Rs2. At the same time, multiply the bottom 16-bit U16 content of Rs1 with the bottom 16-bit U16 content of Rs2. For the UMULX16 instruction, multiply the top 16-bit U16 content of Rs1 with the bottom 16-bit U16 content of Rs2. At the same time, multiply the bottom 16-bit U16 content of Rs1 with the top 16-bit U16 content of Rs2. The two U32 results are then written into an even/odd pair of registers specified by Rd(4,1). Rd(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the 32-bit result calculated from the top part of Rs1 and the even 2d register of the pair contains the 32-bit result calculated from the bottom part of Rs1.

**RV64 Description:** For the UMUL16 instruction, multiply the top 16-bit U16 content of the lower 32-bit word in Rs1 with the top 16-bit U16 content of the lower 32-bit word in Rs2. At the same time, multiply the bottom 16-bit U16 content of the lower 32-bit word in Rs1 with the bottom 16-bit U16 content of the lower 32-bit word in Rs2. For the UMULX16 instruction, multiply the top 16-bit U16 content of the lower 32-bit word in Rs1 with the bottom 16-bit U16 content of the lower 32-bit word in Rs2. At the same time, multiply the bottom 16-bit U16 content of the lower 32-bit word in Rs1 with the top 16-bit U16 content of

the lower 32-bit word in Rs2. The two 32-bit U32 results are then written into Rd. The result calculated from the top 16-bit of the lower 32-bit word in Rs1 is written to Rd.W[1]. And the result calculated from the bottom 16-bit of the lower 32-bit word in Rs1 is written to Rd.W[0]

#### Operations:

```
* RV32:
if (is `UMUL16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[1]; // top
    op1b = Rs1.H[0]; op2b = Rs2.H[0]; // bottom
} else if (is `UMULX16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[0]; // Rs1 top
    op1b = Rs1.H[0]; op2b = Rs2.H[1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    res = aop u* bop;
}
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H] = rest;
R[t_L] = resb;
* RV64:
if (is `UMUL16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[1]; // top
    op1b = Rs1.H[0]; op2b = Rs2.H[0]; // bottom
} else if (is `UMULX16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[0]; // Rs1 top
    op1b = Rs1.H[0]; op2b = Rs2.H[1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    res = aop u* bop;
}
Rd.W[1] = rest;
Rd.W[0] = resb;
```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UMULX16(unsigned int a, unsigned int b)**  
UMULX16 (SIMD Unsigned Crossed 16-bit Multiply)

**Type:** SIMD

#### Syntax:

```
UMUL16 Rd, Rs1, Rs2
UMULX16 Rd, Rs1, Rs2
```

**Purpose:** Do unsigned 16-bit multiplications and generate two 32-bit results simultaneously.

**RV32 Description:** For the UMUL16 instruction, multiply the top 16-bit U16 content of Rs1 with the top 16-bit U16 content of Rs2. At the same time, multiply the bottom 16-bit U16 content of Rs1 with the bottom 16-bit U16 content of Rs2. For the UMULX16 instruction, multiply the top 16-bit U16 content of Rs1 with the bottom 16-bit U16 content of Rs2. At the same time, multiply the bottom 16-bit U16 content of Rs1 with the top 16-bit U16 content of Rs2. The two U32 results are then written into an even/odd pair of registers specified by Rd(4,1). Rd(4,1), i.e., d, determines the even/odd pair group of two registers.

Specifically, the register pair includes register  $2d$  and  $2d+1$ . The odd  $2d+1$  register of the pair contains the 32-bit result calculated from the top part of  $Rs1$  and the even  $2d$  register of the pair contains the 32-bit result calculated from the bottom part of  $Rs1$ .

**RV64 Description:** For the `UMUL16` instruction, multiply the top 16-bit U16 content of the lower 32-bit word in  $Rs1$  with the top 16-bit U16 content of the lower 32-bit word in  $Rs2$ . At the same time, multiply the bottom 16-bit U16 content of the lower 32-bit word in  $Rs1$  with the bottom 16-bit U16 content of the lower 32-bit word in  $Rs2$ . For the `UMULX16` instruction, multiply the top 16-bit U16 content of the lower 32-bit word in  $Rs1$  with the bottom 16-bit U16 content of the lower 32-bit word in  $Rs2$ . At the same time, multiply the bottom 16-bit U16 content of the lower 32-bit word in  $Rs1$  with the top 16-bit U16 content of the lower 32-bit word in  $Rs2$ . The two 32-bit U32 results are then written into  $Rd$ . The result calculated from the top 16-bit of the lower 32-bit word in  $Rs1$  is written to  $Rd.W[1]$ . And the result calculated from the bottom 16-bit of the lower 32-bit word in  $Rs1$  is written to  $Rd.W[0]$

#### Operations:

```
* RV32:
if (is `UMUL16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[1]; // top
    op1b = Rs1.H[0]; op2b = Rs2.H[0]; // bottom
} else if (is `UMULX16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[0]; // Rs1 top
    op1b = Rs1.H[0]; op2b = Rs2.H[1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    res = aop u* bop;
}
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H] = rest;
R[t_L] = resb;
* RV64:
if (is `UMUL16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[1]; // top
    op1b = Rs1.H[0]; op2b = Rs2.H[0]; // bottom
} else if (is `UMULX16`) {
    op1t = Rs1.H[1]; op2t = Rs2.H[0]; // Rs1 top
    op1b = Rs1.H[0]; op2b = Rs2.H[1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    res = aop u* bop;
}
Rd.W[1] = rest;
Rd.W[0] = resb;
```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

### SIMD 8-bit Multiply Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_KHM8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KHMX8(unsigned long a, unsigned long b)
```

```

__STATIC_FORCEINLINE unsigned long long __RV_SMUL8(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long long __RV_SMULX8(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long long __RV_UMUL8(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long long __RV_UMULX8(unsigned int a, unsigned int b)

```

group **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_8B\_MULTIPLY**

SIMD 8-bit Multiply Instructions.

there are 6 SIMD 8-bit Multiply instructions.

## Functions

```

__STATIC_FORCEINLINE unsigned long __RV_KHM8(unsigned long a, unsigned long b)
    KHM8 (SIMD Signed Saturating Q7 Multiply)

```

**Type:** SIMD

**Syntax:**

```

KHM8 Rd, Rs1, Rs2
KHMx8 Rd, Rs1, Rs2

```

**Purpose:** Do Q7xQ7 element multiplications simultaneously. The Q14 results are then reduced to Q7 numbers again.

**Description:** For the KHM8 instruction, multiply the top 8-bit Q7 content of 16-bit chunks in Rs1 with the top 8-bit Q7 content of 16-bit chunks in Rs2. At the same time, multiply the bottom 8-bit Q7 content of 16-bit chunks in Rs1 with the bottom 8-bit Q7 content of 16-bit chunks in Rs2. For the KHMx16 instruction, multiply the top 8-bit Q7 content of 16-bit chunks in Rs1 with the bottom 8-bit Q7 content of 16-bit chunks in Rs2. At the same time, multiply the bottom 8-bit Q7 content of 16-bit chunks in Rs1 with the top 8-bit Q7 content of 16-bit chunks in Rs2. The Q14 results are then right-shifted 7-bits and saturated into Q7 values. The Q7 results are then written into Rd. When both the two Q7 inputs of a multiplication are 0x80, saturation will happen. The result will be saturated to 0x7F and the overflow flag OV will be set.

**Operations:**

```

if (is `KHM8`) {
    op1t = Rs1.B[x+1]; op2t = Rs2.B[x+1]; // top
    op1b = Rs1.B[x]; op2b = Rs2.B[x]; // bottom
} else if (is `KHMx8`) {
    op1t = Rs1.H[x+1]; op2t = Rs2.H[x]; // Rs1 top
    op1b = Rs1.H[x]; op2b = Rs2.H[x+1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    if (0x80 != aop | 0x80 != bop) {
        res = (aop s* bop) >> 7;
    } else {
        res= 0x7F;
        OV = 1;
    }
}
Rd.H[x/2] = concat(rest, resb);
for RV32, x=0,2
for RV64, x=0,2,4,6

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KHM8(unsigned long a, unsigned long b)**  
KHM8 (SIMD Signed Saturating Crossed Q7 Multiply)

**Type:** SIMD

**Syntax:**

```
KHM8 Rd, Rs1, Rs2
KHM8 Rd, Rs1, Rs2
```

**Purpose:** Do Q7xQ7 element multiplications simultaneously. The Q14 results are then reduced to Q7 numbers again.

**Description:** For the KHM8 instruction, multiply the top 8-bit Q7 content of 16-bit chunks in Rs1 with the top 8-bit Q7 content of 16-bit chunks in Rs2. At the same time, multiply the bottom 8-bit Q7 content of 16-bit chunks in Rs1 with the bottom 8-bit Q7 content of 16-bit chunks in Rs2. For the KHM8.16 instruction, multiply the top 8-bit Q7 content of 16-bit chunks in Rs1 with the bottom 8-bit Q7 content of 16-bit chunks in Rs2. At the same time, multiply the bottom 8-bit Q7 content of 16-bit chunks in Rs1 with the top 8-bit Q7 content of 16-bit chunks in Rs2. The Q14 results are then right-shifted 7-bits and saturated into Q7 values. The Q7 results are then written into Rd. When both the two Q7 inputs of a multiplication are 0x80, saturation will happen. The result will be saturated to 0x7F and the overflow flag OV will be set.

**Operations:**

```
if (is `KHM8`) {
    op1t = Rs1.B[x+1]; op2t = Rs2.B[x+1]; // top
    op1b = Rs1.B[x]; op2b = Rs2.B[x]; // bottom
} else if (is `KHM8.16`) {
    op1t = Rs1.H[x+1]; op2t = Rs2.H[x]; // Rs1 top
    op1b = Rs1.H[x]; op2b = Rs2.H[x+1]; // Rs1 bottom
}
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    if (0x80 != aop | 0x80 != bop) {
        res = (aop s* bop) >> 7;
    } else {
        res= 0x7F;
        OV = 1;
    }
}
Rd.H[x/2] = concat(rest, resb);
for RV32, x=0,2
for RV64, x=0,2,4,6
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_SMUL8(unsigned int a, unsigned int b)**  
SMUL8 (SIMD Signed 8-bit Multiply)

**Type:** SIMD

**Syntax:**

```
SMUL8 Rd, Rs1, Rs2
SMULX8 Rd, Rs1, Rs2
```

**Purpose:** Do signed 8-bit multiplications and generate four 16-bit results simultaneously.

**RV32 Description:** For the SMUL8 instruction, multiply the 8-bit data elements of Rs1 with the corresponding 8-bit data elements of Rs2. For the SMULX8 instruction, multiply the first and second 8-bit data elements of Rs1 with the second and first 8-bit data elements of Rs2. At the same time, multiply the third and fourth 8-bit data elements of Rs1 with the fourth and third 8-bit data elements of Rs2. The four 16-bit results are then written into an even/odd pair of registers specified by Rd(4,1). Rd(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the two 16-bit results calculated from the top part of Rs1 and the even 2d register of the pair contains the two 16-bit results calculated from the bottom part of Rs1.

**RV64 Description:** For the SMUL8 instruction, multiply the 8-bit data elements of Rs1 with the corresponding 8-bit data elements of Rs2. For the SMULX8 instruction, multiply the first and second 8-bit data elements of Rs1 with the second and first 8-bit data elements of Rs2. At the same time, multiply the third and fourth 8-bit data elements of Rs1 with the fourth and third 8-bit data elements of Rs2. The four 16-bit results are then written into Rd. The Rd.W[1] contains the two 16-bit results calculated from the top part of Rs1 and the Rd.W[0] contains the two 16-bit results calculated from the bottom part of Rs1.

**Operations:**

```
* RV32:
if (is `SMUL8`) {
    oplt[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x+1]; // top
    oplb[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x]; // bottom
} else if (is `SMULX8`) {
    oplt[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x]; // Rs1 top
    oplb[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x+1]; // Rs1 bottom
}
rest[x/2] = oplt[x/2] s* op2t[x/2];
resb[x/2] = oplb[x/2] s* op2b[x/2];
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H].H[1] = rest[1]; R[t_H].H[0] = resb[1];
R[t_L].H[1] = rest[0]; R[t_L].H[0] = resb[0];
x = 0 and 2

* RV64:
if (is `SMUL8`) {
    oplt[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x+1]; // top
    oplb[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x]; // bottom
} else if (is `SMULX8`) {
    oplt[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x]; // Rs1 top
    oplb[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x+1]; // Rs1 bottom
}
rest[x/2] = oplt[x/2] s* op2t[x/2];
resb[x/2] = oplb[x/2] s* op2b[x/2];
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
Rd.W[1].H[1] = rest[1]; Rd.W[1].H[0] = resb[1];
Rd.W[0].H[1] = rest[0]; Rd.W[0].H[0] = resb[0];
x = 0 and 2
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned int type of value stored in a

- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_SMULX8(unsigned int a, unsigned int b)**  
SMULX8 (SIMD Signed Crossed 8-bit Multiply)

**Type:** SIMD

**Syntax:**

```
SMUL8 Rd, Rs1, Rs2
SMULX8 Rd, Rs1, Rs2
```

**Purpose:** Do signed 8-bit multiplications and generate four 16-bit results simultaneously.

**RV32 Description:** For the SMUL8 instruction, multiply the 8-bit data elements of Rs1 with the corresponding 8-bit data elements of Rs2. For the SMULX8 instruction, multiply the first and second 8-bit data elements of Rs1 with the second and first 8-bit data elements of Rs2. At the same time, multiply the third and fourth 8-bit data elements of Rs1 with the fourth and third 8-bit data elements of Rs2. The four 16-bit results are then written into an even/odd pair of registers specified by Rd(4,1). Rd(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the two 16-bit results calculated from the top part of Rs1 and the even 2d register of the pair contains the two 16-bit results calculated from the bottom part of Rs1.

**RV64 Description:** For the SMUL8 instruction, multiply the 8-bit data elements of Rs1 with the corresponding 8-bit data elements of Rs2. For the SMULX8 instruction, multiply the first and second 8-bit data elements of Rs1 with the second and first 8-bit data elements of Rs2. At the same time, multiply the third and fourth 8-bit data elements of Rs1 with the fourth and third 8-bit data elements of Rs2. The four 16-bit results are then written into Rd. The Rd.W[1] contains the two 16-bit results calculated from the top part of Rs1 and the Rd.W[0] contains the two 16-bit results calculated from the bottom part of Rs1.

**Operations:**

```
* RV32:
if (is `SMUL8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x+1]; // top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x]; // bottom
} else if (is `SMULX8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x]; // Rs1 top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x+1]; // Rs1 bottom
}
rest[x/2] = op1t[x/2] s* op2t[x/2];
resb[x/2] = op1b[x/2] s* op2b[x/2];
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H].H[1] = rest[1]; R[t_H].H[0] = resb[1];
R[t_L].H[1] = rest[0]; R[t_L].H[0] = resb[0];
x = 0 and 2

* RV64:
if (is `SMUL8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x+1]; // top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x]; // bottom
} else if (is `SMULX8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x]; // Rs1 top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x+1]; // Rs1 bottom
}
rest[x/2] = op1t[x/2] s* op2t[x/2];
resb[x/2] = op1b[x/2] s* op2b[x/2];
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
Rd.W[1].H[1] = rest[1]; Rd.W[1].H[0] = resb[1];
```

(continues on next page)



(continued from previous page)

```
Rd.W[0].H[1] = rest[0]; Rd.W[0].H[0] = resb[0];
x = 0 and 2
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UMUL8(unsigned int a, unsigned int b)**  
 UMUL8 (SIMD Unsigned 8-bit Multiply)

**Type:** SIMD

**Syntax:**

```
UMUL8 Rd, Rs1, Rs2
UMULX8 Rd, Rs1, Rs2
```

**Purpose:** Do unsigned 8-bit multiplications and generate four 16-bit results simultaneously.

**RV32 Description:** For the UMUL8 instruction, multiply the unsigned 8-bit data elements of Rs1 with the corresponding unsigned 8-bit data elements of Rs2. For the UMULX8 instruction, multiply the first and second unsigned 8-bit data elements of Rs1 with the second and first unsigned 8-bit data elements of Rs2. At the same time, multiply the third and fourth unsigned 8-bit data elements of Rs1 with the fourth and third unsigned 8-bit data elements of Rs2. The four 16-bit results are then written into an even/odd pair of registers specified by Rd(4,1). Rd(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the two 16-bit results calculated from the top part of Rs1 and the even 2d register of the pair contains the two 16-bit results calculated from the bottom part of Rs1.

**RV64 Description:** For the UMUL8 instruction, multiply the unsigned 8-bit data elements of Rs1 with the corresponding unsigned 8-bit data elements of Rs2. For the UMULX8 instruction, multiply the first and second unsigned 8-bit data elements of Rs1 with the second and first unsigned 8-bit data elements of Rs2. At the same time, multiply the third and fourth unsigned 8-bit data elements of Rs1 with the fourth and third unsigned 8-bit data elements of Rs2. The four 16-bit results are then written into Rd. The Rd.W[1] contains the two 16-bit results calculated from the top part of Rs1 and the Rd.W[0] contains the two 16-bit results calculated from the bottom part of Rs1.

**Operations:**

```
* RV32:
if (is `UMUL8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x+1]; // top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x]; // bottom
} else if (is `UMULX8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x]; // Rs1 top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x+1]; // Rs1 bottom
}
rest[x/2] = op1t[x/2] u* op2t[x/2];
resb[x/2] = op1b[x/2] u* op2b[x/2];
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H].H[1] = rest[1]; R[t_H].H[0] = resb[1];
R[t_L].H[1] = rest[0]; R[t_L].H[0] = resb[0];
x = 0 and 2
* RV64:
```

(continues on next page)

(continued from previous page)

```

if (is `UMUL8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x+1]; // top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x]; // bottom
} else if (is `UMULX8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x]; // Rs1 top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x+1]; // Rs1 bottom
}
rest[x/2] = op1t[x/2] u* op2t[x/2];
resb[x/2] = op1b[x/2] u* op2b[x/2];
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
Rd.W[1].H[1] = rest[1]; Rd.W[1].H[0] = resb[1];
Rd.W[0].H[1] = rest[0]; Rd.W[0].H[0] = resb[0]; x = 0 and 2

```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UMULX8(unsigned int a, unsigned int b)**  
 UMULX8 (SIMD Unsigned Crossed 8-bit Multiply)

**Type:** SIMD

#### Syntax:

```

UMUL8 Rd, Rs1, Rs2
UMULX8 Rd, Rs1, Rs2

```

**Purpose:** Do unsigned 8-bit multiplications and generate four 16-bit results simultaneously.

**RV32 Description:** For the `UMUL8` instruction, multiply the unsigned 8-bit data elements of `Rs1` with the corresponding unsigned 8-bit data elements of `Rs2`. For the `UMULX8` instruction, multiply the first and second unsigned 8-bit data elements of `Rs1` with the second and first unsigned 8-bit data elements of `Rs2`. At the same time, multiply the third and fourth unsigned 8-bit data elements of `Rs1` with the fourth and third unsigned 8-bit data elements of `Rs2`. The four 16-bit results are then written into an even/odd pair of registers specified by `Rd(4,1)`. `Rd(4,1)`, i.e., `d`, determines the even/odd pair group of two registers. Specifically, the register pair includes register `2d` and `2d+1`. The odd `2d+1` register of the pair contains the two 16-bit results calculated from the top part of `Rs1` and the even `2d` register of the pair contains the two 16-bit results calculated from the bottom part of `Rs1`.

**RV64 Description:** For the `UMUL8` instruction, multiply the unsigned 8-bit data elements of `Rs1` with the corresponding unsigned 8-bit data elements of `Rs2`. For the `UMULX8` instruction, multiply the first and second unsigned 8-bit data elements of `Rs1` with the second and first unsigned 8-bit data elements of `Rs2`. At the same time, multiply the third and fourth unsigned 8-bit data elements of `Rs1` with the fourth and third unsigned 8-bit data elements of `Rs2`. The four 16-bit results are then written into `Rd`. The `Rd.W[1]` contains the two 16-bit results calculated from the top part of `Rs1` and the `Rd.W[0]` contains the two 16-bit results calculated from the bottom part of `Rs1`.

#### Operations:

```

* RV32:
if (is `UMUL8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x+1]; // top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x]; // bottom
} else if (is `UMULX8`) {

```

(continues on next page)

(continued from previous page)

```

    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x]; // Rs1 top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x+1]; // Rs1 bottom
}
rest[x/2] = op1t[x/2] u* op2t[x/2];
resb[x/2] = op1b[x/2] u* op2b[x/2];
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H].H[1] = rest[1]; R[t_H].H[0] = resb[1];
R[t_L].H[1] = rest[0]; R[t_L].H[0] = resb[0];
x = 0 and 2
* RV64:
if (is `UMUL8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x+1]; // top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x]; // bottom
} else if (is `UMULX8`) {
    op1t[x/2] = Rs1.B[x+1]; op2t[x/2] = Rs2.B[x]; // Rs1 top
    op1b[x/2] = Rs1.B[x]; op2b[x/2] = Rs2.B[x+1]; // Rs1 bottom
}
rest[x/2] = op1t[x/2] u* op2t[x/2];
resb[x/2] = op1b[x/2] u* op2b[x/2];
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
Rd.W[1].H[1] = rest[1]; Rd.W[1].H[0] = resb[1];
Rd.W[0].H[1] = rest[0]; Rd.W[0].H[0] = resb[0]; x = 0 and 2

```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

## SIMD 16-bit Miscellaneous Instructions

```

__STATIC_FORCEINLINE unsigned long __RV_CLRS16(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_CLO16(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_CLZ16(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_KABS16(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_SMAX16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SMIN16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UMAX16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UMIN16(unsigned long a, unsigned long b)
__RV_SCLIP16(a, b)
__RV_UCLIP16(a, b)

```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_16B\_MISC**

SIMD 16-bit Miscellaneous Instructions.

there are 10 SIMD 16-bit Misc instructions.

## Defines

### **RV\_SCLIP16** (a, b)

SCLIP16 (SIMD 16-bit Signed Clip Value)

**Type:** SIMD

**Syntax:**

```
SCLIP16 Rd, Rs1, imm4u[3:0]
```

**Purpose:** Limit the 16-bit signed integer elements of a register into a signed range simultaneously.

**Description:** This instruction limits the 16-bit signed integer elements stored in Rs1 into a signed integer range between  $2^{\text{imm4u}}-1$  and  $-2^{\text{imm4u}}$ , and writes the limited results to Rd. For example, if imm4u is 3, the 16-bit input values should be saturated between 7 and -8. If saturation is performed, set OV bit to 1.

**Operations:**

```
src = Rs1.H[x];
if (src > (2^imm4u)-1) {
    src = (2^imm4u)-1;
    OV = 1;
} else if (src < -2^imm4u) {
    src = -2^imm4u;
    OV = 1;
}
Rd.H[x] = src
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

### **RV\_UCLIP16** (a, b)

UCLIP16 (SIMD 16-bit Unsigned Clip Value)

**Type:** SIMD

**Syntax:**

```
UCLIP16 Rt, Ra, imm4u
```

**Purpose:** Limit the 16-bit signed elements of a register into an unsigned range simultaneously.

**Description:** This instruction limits the 16-bit signed elements stored in Rs1 into an unsigned integer range between  $2^{\text{imm4u}}-1$  and 0, and writes the limited results to Rd. For example, if imm4u is 3, the 16-bit input values should be saturated between 7 and 0. If saturation is performed, set OV bit to 1.

**Operations:**

```
src = Rs1.H[x];
if (src > (2^imm4u)-1) {
    src = (2^imm4u)-1;
    OV = 1;
}
```

(continues on next page)

(continued from previous page)

```

} else if (src < 0) {
    src = 0;
    OV = 1;
}
Rd.H[x] = src;
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLRS16(unsigned long a)**

CLRS16 (SIMD 16-bit Count Leading Redundant Sign)

**Type:** SIMD

**Syntax:**

```
CLRS16 Rd, Rs1
```

**Purpose:** Count the number of redundant sign bits of the 16-bit elements of a general register.

**Description:** Starting from the bits next to the sign bits of the 16-bit elements of Rs1, this instruction counts the number of redundant sign bits and writes the result to the corresponding 16-bit elements of Rd.

**Operations:**

```

snum[x] = Rs1.H[x];
cnt[x] = 0;
for (i = 14 to 0) {
    if (snum[x](i) == snum[x](15)) {
        cnt[x] = cnt[x] + 1;
    } else {
        break;
    }
}
Rd.H[x] = cnt[x];
for RV32: x=1...0
for RV64: x=3...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLO16(unsigned long a)**

CLO16 (SIMD 16-bit Count Leading One)

**Type:** SIMD

**Syntax:**

```
CLO16 Rd, Rs1
```

**Purpose:** Count the number of leading one bits of the 16-bit elements of a general register.

**Description:** Starting from the most significant bits of the 16-bit elements of Rs1, this instruction counts the number of leading one bits and writes the results to the corresponding 16-bit elements of Rd.

**Operations:**

```
snum[x] = Rs1.H[x];
cnt[x] = 0;
for (i = 15 to 0) {
    if (snum[x](i) == 1) {
        cnt[x] = cnt[x] + 1;
    } else {
        break;
    }
}
Rd.H[x] = cnt[x];
for RV32: x=1...0
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLZ16(unsigned long a)**  
CLZ16 (SIMD 16-bit Count Leading Zero)

**Type:** SIMD

**Syntax:**

```
CLZ16 Rd, Rs1
```

**Purpose:** Count the number of leading zero bits of the 16-bit elements of a general register.

**Description:** Starting from the most significant bits of the 16-bit elements of Rs1, this instruction counts the number of leading zero bits and writes the results to the corresponding 16-bit elements of Rd.

**Operations:**

```
snum[x] = Rs1.H[x];
cnt[x] = 0;
for (i = 15 to 0) {
    if (snum[x](i) == 0) {
        cnt[x] = cnt[x] + 1;
    } else {
        break;
    }
}
Rd.H[x] = cnt[x];
for RV32: x=1...0
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KABS16(unsigned long a)**  
KABS16 (SIMD 16-bit Saturating Absolute)

**Type:** SIMD

**Syntax:**

```
KABS16 Rd, Rs1
```

**Purpose:** Get the absolute value of 16-bit signed integer elements simultaneously.

**Description:** This instruction calculates the absolute value of 16-bit signed integer elements stored in Rs1 and writes the element results to Rd. If the input number is 0x8000, this instruction generates 0x7fff as the output and sets the OV bit to 1.

**Operations:**

```
src = Rs1.H[x];
if (src == 0x8000) {
    src = 0x7fff;
    OV = 1;
} else if (src[15] == 1)
    src = -src;
Rd.H[x] = src;
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SMAX16(unsigned long a, unsigned long b)**  
SMAX16 (SIMD 16-bit Signed Maximum)

**Type:** SIMD

**Syntax:**

```
SMAX16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit signed integer elements finding maximum operations simultaneously.

**Description:** This instruction compares the 16-bit signed integer elements in Rs1 with the 16-bit signed integer elements in Rs2 and selects the numbers that is greater than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.H[x] = (Rs1.H[x] > Rs2.H[x]) ? Rs1.H[x] : Rs2.H[x];
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SMIN16(unsigned long a, unsigned long b)**  
SMIN16 (SIMD 16-bit Signed Minimum)

**Type:** SIMD

**Syntax:**

SMIN16 Rd, Rs1, Rs2
---------------------

**Purpose:** Do 16-bit signed integer elements finding minimum operations simultaneously.

**Description:** This instruction compares the 16-bit signed integer elements in Rs1 with the 16-bit signed integer elements in Rs2 and selects the numbers that is less than the other one. The selected results are written to Rd.

**Operations:**

Rd.H[x] = (Rs1.H[x] < Rs2.H[x]) ? Rs1.H[x] : Rs2.H[x]; for RV32: x=1...0, for RV64: x=3...0
---

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UMAX16(unsigned long a, unsigned long b)**  
UMAX16 (SIMD 16-bit Unsigned Maximum)

**Type:** SIMD

**Syntax:**

UMAX16 Rd, Rs1, Rs2
---------------------

**Purpose:** Do 16-bit unsigned integer elements finding maximum operations simultaneously.

**Description:** This instruction compares the 16-bit unsigned integer elements in Rs1 with the 16-bit unsigned integer elements in Rs2 and selects the numbers that is greater than the other one. The selected results are written to Rd.

**Operations:**

Rd.H[x] = (Rs1.H[x] >u Rs2.H[x]) ? Rs1.H[x] : Rs2.H[x]; for RV32: x=1...0, for RV64: x=3...0
--

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UMIN16(unsigned long a, unsigned long b)**  
UMIN16 (SIMD 16-bit Unsigned Minimum)



**Type:** SIMD

**Syntax:**

```
UMIN16 Rd, Rs1, Rs2
```

**Purpose:** Do 16-bit unsigned integer elements finding minimum operations simultaneously.

**Description:** This instruction compares the 16-bit unsigned integer elements in Rs1 with the 16-bit unsigned integer elements in Rs2 and selects the numbers that is less than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.H[x] = (Rs1.H[x] <u Rs2.H[x]) ? Rs1.H[x] : Rs2.H[x];
for RV32: x=1...0,
for RV64: x=3...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## SIMD 8-bit Miscellaneous Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_CLRS8(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_CLO8(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_CLZ8(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_KABS8(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_SMAX8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SMIN8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UMAX8(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UMIN8(unsigned long a, unsigned long b)
__RV_SCLIP8(a, b)
__RV_UCLIP8(a, b)
```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_8B\_MISC**

SIMD 8-bit Miscellaneous Instructions.

there are 10 SIMD 8-bit Miscellaneous instructions.

## Defines

```
__RV_SCLIP8(a, b)
SCLIP8 (SIMD 8-bit Signed Clip Value)
```

**Type:** SIMD

**Syntax:**

```
SCLIP8 Rd, Rs1, imm3u[2:0]
```

**Purpose:** Limit the 8-bit signed integer elements of a register into a signed range simultaneously.

**Description:** This instruction limits the 8-bit signed integer elements stored in Rs1 into a signed integer range between  $2^{\text{imm3u}}-1$  and  $-2^{\text{imm3u}}$ , and writes the limited results to Rd. For example, if imm3u is 3, the 8-bit input values should be saturated between 7 and -8. If saturation is performed, set OV bit to 1.

**Operations:**

```
src = Rs1.B[x];
if (src > (2^imm3u)-1) {
    src = (2^imm3u)-1;
    OV = 1;
} else if (src < -2^imm3u) {
    src = -2^imm3u;
    OV = 1;
}
Rd.B[x] = src
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

\_\_RV\_UCLIP8 (a, b)

UCLIP8 (SIMD 8-bit Unsigned Clip Value)

**Type:** SIMD

**Syntax:**

```
UCLIP8 Rt, Ra, imm3u
```

**Purpose:** Limit the 8-bit signed elements of a register into an unsigned range simultaneously.

**Description:** This instruction limits the 8-bit signed elements stored in Rs1 into an unsigned integer range between  $2^{\text{imm3u}}-1$  and 0, and writes the limited results to Rd. For example, if imm3u is 3, the 8-bit input values should be saturated between 7 and 0. If saturation is performed, set OV bit to 1.

**Operations:**

```
src = Rs1.H[x];
if (src > (2^imm3u)-1) {
    src = (2^imm3u)-1;
    OV = 1;
} else if (src < 0) {
    src = 0;
    OV = 1;
}
Rd.H[x] = src;
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**Functions**

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLRS8(unsigned long a)**  
CLRS8 (SIMD 8-bit Count Leading Redundant Sign)

**Type:** SIMD

**Syntax:**

```
CLRS8 Rd, Rs1
```

**Purpose:** Count the number of redundant sign bits of the 8-bit elements of a general register.

**Description:** Starting from the bits next to the sign bits of the 8-bit elements of Rs1, this instruction counts the number of redundant sign bits and writes the result to the corresponding 8-bit elements of Rd.

**Operations:**

```
snum[x] = Rs1.B[x];
cnt[x] = 0;
for (i = 6 to 0) {
    if (snum[x](i) == snum[x](7)) {
        cnt[x] = cnt[x] + 1;
    } else {
        break;
    }
}
Rd.B[x] = cnt[x];
for RV32: x=3...0
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLO8(unsigned long a)**  
CLO8 (SIMD 8-bit Count Leading One)

**Type:** SIMD

**Syntax:**

```
CLO8 Rd, Rs1
```

**Purpose:** Count the number of leading one bits of the 8-bit elements of a general register.

**Description:** Starting from the most significant bits of the 8-bit elements of Rs1, this instruction counts the number of leading one bits and writes the results to the corresponding 8-bit elements of Rd.

**Operations:**

```

snum[x] = Rs1.B[x];
cnt[x] = 0;
    for (i = 7 to 0) {
        if (snum[x](i) == 1) {
            cnt[x] = cnt[x] + 1;
        } else {
            break;
        }
    }
Rd.B[x] = cnt[x];
for RV32: x=3...0
for RV64: x=7...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLZ8(unsigned long a)**  
CLZ8 (SIMD 8-bit Count Leading Zero)

**Type:** SIMD

**Syntax:**

```
CLZ8 Rd, Rs1
```

**Purpose:** Count the number of leading zero bits of the 8-bit elements of a general register.

**Description:** Starting from the most significant bits of the 8-bit elements of Rs1, this instruction counts the number of leading zero bits and writes the results to the corresponding 8-bit elements of Rd.

**Operations:**

```

snum[x] = Rs1.B[x];
cnt[x] = 0;
    for (i = 7 to 0) {
        if (snum[x](i) == 0) {
            cnt[x] = cnt[x] + 1;
        } else {
            break;
        }
    }
Rd.B[x] = cnt[x];
for RV32: x=3...0
for RV64: x=7...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KABS8(unsigned long a)**  
KABS8 (SIMD 8-bit Saturating Absolute)

**Type:** SIMD

**Syntax:**

KABS8 Rd, Rs1

**Purpose:** Get the absolute value of 8-bit signed integer elements simultaneously.

**Description:** This instruction calculates the absolute value of 8-bit signed integer elements stored in Rs1 and writes the element results to Rd. If the input number is 0x80, this instruction generates 0x7f as the output and sets the OV bit to 1.

**Operations:**

```
src = Rs1.B[x];
if (src == 0x80) {
    src = 0x7f;
    OV = 1;
} else if (src[7] == 1)
    src = -src;
}
Rd.B[x] = src;
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SMAX8(unsigned long a, unsigned long b)**  
SMAX8 (SIMD 8-bit Signed Maximum)

**Type:** SIMD

**Syntax:**

SMAX8 Rd, Rs1, Rs2

**Purpose:** Do 8-bit signed integer elements finding maximum operations simultaneously.

**Description:** This instruction compares the 8-bit signed integer elements in Rs1 with the 8-bit signed integer elements in Rs2 and selects the numbers that is greater than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.B[x] = (Rs1.B[x] > Rs2.B[x]) ? Rs1.B[x] : Rs2.B[x];
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SMIN8(unsigned long a, unsigned long b)**  
SMIN8 (SIMD 8-bit Signed Minimum)

**Type:** SIMD

**Syntax:**

SMIN8 Rd, Rs1, Rs2

**Purpose:** Do 8-bit signed integer elements finding minimum operations simultaneously.

**Description:** This instruction compares the 8-bit signed integer elements in Rs1 with the 8-bit signed integer elements in Rs2 and selects the numbers that is less than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.B[x] = (Rs1.B[x] < Rs2.B[x]) ? Rs1.B[x] : Rs2.B[x];  
for RV32: x=3...0,  
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UMAX8(unsigned long a, unsigned long b)**  
UMAX8 (SIMD 8-bit Unsigned Maximum)

**Type:** SIMD

**Syntax:**

UMAX8 Rd, Rs1, Rs2

**Purpose:** Do 8-bit unsigned integer elements finding maximum operations simultaneously.

**Description:** This instruction compares the 8-bit unsigned integer elements in Rs1 with the four 8-bit unsigned integer elements in Rs2 and selects the numbers that is greater than the other one. The two selected results are written to Rd.

**Operations:**

```
Rd.B[x] = (Rs1.B[x] >u Rs2.B[x]) ? Rs1.B[x] : Rs2.B[x];  
for RV32: x=3...0,  
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UMIN8(unsigned long a, unsigned long b)**  
UMIN8 (SIMD 8-bit Unsigned Minimum)

**Type:** SIMD

**Syntax:**

UMIN8 Rd, Rs1, Rs2

**Purpose:** Do 8-bit unsigned integer elements finding minimum operations simultaneously.

**Description:** This instruction compares the 8-bit unsigned integer elements in Rs1 with the 8-bit unsigned integer elements in Rs2 and selects the numbers that is less than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.B[x] = (Rs1.B[x] <u Rs2.B[x]) ? Rs1.B[x] : Rs2.B[x];
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## SIMD 8-bit Unpacking Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_SUNPKD810(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_SUNPKD820(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_SUNPKD830(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_SUNPKD831(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_SUNPKD832(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_ZUNPKD810(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_ZUNPKD820(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_ZUNPKD830(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_ZUNPKD831(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_ZUNPKD832(unsigned long a)
```

group NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_8B\_UNPACK

SIMD 8-bit Unpacking Instructions.

there are 8 SIMD 8-bit Unpacking instructions.

## Functions

```
__STATIC_FORCEINLINE unsigned long __RV_SUNPKD810(unsigned long a)
SUNPKD810 (Signed Unpacking Bytes 1 & 0)
```

**Type:** DSP

**Syntax:**

```
SUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}
```

**Purpose:** Unpack byte *x* and byte *y* of 32-bit chunks in a register into two 16-bit signed halfwords of 32-bit chunks in a register.

**Description:** For the SUNPKD8 (*x*) (*\*y\**) instruction, it unpacks byte *x* and byte *y* of 32-bit chunks in Rs1 into two 16-bit signed halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```
Rd.W[m].H[1] = SE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = SE16(Rs1.W[m].B[y])
// SUNPKD810, x=1,y=0
// SUNPKD820, x=2,y=0
// SUNPKD830, x=3,y=0
// SUNPKD831, x=3,y=1
// SUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SUNPKD820(unsigned long a)**  
SUNPKD820 (Signed Unpacking Bytes 2 & 0)

**Type:** DSP

**Syntax:**

```
SUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}
```

**Purpose:** Unpack byte *x* and byte *y* of 32-bit chunks in a register into two 16-bit signed halfwords of 32-bit chunks in a register.

**Description:** For the SUNPKD8 (*x*) (*\*y\**) instruction, it unpacks byte *x* and byte *y* of 32-bit chunks in Rs1 into two 16-bit signed halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```
Rd.W[m].H[1] = SE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = SE16(Rs1.W[m].B[y])
// SUNPKD810, x=1,y=0
// SUNPKD820, x=2,y=0
// SUNPKD830, x=3,y=0
// SUNPKD831, x=3,y=1
// SUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a



**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SUNPKD830(unsigned long a)**  
 SUNPKD830 (Signed Unpacking Bytes 3 & 0)

**Type:** DSP

**Syntax:**

```
SUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}
```

**Purpose:** Unpack byte *x* and byte *y* of 32-bit chunks in a register into two 16-bit signed halfwords of 32-bit chunks in a register.

**Description:** For the SUNPKD8 (*x*) (*\*y\**) instruction, it unpacks byte *x* and byte *y* of 32-bit chunks in Rs1 into two 16-bit signed halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```
Rd.W[m].H[1] = SE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = SE16(Rs1.W[m].B[y])
// SUNPKD810, x=1,y=0
// SUNPKD820, x=2,y=0
// SUNPKD830, x=3,y=0
// SUNPKD831, x=3,y=1
// SUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SUNPKD831(unsigned long a)**  
 SUNPKD831 (Signed Unpacking Bytes 3 & 1)

**Type:** DSP

**Syntax:**

```
SUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}
```

**Purpose:** Unpack byte *x* and byte *y* of 32-bit chunks in a register into two 16-bit signed halfwords of 32-bit chunks in a register.

**Description:** For the SUNPKD8 (*x*) (*\*y\**) instruction, it unpacks byte *x* and byte *y* of 32-bit chunks in Rs1 into two 16-bit signed halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```
Rd.W[m].H[1] = SE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = SE16(Rs1.W[m].B[y])
// SUNPKD810, x=1,y=0
// SUNPKD820, x=2,y=0
// SUNPKD830, x=3,y=0
// SUNPKD831, x=3,y=1
// SUNPKD832, x=3,y=2
```

(continues on next page)

(continued from previous page)

```

for RV32: m=0,
for RV64: m=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SUNPKD832(unsigned long a)**  
 SUNPKD832 (Signed Unpacking Bytes 3 & 2)

**Type:** DSP

**Syntax:**

```

SUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}

```

**Purpose:** Unpack byte *x* and byte *y* of 32-bit chunks in a register into two 16-bit signed halfwords of 32-bit chunks in a register.

**Description:** For the SUNPKD8 (*x*) (*\*y\**) instruction, it unpacks byte *x* and byte *y* of 32-bit chunks in Rs1 into two 16-bit signed halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```

Rd.W[m].H[1] = SE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = SE16(Rs1.W[m].B[y])
// SUNPKD810, x=1,y=0
// SUNPKD820, x=2,y=0
// SUNPKD830, x=3,y=0
// SUNPKD831, x=3,y=1
// SUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_ZUNPKD810(unsigned long a)**  
 ZUNPKD810 (Unsigned Unpacking Bytes 1 & 0)

**Type:** DSP

**Syntax:**

```

ZUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}

```

**Purpose:** Unpack byte *x* and byte *y* of 32-bit chunks in a register into two 16-bit unsigned halfwords of 32-bit chunks in a register.

**Description:** For the ZUNPKD8 (*x*) (*\*y\**) instruction, it unpacks byte *x* and byte *y* of 32-bit chunks in Rs1 into two 16-bit unsigned halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```

Rd.W[m].H[1] = ZE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = ZE16(Rs1.W[m].B[y])
// ZUNPKD810, x=1,y=0
// ZUNPKD820, x=2,y=0
// ZUNPKD830, x=3,y=0
// ZUNPKD831, x=3,y=1
// ZUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_ZUNPKD820(unsigned long a)**  
 ZUNPKD820 (Unsigned Unpacking Bytes 2 & 0)

**Type:** DSP

**Syntax:**

```

ZUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}

```

**Purpose:** Unpack byte x and byte y of 32-bit chunks in a register into two 16-bit unsigned halfwords of 32-bit chunks in a register.

**Description:** For the ZUNPKD8 (x) (\*y\*) instruction, it unpacks byte *x* and byte *y* of 32-bit chunks in Rs1 into two 16-bit unsigned halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```

Rd.W[m].H[1] = ZE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = ZE16(Rs1.W[m].B[y])
// ZUNPKD810, x=1,y=0
// ZUNPKD820, x=2,y=0
// ZUNPKD830, x=3,y=0
// ZUNPKD831, x=3,y=1
// ZUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_ZUNPKD830(unsigned long a)**  
 ZUNPKD830 (Unsigned Unpacking Bytes 3 & 0)

**Type:** DSP

**Syntax:**

```
ZUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}
```

**Purpose:** Unpack byte  $x$  and byte  $y$  of 32-bit chunks in a register into two 16-bit unsigned halfwords of 32-bit chunks in a register.

**Description:** For the ZUNPKD8 ( $x$ ) ( $*y*$ ) instruction, it unpacks byte  $x$  and byte  $y$  of 32-bit chunks in Rs1 into two 16-bit unsigned halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```
Rd.W[m].H[1] = ZE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = ZE16(Rs1.W[m].B[y])
// ZUNPKD810, x=1,y=0
// ZUNPKD820, x=2,y=0
// ZUNPKD830, x=3,y=0
// ZUNPKD831, x=3,y=1
// ZUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

```
__STATIC_FORCEINLINE unsigned long __RV_ZUNPKD831(unsigned long a)
ZUNPKD831 (Unsigned Unpacking Bytes 3 & 1)
```

**Type:** DSP

**Syntax:**

```
ZUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}
```

**Purpose:** Unpack byte  $x$  and byte  $y$  of 32-bit chunks in a register into two 16-bit unsigned halfwords of 32-bit chunks in a register.

**Description:** For the ZUNPKD8 ( $x$ ) ( $*y*$ ) instruction, it unpacks byte  $x$  and byte  $y$  of 32-bit chunks in Rs1 into two 16-bit unsigned halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```
Rd.W[m].H[1] = ZE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = ZE16(Rs1.W[m].B[y])
// ZUNPKD810, x=1,y=0
// ZUNPKD820, x=2,y=0
// ZUNPKD830, x=3,y=0
// ZUNPKD831, x=3,y=1
// ZUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_ZUNPKD832(unsigned long a)**  
 ZUNPKD832 (Unsigned Unpacking Bytes 3 & 2)

**Type:** DSP

**Syntax:**

```
ZUNPKD8xy Rd, Rs1
xy = {10, 20, 30, 31, 32}
```

**Purpose:** Unpack byte x and byte y of 32-bit chunks in a register into two 16-bit unsigned halfwords of 32-bit chunks in a register.

**Description:** For the ZUNPKD8 (x) (\*y\*) instruction, it unpacks byte *x* and byte *y* of 32-bit chunks in Rs1 into two 16-bit unsigned halfwords and writes the results to the top part and the bottom part of 32-bit chunks in Rd.

**Operations:**

```
Rd.W[m].H[1] = ZE16(Rs1.W[m].B[x])
Rd.W[m].H[0] = ZE16(Rs1.W[m].B[y])
// ZUNPKD810, x=1,y=0
// ZUNPKD820, x=2,y=0
// ZUNPKD830, x=3,y=0
// ZUNPKD831, x=3,y=1
// ZUNPKD832, x=3,y=2
for RV32: m=0,
for RV64: m=1..0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

*group* **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_DATA\_PROCESS**  
 SIMD Data Processing Instructions.

## Non-SIMD Instructions

### Non-SIMD Q15 saturation ALU Instructions

```
__STATIC_FORCEINLINE long __RV_KADDH(int a, int b)
__STATIC_FORCEINLINE long __RV_KHMBB(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KHMBT(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KHMTT(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KSUBH(int a, int b)
__STATIC_FORCEINLINE unsigned long __RV_UKADDH(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_UKSUBH(unsigned int a, unsigned int b)
```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_NON\_SIMD\_Q15\_SAT\_ALU**  
 Non-SIMD Q15 saturation ALU Instructions.

there are 7 Non-SIMD Q15 saturation ALU Instructions

## Functions

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KADDH(int a, int b)**  
KADDH (Signed Addition with Q15 Saturation)

**Type:** DSP

**Syntax:**

```
KADDH Rd, Rs1, Rs2
```

**Purpose:** Add the signed lower 32-bit content of two registers with Q15 saturation.

**Description:** The signed lower 32-bit content of Rs1 is added with the signed lower 32-bit content of Rs2. And the result is saturated to the 16-bit signed integer range of  $[-2^{15}, 2^{15}-1]$  and then sign-extended and written to Rd. If saturation happens, this instruction sets the OV flag.

**Operations:**

```
tmp = Rs1.W[0] + Rs2.W[0];
if (tmp > 32767) {
    res = 32767;
    OV = 1;
} else if (tmp < -32768) {
    res = -32768;
    OV = 1;
} else {
    res = tmp;
}
Rd = SE(tmp[15:0]);
```

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KHMBB(unsigned int a, unsigned int b)**  
KHMBB (Signed Saturating Half Multiply B16 x B16)

**Type:** DSP

**Syntax:**

```
KHMxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 number contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then right-shift 15 bits to turn the Q30 result into a Q15 number again and saturate the Q15 result into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then right-shifted 15-bits and saturated into a Q15 value. The Q15 value is then sign-extended and written into Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

**Operations:**

```

aop = Rs1.H[0]; bop = Rs2.H[0]; // KHMBB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KHMBT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KHMTT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult[31:0] = aop * bop;
    res[15:0] = Mresult[30:15];
} else {
    res[15:0] = 0x7FFF;
    OV = 1;
}
Rd = SE32(res[15:0]); // Rv32
Rd = SE64(res[15:0]); // RV64

```

**Return** value stored in long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KHMBT(unsigned int a, unsigned int b)**  
KHMBT (Signed Saturating Half Multiply B16 x T16)

**Type:** DSP

**Syntax:**

```
KHMxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 number contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then right-shift 15 bits to turn the Q30 result into a Q15 number again and saturate the Q15 result into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then right- shifted 15-bits and saturated into a Q15 value. The Q15 value is then sing-extended and written into Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

**Operations:**

```

aop = Rs1.H[0]; bop = Rs2.H[0]; // KHMBB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KHMBT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KHMTT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult[31:0] = aop * bop;
    res[15:0] = Mresult[30:15];
} else {
    res[15:0] = 0x7FFF;
    OV = 1;
}
Rd = SE32(res[15:0]); // Rv32
Rd = SE64(res[15:0]); // RV64

```

**Return** value stored in long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KHMTT(unsigned int a, unsigned int b)**  
KHMTT (Signed Saturating Half Multiply T16 x T16)

**Type:** DSP

**Syntax:**

```
KHMxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 number contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then right-shift 15 bits to turn the Q30 result into a Q15 number again and saturate the Q15 result into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then right- shifted 15-bits and saturated into a Q15 value. The Q15 value is then sing-extended and written into Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

**Operations:**

```
aop = Rs1.H[0]; bop = Rs2.H[0]; // KHMBB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KHMBT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KHMTT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult[31:0] = aop * bop;
    res[15:0] = Mresult[30:15];
} else {
    res[15:0] = 0x7FFF;
    OV = 1;
}
Rd = SE32(res[15:0]); // Rv32
Rd = SE64(res[15:0]); // RV64
```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KSUBH(int a, int b)**  
KSUBH (Signed Subtraction with Q15 Saturation)

**Type:** DSP

**Syntax:**

```
KSUBH Rd, Rs1, Rs2
```

**Purpose:** Subtract the signed lower 32-bit content of two registers with Q15 saturation.

**Description:** The signed lower 32-bit content of Rs2 is subtracted from the signed lower 32-bit content of Rs1. And the result is saturated to the 16-bit signed integer range of  $[-2^{15}, 2^{15}-1]$  and then sign-extended and written to Rd. If saturation happens, this instruction sets the OV flag.



**Operations:**

```

tmp = Rs1.W[0] - Rs2.W[0];
if (tmp > (2^15)-1) {
    res = (2^15)-1;
    OV = 1;
} else if (tmp < -2^15) {
    res = -2^15;
    OV = 1;
} else {
    res = tmp;
}
Rd = SE(res[15:0]);

```

**Return** value stored in long type**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKADDH(unsigned int a, unsigned int b)**  
 UKADDH (Unsigned Addition with U16 Saturation)

**Type:** DSP**Syntax:**

```
UKADDH Rd, Rs1, Rs2
```

**Purpose:** Add the unsigned lower 32-bit content of two registers with U16 saturation.

**Description:** The unsigned lower 32-bit content of Rs1 is added with the unsigned lower 32-bit content of Rs2. And the result is saturated to the 16-bit unsigned integer range of [0, 2^16-1] and then sign-extended and written to Rd. If saturation happens, this instruction sets the OV flag.

**Operations:**

```

tmp = Rs1.W[0] + Rs2.W[0];
if (tmp > (2^16)-1) {
    tmp = (2^16)-1;
    OV = 1;
}
Rd = SE(tmp[15:0]);

```

**Return** value stored in unsigned long type**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSUBH(unsigned int a, unsigned int b)**  
 UKSUBH (Unsigned Subtraction with U16 Saturation)

**Type:** DSP**Syntax:**

UKSUBH Rd, Rs1, Rs2

**Purpose:** Subtract the unsigned lower 32-bit content of two registers with U16 saturation.

**Description:** The unsigned lower 32-bit content of Rs2 is subtracted from the unsigned lower 32-bit content of Rs1. And the result is saturated to the 16-bit unsigned integer range of  $[0, 2^{16}-1]$  and then sign-extended and written to Rd. If saturation happens, this instruction sets the OV flag.

**Operations:**

```
tmp = Rs1.W[0] - Rs2.W[0];
if (tmp > (2^16)-1) {
    tmp = (2^16)-1;
    OV = 1;
}
else if (tmp < 0) {
    tmp = 0;
    OV = 1;
}
Rd = SE(tmp[15:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

## Non-SIMD Q31 saturation ALU Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_KABSW(signed long a)
__STATIC_FORCEINLINE long __RV_KADDW(int a, int b)
__STATIC_FORCEINLINE long __RV_KDMBB(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KDMBT(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KDMTT(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KDMABB(long t, unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KDMABT(long t, unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KDMATT(long t, unsigned int a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KSLLW(long a, unsigned int b)
__STATIC_FORCEINLINE long __RV_KSLRAW(int a, int b)
__STATIC_FORCEINLINE long __RV_KSLRAW_U(int a, int b)
__STATIC_FORCEINLINE long __RV_KSUBW(int a, int b)
__STATIC_FORCEINLINE unsigned long __RV_UKADDW(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_UKSUBW(unsigned int a, unsigned int b)
__RV_KSLLIW(a, b)
```

group **NMSIS\_Core\_DSP\_Intrinsic\_NON\_SIMD\_Q31\_SAT\_ALU**

Non-SIMD Q31 saturation ALU Instructions.

there are Non-SIMD Q31 saturation ALU Instructions

## Defines

**\_\_RV\_KSLLIW**(a, b)

KSLLIW (Saturating Shift Left Logical Immediate for Word)

**Type:** DSP

**Syntax:**

```
KSLLIW Rd, Rs1, imm5u
```

**Purpose:** Do logical left shift operation with saturation on a 32-bit word. The shift amount is an immediate value.

**Description:** The first word data in Rs1 is left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the imm5u constant. Any shifted value greater than  $2^{31}-1$  is saturated to  $2^{31}-1$ . Any shifted value smaller than  $-2^{31}$  is saturated to  $-2^{31}$ . And the saturated result is sign-extended and written to Rd. If any saturation is performed, set OV bit to 1.

**Operations:**

```
sa = imm5u;
res[(31+sa):0] = Rs1.W[0] << sa;
if (res > (2^31)-1) {
    res = 0x7fffffff; OV = 1;
} else if (res < -2^31) {
    res = 0x80000000; OV = 1;
}
Rd[31:0] = res[31:0]; // RV32
Rd[63:0] = SE(res[31:0]); // RV64
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned int type of value stored in b

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KABSW**(signed long a)

KABSW (Scalar 32-bit Absolute Value with Saturation)

**Type:** DSP

**Syntax:**

```
KABSW Rd, Rs1
```

**Purpose:** Get the absolute value of a signed 32-bit integer in a general register.

**Description:** This instruction calculates the absolute value of a signed 32-bit integer stored in Rs1. The result is sign-extended (for RV64) and written to Rd. This instruction with the minimum negative integer

input of 0x80000000 will produce a saturated output of maximum positive integer of 0x7fffffff and the OV flag will be set to 1.

#### Operations:

```
if (Rs1.W[0] >= 0) {
    res = Rs1.W[0];
} else {
    If (Rs1.W[0] == 0x80000000) {
        res = 0x7fffffff;
        OV = 1;
    } else {
        res = -Rs1.W[0];
    }
}
Rd = SE32(res);
```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: signed long type of value stored in a

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KADDW(int a, int b)**

KADDW (Signed Addition with Q31 Saturation)

**Type:** DSP

#### Syntax:

```
KADDW Rd, Rs1, Rs2
```

**Purpose:** Add the lower 32-bit signed content of two registers with Q31 saturation.

**Description:** The lower 32-bit signed content of Rs1 is added with the lower 32-bit signed content of Rs2. And the result is saturated to the 32-bit signed integer range of  $[-2^{31}, 2^{31}-1]$  and then sign-extended and written to Rd. If saturation happens, this instruction sets the OV flag.

#### Operations:

```
tmp = Rs1.W[0] + Rs2.W[0];
if (tmp > (2^31)-1) {
    res = (2^31)-1;
    OV = 1;
} else if (tmp < -2^31) {
    res = -2^31;
    OV = 1;
} else {
    res = tmp;
}
Rd = res[31:0]; // RV32
Rd = SE(res[31:0]) // RV64
```

**Return** value stored in long type

#### Parameters

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KDMBB(unsigned int a, unsigned int b)**  
 KDMBB (Signed Saturating Double Multiply B16 x B16)

**Type:** DSP

**Syntax:**

```
KDMxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then double and saturate the Q31 result. The result is written into the destination register for RV32 or sign-extended to 64-bits and written into the destination register for RV64. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then doubled and saturated into a Q31 value. The Q31 value is then written into Rd (sign-extended in RV64). When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFFFFFF and the overflow flag OV will be set.

**Operations:**

```
aop = Rs1.H[0]; bop = Rs2.H[0]; // KDMBB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KDMBT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KDMTT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult = aop * bop;
    resQ31 = Mresult << 1;
    Rd = resQ31; // RV32
    Rd = SE(resQ31); // RV64
} else {
    resQ31 = 0x7FFFFFFF;
    Rd = resQ31; // RV32
    Rd = SE(resQ31); // RV64
    OV = 1;
}
```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KDMBT(unsigned int a, unsigned int b)**  
 KDMBT (Signed Saturating Double Multiply B16 x T16)

**Type:** DSP

**Syntax:**

```
KDMxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then double and saturate the Q31 result. The result is written into the destination register for RV32 or sign-extended to 64-bits and written into the destination register for RV64. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then doubled and

saturated into a Q31 value. The Q31 value is then written into Rd (sign-extended in RV64). When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFFFFFF and the overflow flag OV will be set.

#### Operations:

```
aop = Rs1.H[0]; bop = Rs2.H[0]; // KDMBB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KDMBT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KDMTT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult = aop * bop;
    resQ31 = Mresult << 1;
    Rd = resQ31; // RV32
    Rd = SE(resQ31); // RV64
} else {
    resQ31 = 0x7FFFFFFF;
    Rd = resQ31; // RV32
    Rd = SE(resQ31); // RV64
    OV = 1;
}
```

**Return** value stored in long type

#### Parameters

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KDMTT(unsigned int a, unsigned int b)**

KDMTT (Signed Saturating Double Multiply T16 x T16)

**Type:** DSP

#### Syntax:

```
KDMxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then double and saturate the Q31 result. The result is written into the destination register for RV32 or sign-extended to 64-bits and written into the destination register for RV64. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then doubled and saturated into a Q31 value. The Q31 value is then written into Rd (sign-extended in RV64). When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFFFFFF and the overflow flag OV will be set.

#### Operations:

```
aop = Rs1.H[0]; bop = Rs2.H[0]; // KDMBB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KDMBT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KDMTT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult = aop * bop;
    resQ31 = Mresult << 1;
    Rd = resQ31; // RV32
    Rd = SE(resQ31); // RV64
} else {
```

(continues on next page)

(continued from previous page)

```

    resQ31 = 0x7FFFFFFF;
    Rd = resQ31; // RV32
    Rd = SE(resQ31); // RV64
    OV = 1;
}

```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KDMABB(long t, unsigned int a, unsigned int b)**  
KDMABB (Signed Saturating Double Multiply Addition B16 x B16)

**Type:** DSP

**Syntax:**

```
KDMAxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then double and saturate the Q31 result, add the result with the sign-extended lower 32-bit chunk destination register and write the saturated addition result into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then doubled and saturated into a Q31 value. The Q31 value is then added with the content of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV flag is set to 1. The result after saturation is written to Rd. When both the two Q15 inputs are 0x8000, saturation will happen and the overflow flag OV will be set.

**Operations:**

```

aop = Rs1.H[0]; bop = Rs2.H[0]; // KDMABB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KDMABT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KDMATT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult = aop * bop;
    resQ31 = Mresult << 1;
} else {
    resQ31 = 0x7FFFFFFF;
    OV = 1;
}
resadd = Rd + resQ31; // RV32
resadd = Rd.W[0] + resQ31; // RV64
if (resadd > (2^31)-1) {
    resadd = (2^31)-1;
    OV = 1;
} else if (resadd < -2^31) {
    resadd = -2^31;
    OV = 1;
}
Rd = resadd; // RV32
Rd = SE(resadd); // RV64

```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KDMABT(long t, unsigned int a, unsigned int b)**  
KDMABT (Signed Saturating Double Multiply Addition B16 x T16)

**Type:** DSP

**Syntax:**

KDMAxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
---------------------------------------

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then double and saturate the Q31 result, add the result with the sign-extended lower 32-bit chunk destination register and write the saturated addition result into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then doubled and saturated into a Q31 value. The Q31 value is then added with the content of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV flag is set to 1. The result after saturation is written to Rd. When both the two Q15 inputs are 0x8000, saturation will happen and the overflow flag OV will be set.

**Operations:**

```
aop = Rs1.H[0]; bop = Rs2.H[0]; // KDMABB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KDMABT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KDMATT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult = aop * bop;
    resQ31 = Mresult << 1;
} else {
    resQ31 = 0x7FFFFFFF;
    OV = 1;
}
resadd = Rd + resQ31; // RV32
resadd = Rd.W[0] + resQ31; // RV64
if (resadd > (2^31)-1) {
    resadd = (2^31)-1;
    OV = 1;
} else if (resadd < -2^31) {
    resadd = -2^31;
    OV = 1;
}
Rd = resadd; // RV32
Rd = SE(resadd); // RV64
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned int type of value stored in a



- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KDMATT(long t, unsigned int a, unsigned int b)**  
KDMATT (Signed Saturating Double Multiply Addition T16 x T16)

**Type:** DSP

**Syntax:**

KDMAxy Rd, Rs1, Rs2 (xy = BB, BT, TT)
---------------------------------------

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the lower 32-bit chunk in registers and then double and saturate the Q31 result, add the result with the sign-extended lower 32-bit chunk destination register and write the saturated addition result into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs1 with the top or bottom 16-bit Q15 content of the lower 32-bit portion in Rs2. The Q30 result is then doubled and saturated into a Q31 value. The Q31 value is then added with the content of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV flag is set to 1. The result after saturation is written to Rd. When both the two Q15 inputs are 0x8000, saturation will happen and the overflow flag OV will be set.

**Operations:**

```
aop = Rs1.H[0]; bop = Rs2.H[0]; // KDMABB
aop = Rs1.H[0]; bop = Rs2.H[1]; // KDMABT
aop = Rs1.H[1]; bop = Rs2.H[1]; // KDMATT
If (0x8000 != aop | 0x8000 != bop) {
    Mresult = aop * bop;
    resQ31 = Mresult << 1;
} else {
    resQ31 = 0x7FFFFFFF;
    OV = 1;
}
resadd = Rd + resQ31; // RV32
resadd = Rd.W[0] + resQ31; // RV64
if (resadd > (2^31)-1) {
    resadd = (2^31)-1;
    OV = 1;
} else if (resadd < -2^31) {
    resadd = -2^31;
    OV = 1;
}
Rd = resadd; // RV32
Rd = SE(resadd); // RV64
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KSLLW(long a, unsigned int b)**  
KSLLW (Saturating Shift Left Logical for Word)

**Type:** DSP

**Syntax:**

```
KSLRW Rd, Rs1, Rs2
```

**Purpose:** Do logical left shift operation with saturation on a 32-bit word. The shift amount is a variable from a GPR.

**Description:** The first word data in Rs1 is left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the low-order 5-bits of the value in the Rs2 register. Any shifted value greater than  $2^{31}-1$  is saturated to  $2^{31}-1$ . Any shifted value smaller than  $-2^{31}$  is saturated to  $-2^{31}$ . And the saturated result is sign-extended and written to Rd. If any saturation is performed, set OV bit to 1.

**Operations:**

```
sa = Rs2[4:0];
res[(31+sa):0] = Rs1.W[0] << sa;
if (res > (2^31)-1) {
    res = 0x7fffffff; OV = 1;
} else if (res < -2^31) {
    res = 0x80000000; OV = 1;
}
Rd[31:0] = res[31:0]; // RV32
Rd[63:0] = SE(res[31:0]); // RV64
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned int type of value stored in b

```
__STATIC_FORCEINLINE long __RV_KSLRAW(int a, int b)
KSLRAW (Shift Left Logical with Q31 Saturation or Shift Right Arithmetic)
```

**Type:** DSP

**Syntax:**

```
KSLRAW Rd, Rs1, Rs2
```

**Purpose:** Perform a logical left (positive) or arithmetic right (negative) shift operation with Q31 saturation for the left shift on a 32-bit data.

**Description:** The lower 32-bit content of Rs1 is left-shifted logically or right-shifted arithmetically based on the value of Rs2[5:0]. Rs2[5:0] is in the signed range of [-25, 25-1]. A positive Rs2[5:0] means logical left shift and a negative Rs2[5:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[5:0] clamped to the actual shift range of [0, 31]. The left-shifted result is saturated to the 32-bit signed integer range of  $[-2^{31}, 2^{31}-1]$ . After the shift operation, the final result is bit-31 sign-extended and written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:6] will not affected the operation of this instruction.

**Operations:**

```
if (Rs2[5:0] < 0) {
    sa = -Rs2[5:0];
    sa = (sa == 32)? 31 : sa;
    res[31:0] = Rs1.W[0] >>(arith) sa;
} else {
    sa = Rs2[5:0];
```

(continues on next page)

(continued from previous page)

```

tmp = Rs1.W[0] <<(logic) sa;
if (tmp > (2^31)-1) {
    res[31:0] = (2^31)-1;
    OV = 1;
} else if (tmp < -2^31) {
    res[31:0] = -2^31;
    OV = 1;
} else {
    res[31:0] = tmp[31:0];
}
}
Rd = res[31:0]; // RV32
Rd = SE64(res[31:0]); // RV64

```

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KSLRAW U(int a, int b)**

KSLRAW.u (Shift Left Logical with Q31 Saturation or Rounding Shift Right Arithmetic)

**Type:** DSP

**Syntax:**

```
KSLRAW.u Rd, Rs1, Rs2
```

**Purpose:** Perform a logical left (positive) or arithmetic right (negative) shift operation with Q31 saturation for the left shift and a rounding up operation for the right shift on a 32-bit data.

**Description:** The lower 32-bit content of Rs1 is left-shifted logically or right-shifted arithmetically based on the value of Rs2[5:0]. Rs2[5:0] is in the signed range of [-25, 25-1]. A positive Rs2[5:0] means logical left shift and a negative Rs2[5:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[5:0] clamped to the actual shift range of [0, 31]. The left-shifted result is saturated to the 32-bit signed integer range of [-2<sup>31</sup>, 2<sup>31</sup>-1]. The right-shifted result is added a 1 to the most significant discarded bit position for rounding effect. After the shift, saturation, or rounding, the final result is bit-31 sign-extended and written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:6] will not affect the operation of this instruction.

**Operations:**

```

if (Rs2[5:0] < 0) {
    sa = -Rs2[5:0];
    sa = (sa == 32)? 31 : sa;
    res[31:-1] = SE33(Rs1[31:(sa-1)]) + 1;
    rst[31:0] = res[31:0];
} else {
    sa = Rs2[5:0];
    tmp = Rs1.W[0] <<(logic) sa;
    if (tmp > (2^31)-1) {
        rst[31:0] = (2^31)-1;
        OV = 1;
    } else if (tmp < -2^31) {
        rst[31:0] = -2^31;
    }
}

```

(continues on next page)

(continued from previous page)

```

    OV = 1
  } else {
    rst[31:0] = tmp[31:0];
  }
}
Rd = rst[31:0]; // RV32
Rd = SE64(rst[31:0]); // RV64

```

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KSUBW(int a, int b)**  
 KSUBW (Signed Subtraction with Q31 Saturation)

**Type:** DSP

**Syntax:**

```
KSUBW Rd, Rs1, Rs2
```

**Purpose:** Subtract the signed lower 32-bit content of two registers with Q31 saturation.

**Description:** The signed lower 32-bit content of Rs2 is subtracted from the signed lower 32-bit content of Rs1. And the result is saturated to the 32-bit signed integer range of  $[-2^{31}, 2^{31}-1]$  and then sign-extended and written to Rd. If saturation happens, this instruction sets the OV flag.

**Operations:**

```

tmp = Rs1.W[0] - Rs2.W[0];
if (tmp > (2^31)-1) {
  res = (2^31)-1;
  OV = 1;
} else if (tmp < -2^31) {
  res = -2^31;
  OV = 1;
} else {
  res = tmp;
}
Rd = res[31:0]; // RV32
Rd = SE(res[31:0]); // RV64

```

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKADDW(unsigned int a, unsigned int b)**  
 UKADDW (Unsigned Addition with U32 Saturation)

**Type:** DSP

**Syntax:**

```
UKADDW Rd, Rs1, Rs2
```

**Purpose:** Add the unsigned lower 32-bit content of two registers with U32 saturation.

**Description:** The unsigned lower 32-bit content of Rs1 is added with the unsigned lower 32-bit content of Rs2. And the result is saturated to the 32-bit unsigned integer range of  $[0, 2^{32}-1]$  and then sign-extended and written to Rd. If saturation happens, this instruction sets the OV flag.

**Operations:**

```
tmp = Rs1.W[0] + Rs2.W[0];
if (tmp > (2^32)-1) {
    tmp[31:0] = (2^32)-1;
    OV = 1;
}
Rd = tmp[31:0]; // RV32
Rd = SE(tmp[31:0]); // RV64
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSUBW(unsigned int a, unsigned int b)**  
UKSUBW (Unsigned Subtraction with U32 Saturation)

**Type:** DSP

**Syntax:**

```
UKSUBW Rd, Rs1, Rs2
```

**Purpose:** Subtract the unsigned lower 32-bit content of two registers with unsigned 32-bit saturation.

**Description:** The unsigned lower 32-bit content of Rs2 is subtracted from the unsigned lower 32-bit content of Rs1. And the result is saturated to the 32-bit unsigned integer range of  $[0, 2^{32}-1]$  and then sign-extended and written to Rd. If saturation happens, this instruction sets the OV flag.

**Operations:**

```
tmp = Rs1.W[0] - Rs2.W[0];
if (tmp < 0) {
    tmp[31:0] = 0;
    OV = 1;
}
Rd = tmp[31:0]; // RV32
Rd = SE(tmp[31:0]); // RV64
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

### 32-bit Computation Instructions

```
__STATIC_FORCEINLINE long __RV_MAXW(int a, int b)
__STATIC_FORCEINLINE long __RV_MINW(int a, int b)
__STATIC_FORCEINLINE unsigned long long __RV_MULR64(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_MULSR64(long a, long b)
__STATIC_FORCEINLINE long __RV_RADDW(int a, int b)
__STATIC_FORCEINLINE long __RV_RSUBW(int a, int b)
__STATIC_FORCEINLINE unsigned long __RV_URADDW(unsigned int a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_URSUBW(unsigned int a, unsigned int b)
```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_32B\_COMPUTATION**

32-bit Computation Instructions

there are 8 32-bit Computation Instructions

#### Functions

```
__STATIC_FORCEINLINE long __RV_MAXW(int a, int b)
MAXW (32-bit Signed Word Maximum)
```

**Type:** DSP

**Syntax:**

MAXW Rd, Rs1, Rs2
-------------------

**Purpose:** Get the larger value from the 32-bit contents of two general registers.

**Description:** This instruction compares two signed 32-bit integers stored in Rs1 and Rs2, picks the larger value as the result, and writes the result to Rd.

**Operations:**

<pre>if (Rs1.W[0] &gt;= Rs2.W[0]) {     Rd = SE(Rs1.W[0]); } else {     Rd = SE(Rs2.W[0]); }</pre>
--

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

```
__STATIC_FORCEINLINE long __RV_MINW(int a, int b)
MINW (32-bit Signed Word Minimum)
```

**Type:** DSP

**Syntax:**

```
MINW Rd, Rs1, Rs2
```

**Purpose:** Get the smaller value from the 32-bit contents of two general registers.

**Description:** This instruction compares two signed 32-bit integers stored in Rs1 and Rs2, picks the smaller value as the result, and writes the result to Rd.

**Operations:**

```
if (Rs1.W[0] >= Rs2.W[0]) { Rd = SE(Rs2.W[0]); } else { Rd = SE(Rs1.W[0]); }
```

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

```
__STATIC_FORCEINLINE unsigned long long __RV_MULR64(unsigned long a, unsigned long b)
    MULR64 (Multiply Word Unsigned to 64-bit Data)
```

**Type:** DSP

**Syntax:**

```
MULR64 Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit unsigned integer contents of two registers and write the 64-bit result.

**RV32 Description:** This instruction multiplies the 32-bit content of Rs1 with that of Rs2 and writes the 64-bit multiplication result to an even/odd pair of registers containing Rd. Rd(4,1) index d determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result. The lower 32-bit contents of Rs1 and Rs2 are treated as unsigned integers.

**RV64 Description:** This instruction multiplies the lower 32-bit content of Rs1 with that of Rs2 and writes the 64-bit multiplication result to Rd. The lower 32-bit contents of Rs1 and Rs2 are treated as unsigned integers.

**Operations:**

```
RV32:
Mresult = CONCAT(1'b0, Rs1) u* CONCAT(1'b0, Rs2);
R[Rd(4,1).1(0)][31:0] = Mresult[63:32];
R[Rd(4,1).0(0)][31:0] = Mresult[31:0];
RV64:
Rd = Mresult[63:0];
Mresult = CONCAT(1'b0, Rs1.W[0]) u* CONCAT(1'b0, Rs2.W[0]);
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long long __RV_MULSR64(long a, long b)
    MULSR64 (Multiply Word Signed to 64-bit Data)
```

**Type:** DSP**Syntax:**

```
MULSR64 Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit signed integer contents of two registers and write the 64-bit result.

**RV32 Description:** This instruction multiplies the lower 32-bit content of Rs1 with the lower 32-bit content of Rs2 and writes the 64-bit multiplication result to an even/odd pair of registers containing Rd. Rd(4,1) index d determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result. The lower 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**RV64 Description:** This instruction multiplies the lower 32-bit content of Rs1 with the lower 32-bit content of Rs2 and writes the 64-bit multiplication result to Rd. The lower 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
RV32:
Mresult = Ra s* Rb;
R[Rd(4,1).1(0)][31:0] = Mresult[63:32];
R[Rd(4,1).0(0)][31:0] = Mresult[31:0];
RV64:
Mresult = Ra.W[0] s* Rb.W[0];
Rd = Mresult[63:0];
```

**Return** value stored in long long type**Parameters**

- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_RADDW(int a, int b)
```

RADDW (32-bit Signed Halving Addition)

**Type:** DSP**Syntax:**

```
RADDW Rd, Rs1, Rs2
```

**Purpose:** Add 32-bit signed integers and the results are halved to avoid overflow or saturation.

**Description:** This instruction adds the first 32-bit signed integer in Rs1 with the first 32-bit signed integer in Rs2. The result is first arithmetically right-shifted by 1 bit and then sign-extended and written to Rd.

**Examples:**

```
* Rs1 = 0x7FFFFFFF, Rs2 = 0x7FFFFFFF, Rd = 0x7FFFFFFF
* Rs1 = 0x80000000, Rs2 = 0x80000000, Rd = 0x80000000
* Rs1 = 0x40000000, Rs2 = 0x80000000, Rd = 0xE0000000
```

**Operations:**



```
RV32:
Rd[31:0] = (Rs1[31:0] + Rs2[31:0]) s>> 1;
RV64:
resw[31:0] = (Rs1[31:0] + Rs2[31:0]) s>> 1;
Rd[63:0] = SE(resw[31:0]);
```

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

```
__STATIC_FORCEINLINE long __RV_RSUBW(int a, int b)
RSUBW (32-bit Signed Halving Subtraction)
```

**Type:** DSP

**Syntax:**

```
RSUBW Rd, Rs1, Rs2
```

**Purpose:** Subtract 32-bit signed integers and the result is halved to avoid overflow or saturation.

**Description:** This instruction subtracts the first 32-bit signed integer in Rs2 from the first 32-bit signed integer in Rs1. The result is first arithmetically right-shifted by 1 bit and then sign-extended and written to Rd.

**Examples:**

```
* Rs1 = 0x7FFFFFFF, Rs2 = 0x80000000, Rd = 0x7FFFFFFF
* Rs1 = 0x80000000, Rs2 = 0x7FFFFFFF, Rd = 0x80000000
* Rs1 = 0x80000000, Rs2 = 0x40000000, Rd = 0xA0000000
```

**Operations:**

```
RV32:
Rd[31:0] = (Rs1[31:0] - Rs2[31:0]) s>> 1;
RV64:
resw[31:0] = (Rs1[31:0] - Rs2[31:0]) s>> 1;
Rd[63:0] = SE(resw[31:0]);
```

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: int type of value stored in b

```
__STATIC_FORCEINLINE unsigned long __RV_URADDW(unsigned int a, unsigned int b)
URADDW (32-bit Unsigned Halving Addition)
```

**Type:** DSP

**Syntax:**

```
URADDW Rd, Rs1, Rs2
```

**Purpose:** Add 32-bit unsigned integers and the results are halved to avoid overflow or saturation.

**Description:** This instruction adds the first 32-bit unsigned integer in Rs1 with the first 32-bit unsigned integer in Rs2. The result is first logically right-shifted by 1 bit and then sign-extended and written to Rd.

**Examples:**

```
* Ra = 0x7FFFFFFF, Rb = 0x7FFFFFFF Rt = 0x7FFFFFFF
* Ra = 0x80000000, Rb = 0x80000000 Rt = 0x80000000
* Ra = 0x40000000, Rb = 0x80000000 Rt = 0x60000000
```

**Operations:**

```
* RV32:
Rd[31:0] = (Rs1[31:0] + Rs2[31:0]) u>> 1;
* RV64:
resw[31:0] = (Rs1[31:0] + Rs2[31:0]) u>> 1;
Rd[63:0] = SE(resw[31:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URSUBW(unsigned int a, unsigned int b)**  
URSUBW (32-bit Unsigned Halving Subtraction)

**Type:** DSP

**Syntax:**

```
URSUBW Rd, Rs1, Rs2
```

**Purpose:** Subtract 32-bit unsigned integers and the result is halved to avoid overflow or saturation.

**Description:** This instruction subtracts the first 32-bit signed integer in Rs2 from the first 32-bit signed integer in Rs1. The result is first logically right-shifted by 1 bit and then sign-extended and written to Rd.

**Examples:**

```
* Ra = 0x7FFFFFFF, Rb = 0x80000000 Rt = 0xFFFFFFFF
* Ra = 0x80000000, Rb = 0x7FFFFFFF Rt = 0x00000000
* Ra = 0x80000000, Rb = 0x40000000 Rt = 0x20000000
```

**Operations:**

```
* RV32:
Rd[31:0] = (Rs1[31:0] - Rs2[31:0]) u>> 1;
* RV64:
resw[31:0] = (Rs1[31:0] - Rs2[31:0]) u>> 1;
Rd[63:0] = SE(resw[31:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned int type of value stored in a
- [in] b: unsigned int type of value stored in b

## OV (Overflow) flag Set/Clear Instructions

```
__STATIC_FORCEINLINE void __RV_CLROV(void)
__STATIC_FORCEINLINE unsigned long __RV_RDOV(void)
```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_OV\_FLAG\_SC**

OV (Overflow) flag Set/Clear Instructions.

The following table lists the user instructions related to Overflow (OV) flag manipulation. there are 2 OV (Overflow) flag Set/Clear Instructions

### Functions

```
__STATIC_FORCEINLINE void __RV_CLROV(void)
    CLROV (Clear OV flag)
```

**Type:** DSP

**Syntax:**

```
CLROV # pseudo mnemonic
```

**Purpose:** This pseudo instruction is an alias to CSRRCI x0, ucode, 1 instruction.

```
__STATIC_FORCEINLINE unsigned long __RV_RDOV(void)
    RDOV (Read OV flag)
```

**Type:** DSP

**Syntax:**

```
RDOV Rd # pseudo mnemonic
```

**Purpose:** This pseudo instruction is an alias to CSRR Rd, ucode instruction which maps to the real instruction of CSRRS Rd, ucode, x0.

**Return** value stored in unsigned long type

## Non-SIMD Miscellaneous Instructions

```
__STATIC_FORCEINLINE long __RV_AVE(long a, long b)
__STATIC_FORCEINLINE unsigned long __RV_BITREV(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_BPICK(unsigned long a, unsigned long b, unsigned long c)
__STATIC_FORCEINLINE unsigned long __RV_MADDR32(unsigned long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_MSUBR32(unsigned long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SRA_U(long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SWAP8(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_SWAP16(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_WEXT(long long a, unsigned int b)
__RV_BITREVI(a, b)
```

\_\_RV\_INSB (t, a, b)

\_\_RV\_SRAI\_U (a, b)

\_\_RV\_WEXTI (a, b)

*group* **NMSIS\_Core\_DSP\_Intrinsic\_NON\_SIMD\_MISC**

Non-SIMD Miscellaneous Instructions.

There are 13 Miscellaneous Instructions here.

## Defines

\_\_RV\_BITREVI (a, b)

BITREVI (Bit Reverse Immediate)

**Type:** DSP

**Syntax:**

```
(RV32) BITREVI Rd, Rs1, imm[4:0]
(RV64) BITREVI Rd, Rs1, imm[5:0]
```

**Purpose:** Reverse the bit positions of the source operand within a specified width starting from bit 0. The reversed width is an immediate value.

**Description:** This instruction reverses the bit positions of the content of Rs1. The reversed bit width is calculated as imm[4:0]+1 (RV32) or imm[5:0]+1 (RV64). The upper bits beyond the reversed width are filled with zeros. After the bit reverse operation, the result is written to Rd.

**Operations:**

```
msb = imm[4:0]; (RV32)
msb = imm[5:0]; (RV64)
rev[0:msb] = Rs1[msb:0];
Rd = ZE(rev[msb:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

\_\_RV\_INSB (t, a, b)

INSB (Insert Byte)

**Type:** DSP

**Syntax:**

```
(RV32) INSB Rd, Rs1, imm[1:0]
(RV64) INSB Rd, Rs1, imm[2:0]
```

**Purpose:** Insert byte 0 of a 32-bit or 64-bit register into one of the byte elements of another register.

**Description:** This instruction inserts byte 0 of Rs1 into byte imm[1:0] (RV32) or imm[2:0] (RV64) of Rd.

**Operations:**

```
bpos = imm[1:0]; (RV32)
bpos = imm[2:0]; (RV64)
Rd.B[bpos] = Rs1.B[0]
```

**Return** value stored in unsigned long type

**Parameters**

- [in] t: unsigned long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_RV\_SRAI\_U** (a, b)

SRAI.u (Rounding Shift Right Arithmetic Immediate)

**Type:** DSP

**Syntax:**

```
SRAI.u Rd, Rs1, imm6u[4:0] (RV32)
SRAI.u Rd, Rs1, imm6u[5:0] (RV64)
```

**Purpose:** Perform an arithmetic right shift operation with rounding. The shift amount is an immediate value.

**Description:** This instruction right-shifts the content of Rs1 arithmetically. The shifted out bits are filled with the sign-bit and the shift amount is specified by the imm6u[4:0] (RV32) or imm6u[5:0] (RV64) constant. For the rounding operation, a value of 1 is added to the most significant discarded bit of the data to calculate the final result. And the result is written to Rd.

**Operations:**

```
* RV32:
sa = imm6u[4:0];
if (sa > 0) {
    res[31:-1] = SE33(Rs1[31:(sa-1)]) + 1;
    Rd = res[31:0];
} else {
    Rd = Rs1;
}

* RV64:
sa = imm6u[5:0];
if (sa > 0) {
    res[63:-1] = SE65(Rs1[63:(sa-1)]) + 1;
    Rd = res[63:0];
} else {
    Rd = Rs1;
}
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_RV\_WEXTI** (a, b)

WEXTI (Extract Word from 64-bit Immediate)

**Type:** DSP**Syntax:**

WEXTI Rd, Rs1, #LSBloc
------------------------

**Purpose:** Extract a 32-bit word from a 64-bit value stored in an even/odd pair of registers (RV32) or a register (RV64) starting from a specified immediate LSB bit position.

**RV32 Description:** This instruction extracts a 32-bit word from a 64-bit value of an even/odd pair of registers specified by Rs1(4,1) starting from a specified immediate LSB bit position, #LSBloc. The extracted word is written to Rd. Rs1(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the 64-bit value and the even 2d register of the pair contains the low 32-bit of the 64-bit value.

**RV64 Description:** This instruction extracts a 32-bit word from a 64-bit value in Rs1 starting from a specified immediate LSB bit position, #LSBloc. The extracted word is sign-extended and written to lower 32-bit of Rd.

**Operations:**

<pre>* RV32: Idx0 = CONCAT(Rs1(4,1), 1'b0); Idx1 = CONCAT(Rs2(4,1), 1'b1); src[63:0] = Concat(R[Idx1], R[Idx0]); Rd = src[31+LSBloc:LSBloc]; * RV64: ExtractW = Rs1[31+LSBloc:LSBloc]; Rd = SE(ExtractW)</pre>
--

**Return** value stored in unsigned long type**Parameters**

- [in] a: long long type of value stored in a
- [in] b: unsigned int type of value stored in b

## Functions

**\_\_STATIC\_FORCEINLINE** long **\_\_RV\_AVE**(long a, long b)

AVE (Average with Rounding)

**Type:** DSP**Syntax:**

AVE Rd, Rs1, Rs2
------------------

**Purpose:** Calculate the average of the contents of two general registers.

**Description:** This instruction calculates the average value of two signed integers stored in Rs1 and Rs2, rounds up a half-integer result to the nearest integer, and writes the result to Rd.

**Operations:**

```

Sum = CONCAT(Rs1[MSB],Rs1[MSB:0]) + CONCAT(Rs2[MSB],Rs2[MSB:0]) + 1;
Rd = Sum[(MSB+1):1];
for RV32: MSB=31,
for RV64: MSB=63

```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_BITREV(unsigned long a, unsigned long b)**  
BITREV (Bit Reverse)

**Type:** DSP

**Syntax:**

```
BITREV Rd, Rs1, Rs2
```

**Purpose:** Reverse the bit positions of the source operand within a specified width starting from bit 0. The reversed width is a variable from a GPR.

**Description:** This instruction reverses the bit positions of the content of Rs1. The reversed bit width is calculated as Rs2[4:0]+1 (RV32) or Rs2[5:0]+1 (RV64). The upper bits beyond the reversed width are filled with zeros. After the bit reverse operation, the result is written to Rd.

**Operations:**

```

msb = Rs2[4:0]; (for RV32)
msb = Rs2[5:0]; (for RV64)
rev[0:msb] = Rs1[msb:0];
Rd = ZE(rev[msb:0]);

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_BPICK(unsigned long a, unsigned long b, unsigned long c)**  
BPICK (Bit-wise Pick)

**Type:** DSP

**Syntax:**

```
BPICK Rd, Rs1, Rs2, Rc
```

**Purpose:** Select from two source operands based on a bit mask in the third operand.

**Description:** This instruction selects individual bits from Rs1 or Rs2, based on the bit mask value in Rc. If a bit in Rc is 1, the corresponding bit is from Rs1; otherwise, the corresponding bit is from Rs2. The selection results are written to Rd.

**Operations:**

```
Rd[x] = Rc[x] ? Rs1[x] : Rs2[x];
for RV32, x=31...0
for RV64, x=63...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b
- [in] c: unsigned long type of value stored in c

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_MADDR32(unsigned long t, unsigned long a, unsigned long b)**  
MADDR32 (Multiply and Add to 32-Bit Word)

**Type:** DSP

**Syntax:**

```
MADDR32 Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit contents of two registers and add the lower 32-bit multiplication result to the 32-bit content of a destination register. Write the final result back to the destination register.

**Description:** This instruction multiplies the lower 32-bit content of Rs1 with that of Rs2. It adds the lower 32-bit multiplication result to the lower 32-bit content of Rd and writes the final result (RV32) or sign-extended result (RV64) back to Rd. The contents of Rs1 and Rs2 can be either signed or unsigned integers.

**Operations:**

```
RV32:
Mresult = Rs1 * Rs2;
Rd = Rd + Mresult.W[0];
RV64:
Mresult = Rs1.W[0] * Rs2.W[0];
tres[31:0] = Rd.W[0] + Mresult.W[0];
Rd = SE64(tres[31:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] t: unsigned long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_MSUBR32(unsigned long t, unsigned long a, unsigned long b)**  
MSUBR32 (Multiply and Subtract from 32-Bit Word)

**Type:** DSP

**Syntax:**

```
MSUBR32 Rd, Rs1, Rs2
```



**Purpose:** Multiply the 32-bit contents of two registers and subtract the lower 32-bit multiplication result from the 32-bit content of a destination register. Write the final result back to the destination register.

**Description:** This instruction multiplies the lower 32-bit content of Rs1 with that of Rs2, subtracts the lower 32-bit multiplication result from the lower 32-bit content of Rd, then writes the final result (RV32) or sign-extended result (RV64) back to Rd. The contents of Rs1 and Rs2 can be either signed or unsigned integers.

**Operations:**

```
RV32:
Mresult = Rs1 * Rs2;
Rd = Rd - Mresult.W[0];
RV64:
Mresult = Rs1.W[0] * Rs2.W[0];
tres[31:0] = Rd.W[0] - Mresult.W[0];
Rd = SE64(tres[31:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] t: unsigned long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_SRA_U(long a, unsigned int b)
SRA.u (Rounding Shift Right Arithmetic)
```

**Type:** DSP

**Syntax:**

```
SRA.u Rd, Rs1, Rs2
```

**Purpose:** Perform an arithmetic right shift operation with rounding. The shift amount is a variable from a GPR.

**Description:** This instruction right-shifts the content of Rs1 arithmetically. The shifted out bits are filled with the sign-bit and the shift amount is specified by the low-order 5-bits (RV32) or 6-bits (RV64) of the Rs2 register. For the rounding operation, a value of 1 is added to the most significant discarded bit of the data to calculate the final result. And the result is written to Rd.

**Operations:**

```
* RV32:
sa = Rs2[4:0];
if (sa > 0) {
    res[31:-1] = SE33(Rs1[31:(sa-1)]) + 1;
    Rd = res[31:0];
} else {
    Rd = Rs1;
}
* RV64:
sa = Rs2[5:0];
if (sa > 0) {
    res[63:-1] = SE65(Rs1[63:(sa-1)]) + 1;
    Rd = res[63:0];
```

(continues on next page)

(continued from previous page)

```

} else {
    Rd = Rs1;
}

```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SWAP8(unsigned long a)**  
 SWAP8 (Swap Byte within Halfword)

**Type:** DSP

**Syntax:**

```
SWAP8 Rd, Rs1
```

**Purpose:** Swap the bytes within each halfword of a register.

**Description:** This instruction swaps the bytes within each halfword of Rs1 and writes the result to Rd.

**Operations:**

```

Rd.H[x] = CONCAT(Rs1.H[x][7:0],Rs1.H[x][15:8]);
for RV32: x=1...0,
for RV64: x=3...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SWAP16(unsigned long a)**  
 SWAP16 (Swap Halfword within Word)

**Type:** DSP

**Syntax:**

```
SWAP16 Rd, Rs1
```

**Purpose:** Swap the 16-bit halfwords within each word of a register.

**Description:** This instruction swaps the 16-bit halfwords within each word of Rs1 and writes the result to Rd.

**Operations:**

```

Rd.W[x] = CONCAT(Rs1.W[x][15:0],Rs1.H[x][31:16]);
for RV32: x=0,
for RV64: x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_WEXT(long long a, unsigned int b)**  
WEXT (Extract Word from 64-bit)

**Type:** DSP

**Syntax:**

```
WEXT Rd, Rs1, Rs2
```

**Purpose:** Extract a 32-bit word from a 64-bit value stored in an even/odd pair of registers (RV32) or a register (RV64) starting from a specified LSB bit position in a register.

**RV32 Description:** This instruction extracts a 32-bit word from a 64-bit value of an even/odd pair of registers specified by Rs1(4,1) starting from a specified LSB bit position, specified in Rs2[4:0]. The extracted word is written to Rd. Rs1(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the 64-bit value and the even 2d register of the pair contains the low 32-bit of the 64-bit value.

**Operations:**

```
* RV32:
Idx0 = CONCAT(Rs1(4,1), 1'b0); Idx1 = CONCAT(Rs1(4,1), 1'b1);
src[63:0] = Concat(R[Idx1], R[Idx0]);
LSBloc = Rs2[4:0];
Rd = src[31+LSBloc:LSBloc];
* RV64:
LSBloc = Rs2[4:0];
ExtractW = Rs1[31+LSBloc:LSBloc];
Rd = SE(ExtractW)
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: long long type of value stored in a
- [in] b: unsigned int type of value stored in b

*group* **NMSIS\_Core\_DSP\_Intrinsic\_NON\_SIMD**  
Non-SIMD Instructions.

## Partial-SIMD Data Processing Instructions

### SIMD 16-bit Packing Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_PKBB16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_PKBT16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_PKTT16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_PKTB16(unsigned long a, unsigned long b)
```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_SIMD\_16B\_PACK**  
SIMD 16-bit Packing Instructions.

there are 4 SIMD16-bit Packing Instructions.

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PKBB16(unsigned long a, unsigned long b)**  
 PKBB16 (Pack Two 16-bit Data from Both Bottom Half)

**Type:** DSP

**Syntax:**

```
PKBB16 Rd, Rs1, Rs2
PKBT16 Rd, Rs1, Rs2
PKTT16 Rd, Rs1, Rs2
PKTB16 Rd, Rs1, Rs2
```

**Purpose:** Pack 16-bit data from 32-bit chunks in two registers.

- PKBB16: bottom.bottom
- PKBT16 bottom.top
- PKTT16 top.top
- PKTB16 top.bottom

**Description:** (PKBB16) moves Rs1.W[x][15:0] to Rd.W[x][31:16] and moves Rs2.W[x][15:0] to Rd.W[x][15:0]. (PKBT16) moves Rs1.W[x][15:0] to Rd.W[x][31:16] and moves Rs2.W[x][31:16] to Rd.W[x][15:0]. (PKTT16) moves Rs1.W[x][31:16] to Rd.W[x][31:16] and moves Rs2.W[x][31:16] to Rd.W[x][15:0]. (PKTB16) moves Rs1.W[x][31:16] to Rd.W[x][31:16] and moves Rs2.W[x][15:0] to Rd.W[x][15:0].

**Operations:**

```
Rd.W[x][31:0] = CONCAT(Rs1.W[x][15:0], Rs2.W[x][15:0]); // PKBB16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][15:0], Rs2.W[x][31:16]); // PKBT16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][31:16], Rs2.W[x][15:0]); // PKTB16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][31:16], Rs2.W[x][31:16]); // PKTT16
for RV32: x=0,
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PKBT16(unsigned long a, unsigned long b)**  
 PKBT16 (Pack Two 16-bit Data from Bottom and Top Half)

**Type:** DSP

**Syntax:**

```
PKBB16 Rd, Rs1, Rs2
PKBT16 Rd, Rs1, Rs2
PKTT16 Rd, Rs1, Rs2
PKTB16 Rd, Rs1, Rs2
```

**Purpose:** Pack 16-bit data from 32-bit chunks in two registers.

- PKBB16: bottom.bottom

- PKBT16 bottom.top
- PKTT16 top.top
- PKTB16 top.bottom

**Description:** (PKBB16) moves Rs1.W[x][15:0] to Rd.W[x][31:16] and moves Rs2.W[x] [15:0] to Rd.W[x] [15:0]. (PKBT16) moves Rs1.W[x] [15:0] to Rd.W[x] [31:16] and moves Rs2.W[x] [31:16] to Rd.W[x] [15:0]. (PKTT16) moves Rs1.W[x] [31:16] to Rd.W[x] [31:16] and moves Rs2.W[x] [31:16] to Rd.W[x] [15:0]. (PKTB16) moves Rs1.W[x] [31:16] to Rd.W[x] [31:16] and moves Rs2.W[x] [15:0] to Rd.W[x] [15:0].

**Operations:**

```
Rd.W[x][31:0] = CONCAT(Rs1.W[x][15:0], Rs2.W[x][15:0]); // PKBB16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][15:0], Rs2.W[x][31:16]); // PKBT16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][31:16], Rs2.W[x][15:0]); // PKTB16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][31:16], Rs2.W[x][31:16]); // PKTT16
for RV32: x=0,
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PKTT16(unsigned long a, unsigned long b)**  
PKTT16 (Pack Two 16-bit Data from Both Top Half)

**Type:** DSP

**Syntax:**

```
PKBB16 Rd, Rs1, Rs2
PKBT16 Rd, Rs1, Rs2
PKTT16 Rd, Rs1, Rs2
PKTB16 Rd, Rs1, Rs2
```

**Purpose:** Pack 16-bit data from 32-bit chunks in two registers.

- PKBB16: bottom.bottom
- PKBT16 bottom.top
- PKTT16 top.top
- PKTB16 top.bottom

**Description:** (PKBB16) moves Rs1.W[x][15:0] to Rd.W[x][31:16] and moves Rs2.W[x] [15:0] to Rd.W[x] [15:0]. (PKBT16) moves Rs1.W[x] [15:0] to Rd.W[x] [31:16] and moves Rs2.W[x] [31:16] to Rd.W[x] [15:0]. (PKTT16) moves Rs1.W[x] [31:16] to Rd.W[x] [31:16] and moves Rs2.W[x] [31:16] to Rd.W[x] [15:0]. (PKTB16) moves Rs1.W[x] [31:16] to Rd.W[x] [31:16] and moves Rs2.W[x] [15:0] to Rd.W[x] [15:0].

**Operations:**

```
Rd.W[x][31:0] = CONCAT(Rs1.W[x][15:0], Rs2.W[x][15:0]); // PKBB16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][15:0], Rs2.W[x][31:16]); // PKBT16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][31:16], Rs2.W[x][15:0]); // PKTB16
```

(continues on next page)

(continued from previous page)

```

Rd.W[x][31:0] = CONCAT(Rs1.W[x][31:16], Rs2.W[x][31:16]); // PKTT16
for RV32: x=0,
for RV64: x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PKTB16(unsigned long a, unsigned long b)**  
 PKTB16 (Pack Two 16-bit Data from Top and Bottom Half)

**Type:** DSP

**Syntax:**

```

PKBB16 Rd, Rs1, Rs2
PKBT16 Rd, Rs1, Rs2
PKTT16 Rd, Rs1, Rs2
PKTB16 Rd, Rs1, Rs2

```

**Purpose:** Pack 16-bit data from 32-bit chunks in two registers.

- PKBB16: bottom.bottom
- PKBT16 bottom.top
- PKTT16 top.top
- PKTB16 top.bottom

**Description:** (PKBB16) moves Rs1.W[x][15:0] to Rd.W[x][31:16] and moves Rs2.W[x][15:0] to Rd.W[x][15:0]. (PKBT16) moves Rs1.W[x][15:0] to Rd.W[x][31:16] and moves Rs2.W[x][31:16] to Rd.W[x][15:0]. (PKTT16) moves Rs1.W[x][31:16] to Rd.W[x][31:16] and moves Rs2.W[x][31:16] to Rd.W[x][15:0]. (PKTB16) moves Rs1.W[x][31:16] to Rd.W[x][31:16] and moves Rs2.W[x][15:0] to Rd.W[x][15:0].

**Operations:**

```

Rd.W[x][31:0] = CONCAT(Rs1.W[x][15:0], Rs2.W[x][15:0]); // PKBB16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][15:0], Rs2.W[x][31:16]); // PKBT16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][31:16], Rs2.W[x][15:0]); // PKTB16
Rd.W[x][31:0] = CONCAT(Rs1.W[x][31:16], Rs2.W[x][31:16]); // PKTT16
for RV32: x=0,
for RV64: x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## Signed MSW 32x32 Multiply and Add Instructions

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAC(long t, long a, long b)**

```

__STATIC_FORCEINLINE long __RV_KMMAC_U(long t, long a, long b)
__STATIC_FORCEINLINE long __RV_KMMSB(long t, long a, long b)
__STATIC_FORCEINLINE long __RV_KMMSB_U(long t, long a, long b)
__STATIC_FORCEINLINE long __RV_KWMMUL(long a, long b)
__STATIC_FORCEINLINE long __RV_KWMMUL_U(long a, long b)
__STATIC_FORCEINLINE long __RV_SMMUL(long a, long b)
__STATIC_FORCEINLINE long __RV_SMMUL_U(long a, long b)

```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_SIGNED\_MSW\_32X32\_MAC**

Signed MSW 32x32 Multiply and Add Instructions.

there are 8 Signed MSW 32x32 Multiply and Add Instructions

## Functions

```

__STATIC_FORCEINLINE long __RV_KMMAC(long t, long a, long b)
    KMMAC (SIMD Saturating MSW Signed Multiply Word and Add)

```

**Type:** SIMD

**Syntax:**

```

KMMAC Rd, Rs1, Rs2
KMMAC.u Rd, Rs1, Rs2

```

**Purpose:** Multiply the signed 32-bit integer elements of two registers and add the most significant 32-bit results with the signed 32-bit integer elements of a third register. The addition results are saturated first and then written back to the third register. The `.u` form performs an additional rounding up operation on the multiplication results before adding the most significant 32-bit part of the results.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed 32-bit elements of Rs2 and adds the most significant 32-bit multiplication results with the signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to Rd. The `.u` form of the instruction additionally rounds up the most significant 32-bit of the 64-bit multiplication results by adding a 1 to bit 31 of the results.

**Operations:**

```

Mres[x][63:0] = Rs1.W[x] * Rs2.W[x];
if (` .u ` form) {
    Round[x][32:0] = Mres[x][63:31] + 1;
    res[x] = Rd.W[x] + Round[x][32:1];
} else {
    res[x] = Rd.W[x] + Mres[x][63:32];
}
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];

```

(continues on next page)

(continued from previous page)

```
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAC\_U(long t, long a, long b)**  
KMMAC.u (SIMD Saturating MSW Signed Multiply Word and Add with Rounding)

**Type:** SIMD

**Syntax:**

```
KMMAC Rd, Rs1, Rs2
KMMAC.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of two registers and add the most significant 32-bit results with the signed 32-bit integer elements of a third register. The addition results are saturated first and then written back to the third register. The .u form performs an additional rounding up operation on the multiplication results before adding the most significant 32-bit part of the results.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed 32-bit elements of Rs2 and adds the most significant 32-bit multiplication results with the signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to Rd. The .u form of the instruction additionally rounds up the most significant 32-bit of the 64-bit multiplication results by adding a 1 to bit 31 of the results.

**Operations:**

```
Mres[x][63:0] = Rs1.W[x] * Rs2.W[x];
if (`.u` form) {
    Round[x][32:0] = Mres[x][63:31] + 1;
    res[x] = Rd.W[x] + Round[x][32:1];
} else {
    res[x] = Rd.W[x] + Mres[x][63:32];
}
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**



- [in] t: long type of value stored in t
- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMSB(long t, long a, long b)**

KMMSB (SIMD Saturating MSW Signed Multiply Word and Subtract)

**Type:** SIMD

**Syntax:**

```
KMMSB Rd, Rs1, Rs2
KMMSB.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of two registers and subtract the most significant 32-bit results from the signed 32-bit elements of a third register. The subtraction results are written to the third register. The .u form performs an additional rounding up operation on the multiplication results before subtracting the most significant 32-bit part of the results.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed 32-bit elements of Rs2 and subtracts the most significant 32-bit multiplication results from the signed 32-bit elements of Rd. If the subtraction result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to Rd. The .u form of the instruction additionally rounds up the most significant 32-bit of the 64-bit multiplication results by adding a 1 to bit 31 of the results.

**Operations:**

```
Mres[x][63:0] = Rs1.W[x] * Rs2.W[x];
if (`.u` form) {
    Round[x][32:0] = Mres[x][63:31] + 1;
    res[x] = Rd.W[x] - Round[x][32:1];
} else {
    res[x] = Rd.W[x] - Mres[x][63:32];
}
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMSB\_U(long t, long a, long b)**

KMMSB.u (SIMD Saturating MSW Signed Multiply Word and Subtraction with Rounding)

**Type:** SIMD

**Syntax:**

```
KMMSB Rd, Rs1, Rs2
KMMSB.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of two registers and subtract the most significant 32-bit results from the signed 32-bit elements of a third register. The subtraction results are written to the third register. The `.u` form performs an additional rounding up operation on the multiplication results before subtracting the most significant 32-bit part of the results.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed 32-bit elements of Rs2 and subtracts the most significant 32-bit multiplication results from the signed 32-bit elements of Rd. If the subtraction result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to Rd. The `.u` form of the instruction additionally rounds up the most significant 32-bit of the 64-bit multiplication results by adding a 1 to bit 31 of the results.

**Operations:**

```
Mres[x][63:0] = Rs1.W[x] * Rs2.W[x];
if (`.u` form) {
    Round[x][32:0] = Mres[x][63:31] + 1;
    res[x] = Rd.W[x] - Round[x][32:1];
} else {
    res[x] = Rd.W[x] - Mres[x][63:32];
}
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_KWMMUL(long a, long b)
KWMMUL (SIMD Saturating MSW Signed Multiply Word & Double)
```

**Type:** SIMD

**Syntax:**

```
KWMMUL Rd, Rs1, Rs2
KWMMUL.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of two registers, shift the results left 1-bit, saturate, and write the most significant 32-bit results to a register. The `.u` form additionally rounds up the multiplication results from the most significant discarded bit.

**Description:** This instruction multiplies the 32-bit elements of Rs1 with the 32-bit elements of Rs2. It then shifts the multiplication results one bit to the left and takes the most significant 32-bit results. If the shifted result is greater than  $2^{31}-1$ , it is saturated to  $2^{31}-1$  and the OV flag is set to 1. The final element result is written to Rd. The 32-bit elements of Rs1 and Rs2 are treated as signed integers. The `.u` form of the instruction additionally rounds up the 64-bit multiplication results by adding a 1 to bit 30 before the shift and saturation operations.

**Operations:**

```
if ((0x80000000 != Rs1.W[x]) | (0x80000000 != Rs2.W[x])) {
    Mres[x][63:0] = Rs1.W[x] * Rs2.W[x];
    if (`.u` form) {
        Round[x][33:0] = Mres[x][63:30] + 1;
        Rd.W[x] = Round[x][32:1];
    } else {
        Rd.W[x] = Mres[x][62:31];
    }
} else {
    Rd.W[x] = 0x7fffffff;
    OV = 1;
}
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KWMUL\_U(long a, long b)**

KWMUL.u (SIMD Saturating MSW Signed Multiply Word & Double with Rounding)

**Type:** SIMD

**Syntax:**

```
KWMUL Rd, Rs1, Rs2
KWMUL.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of two registers, shift the results left 1-bit, saturate, and write the most significant 32-bit results to a register. The `.u` form additionally rounds up the multiplication results from the most significant discarded bit.

**Description:** This instruction multiplies the 32-bit elements of Rs1 with the 32-bit elements of Rs2. It then shifts the multiplication results one bit to the left and takes the most significant 32-bit results. If the shifted result is greater than  $2^{31}-1$ , it is saturated to  $2^{31}-1$  and the OV flag is set to 1. The final element result is written to Rd. The 32-bit elements of Rs1 and Rs2 are treated as signed integers. The `.u` form of the instruction additionally rounds up the 64-bit multiplication results by adding a 1 to bit 30 before the shift and saturation operations.

**Operations:**

```
if ((0x80000000 != Rs1.W[x]) | (0x80000000 != Rs2.W[x])) {
    Mres[x][63:0] = Rs1.W[x] * Rs2.W[x];
    if (`.u` form) {
        Round[x][33:0] = Mres[x][63:30] + 1;
    }
}
```

(continues on next page)

(continued from previous page)

```

    Rd.W[x] = Round[x][32:1];
  } else {
    Rd.W[x] = Mres[x][62:31];
  }
} else {
  Rd.W[x] = 0x7fffffff;
  OV = 1;
}
for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

#### Parameters

- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMMUL(long a, long b)**  
SMMUL (SIMD MSW Signed Multiply Word)

**Type:** SIMD

#### Syntax:

```

SMMUL Rd, Rs1, Rs2
SMMUL.u Rd, Rs1, Rs2

```

**Purpose:** Multiply the 32-bit signed integer elements of two registers and write the most significant 32-bit results to the corresponding 32-bit elements of a register. The `.u` form performs an additional rounding up operation on the multiplication results before taking the most significant 32-bit part of the results.

**Description:** This instruction multiplies the 32-bit elements of Rs1 with the 32-bit elements of Rs2 and writes the most significant 32-bit multiplication results to the corresponding 32-bit elements of Rd. The 32-bit elements of Rs1 and Rs2 are treated as signed integers. The `.u` form of the instruction rounds up the most significant 32-bit of the 64-bit multiplication results by adding a 1 to bit 31 of the results.

- For `smmul/RV32` instruction, it is an alias to `mulh/RV32` instruction.

#### Operations:

```

Mres[x][63:0] = Rs1.W[x] * Rs2.W[x];
if (`.u` form) {
  Round[x][32:0] = Mres[x][63:31] + 1;
  Rd.W[x] = Round[x][32:1];
} else {
  Rd.W[x] = Mres[x][63:32];
}
for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

#### Parameters

- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMMUL\_U(long a, long b)**  
 SMMUL.u (SIMD MSW Signed Multiply Word with Rounding)

**Type:** SIMD

**Syntax:**

```
SMMUL Rd, Rs1, Rs2
SMMUL.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit signed integer elements of two registers and write the most significant 32-bit results to the corresponding 32-bit elements of a register. The .u form performs an additional rounding up operation on the multiplication results before taking the most significant 32-bit part of the results.

**Description:** This instruction multiplies the 32-bit elements of Rs1 with the 32-bit elements of Rs2 and writes the most significant 32-bit multiplication results to the corresponding 32-bit elements of Rd. The 32-bit elements of Rs1 and Rs2 are treated as signed integers. The .u form of the instruction rounds up the most significant 32-bit of the 64-bit multiplication results by adding a 1 to bit 31 of the results.

- For smmul/RV32 instruction, it is an alias to mulh/RV32 instruction.

**Operations:**

```
Mres[x][63:0] = Rs1.W[x] * Rs2.W[x];
if (`.u` form) {
    Round[x][32:0] = Mres[x][63:31] + 1;
    Rd.W[x] = Round[x][32:1];
} else {
    Rd.W[x] = Mres[x][63:32];
}
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

## Signed MSW 32x16 Multiply and Add Instructions

```
__STATIC_FORCEINLINE long __RV_KMMAWB(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMAWB_U(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMAWB2(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMAWB2_U(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMAWT(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMAWT_U(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMAWT2(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMAWT2_U(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMWB2(long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMWB2_U(long a, unsigned long b)
```

```

__STATIC_FORCEINLINE long __RV_KMMWT2(long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMMWT2_U(long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMMWB(long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMMWB_U(long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMMWT(long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMMWT_U(long a, unsigned long b)

```

group **NMSIS\_Core\_DSP\_Intrinsic\_SIGNED\_MSW\_32X16\_MAC**

Signed MSW 32x16 Multiply and Add Instructions.

there are 15 Signed MSW 32x16 Multiply and Add Instructions

## Functions

```

__STATIC_FORCEINLINE long __RV_KMMAWB(long t, unsigned long a, unsigned long b)
KMMAWB (SIMD Saturating MSW Signed Multiply Word and Bottom Half and Add)

```

**Type:** SIMD

**Syntax:**

```

KMMAWB Rd, Rs1, Rs2
KMMAWB.u Rd, Rs1, Rs2

```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the bottom 16-bit of the corresponding 32-bit elements of another register and add the most significant 32-bit results with the corresponding signed 32-bit elements of a third register. The addition result is written to the corresponding 32-bit elements of the third register. The `.u` form rounds up the multiplication results from the most significant discarded bit before the addition operations.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed bottom 16-bit content of the corresponding 32-bit elements of Rs2 and adds the most significant 32-bit multiplication results with the corresponding signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit multiplication results by adding a 1 to bit 15 of the result before the addition operations.

**Operations:**

```

Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[0];
if (` .u ` form) {
    Round[x][32:0] = Mres[x][47:15] + 1;
    res[x] = Rd.W[x] + Round[x][32:1];
} else {
    res[x] = Rd.W[x] + Mres[x][47:16];
}
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];

```

(continues on next page)

(continued from previous page)

```
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAWB U(long t, unsigned long a, unsigned long b)**  
KMMAWB.u (SIMD Saturating MSW Signed Multiply Word and Bottom Half and Add with Rounding)

**Type:** SIMD

**Syntax:**

```
KMMAWB Rd, Rs1, Rs2
KMMAWB.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the bottom 16-bit of the corresponding 32-bit elements of another register and add the most significant 32-bit results with the corresponding signed 32-bit elements of a third register. The addition result is written to the corresponding 32-bit elements of the third register. The .u form rounds up the multiplication results from the most significant discarded bit before the addition operations.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed bottom 16-bit content of the corresponding 32-bit elements of Rs2 and adds the most significant 32-bit multiplication results with the corresponding signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to the corresponding 32-bit elements of Rd. The .u form of the instruction rounds up the most significant 32-bit of the 48-bit multiplication results by adding a 1 to bit 15 of the result before the addition operations.

**Operations:**

```
Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[0];
if (` .u ` form) {
    Round[x][32:0] = Mres[x][47:15] + 1;
    res[x] = Rd.W[x] + Round[x][32:1];
} else {
    res[x] = Rd.W[x] + Mres[x][47:16];
}
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAWB2(long t, unsigned long a, unsigned long b)**  
 KMMAWB2 (SIMD Saturating MSW Signed Multiply Word and Bottom Half & 2 and Add)

**Type:** SIMD

**Syntax:**

```
KMMAWB2 Rd, Rs1, Rs2
KMMAWB2.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit elements of one register and the bottom 16-bit of the corresponding 32-bit elements of another register, double the multiplication results and add the saturated most significant 32-bit results with the corresponding signed 32-bit elements of a third register. The saturated addition result is written to the corresponding 32-bit elements of the third register. The .u form rounds up the multiplication results from the most significant discarded bit before the addition operations.

**Description:** This instruction multiplies the signed 32-bit Q31 elements of Rs1 with the signed bottom 16-bit Q15 content of the corresponding 32-bit elements of Rs2, doubles the Q46 results to Q47 numbers and adds the saturated most significant 32-bit Q31 multiplication results with the corresponding signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to the corresponding 32-bit elements of Rd. The .u form of the instruction rounds up the most significant 32-bit of the 48-bit Q47 multiplication results by adding a 1 to bit 15 (i.e., bit 14 before doubling) of the result before the addition operations.

**Operations:**

```
if ((Rs1.W[x] == 0x80000000) & (Rs2.W[x].H[0] == 0x8000)) {
    addop.W[x] = 0x7fffffff;
    OV = 1;
} else {
    Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[0];
    if (`.u` form) {
        Mres[x][47:14] = Mres[x][47:14] + 1;
    }
    addop.W[x] = Mres[x][46:15]; // doubling
}
res[x] = Rd.W[x] + addop.W[x];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**



- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAWB2\_U(long t, unsigned long a, unsigned long b)**  
 KMMAWB2.u (SIMD Saturating MSW Signed Multiply Word and Bottom Half & 2 and Add with Rounding)

**Type:** SIMD

**Syntax:**

```
KMMAWB2 Rd, Rs1, Rs2
KMMAWB2.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit elements of one register and the bottom 16-bit of the corresponding 32-bit elements of another register, double the multiplication results and add the saturated most significant 32-bit results with the corresponding signed 32-bit elements of a third register. The saturated addition result is written to the corresponding 32-bit elements of the third register. The .u form rounds up the multiplication results from the most significant discarded bit before the addition operations.

**Description:** This instruction multiplies the signed 32-bit Q31 elements of Rs1 with the signed bottom 16-bit Q15 content of the corresponding 32-bit elements of Rs2, doubles the Q46 results to Q47 numbers and adds the saturated most significant 32-bit Q31 multiplication results with the corresponding signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to the corresponding 32-bit elements of Rd. The .u form of the instruction rounds up the most significant 32-bit of the 48-bit Q47 multiplication results by adding a 1 to bit 15 (i.e., bit 14 before doubling) of the result before the addition operations.

**Operations:**

```
if ((Rs1.W[x] == 0x80000000) & (Rs2.W[x].H[0] == 0x8000)) {
    addop.W[x] = 0x7fffffff;
    OV = 1;
} else {
    Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[0];
    if (`u` form) {
        Mres[x][47:14] = Mres[x][47:14] + 1;
    }
    addop.W[x] = Mres[x][46:15]; // doubling
}
res[x] = Rd.W[x] + addop.W[x];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAWT(long t, unsigned long a, unsigned long b)**  
KMMAWT (SIMD Saturating MSW Signed Multiply Word and Top Half and Add)

**Type:** SIMD

**Syntax:**

```
KMMAWT Rd, Rs1, Rs2
KMMAWT.u Rd Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the signed top 16-bit of the corresponding 32-bit elements of another register and add the most significant 32-bit results with the corresponding signed 32-bit elements of a third register. The addition results are written to the corresponding 32-bit elements of the third register. The .u form rounds up the multiplication results from the most significant discarded bit before the addition operations.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed top 16-bit of the corresponding 32-bit elements of Rs2 and adds the most significant 32-bit multiplication results with the corresponding signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to the corresponding 32-bit elements of Rd. The .u form of the instruction rounds up the most significant 32-bit of the 48-bit multiplication results by adding a 1 to bit 15 of the result before the addition operations.

**Operations:**

```
Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[1];
if (` .u ` form) {
    Round[x][32:0] = Mres[x][47:15] + 1;
    res[x] = Rd.W[x] + Round[x][32:1];
} else {
    res[x] = Rd.W[x] + Mres[x][47:16];
}
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1..0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAWT\_U(long t, unsigned long a, unsigned long b)**  
 KMMAWT.u (SIMD Saturating MSW Signed Multiply Word and Top Half and Add with Rounding)

**Type:** SIMD

**Syntax:**

```
KMMAWT Rd, Rs1, Rs2
KMMAWT.u Rd Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the signed top 16-bit of the corresponding 32-bit elements of another register and add the most significant 32-bit results with the corresponding signed 32-bit elements of a third register. The addition results are written to the corresponding 32-bit elements of the third register. The .u form rounds up the multiplication results from the most significant discarded bit before the addition operations.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed top 16-bit of the corresponding 32-bit elements of Rs2 and adds the most significant 32-bit multiplication results with the corresponding signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to the corresponding 32-bit elements of Rd. The .u form of the instruction rounds up the most significant 32-bit of the 48-bit multiplication results by adding a 1 to bit 15 of the result before the addition operations.

**Operations:**

```
Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[1];
if (`.u` form) {
    Round[x][32:0] = Mres[x][47:15] + 1;
    res[x] = Rd.W[x] + Round[x][32:1];
} else {
    res[x] = Rd.W[x] + Mres[x][47:16];
}
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAWT2(long t, unsigned long a, unsigned long b)**  
 KMMAWT2 (SIMD Saturating MSW Signed Multiply Word and Top Half & 2 and Add)

**Type:** SIMD

**Syntax:**

```
KMMAWT2 Rd, Rs1, Rs2
KMMAWT2.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit elements of one register and the top 16-bit of the corresponding 32-bit elements of another register, double the multiplication results and add the saturated most significant 32-bit results with the corresponding signed 32-bit elements of a third register. The saturated addition result is written to the corresponding 32-bit elements of the third register. The `.u` form rounds up the multiplication results from the most significant discarded bit before the addition operations.

**Description:** This instruction multiplies the signed 32-bit Q31 elements of Rs1 with the signed top 16-bit Q15 content of the corresponding 32-bit elements of Rs2, doubles the Q46 results to Q47 numbers and adds the saturated most significant 32-bit Q31 multiplication results with the corresponding signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit Q47 multiplication results by adding a 1 to bit 15 (i.e., bit 14 before doubling) of the result before the addition operations.

#### Operations:

```
if ((Rs1.W[x] == 0x80000000) & (Rs2.W[x].H[1] == 0x8000)) {
    addop.W[x] = 0x7fffffff;
    OV = 1;
} else {
    Mres[x][47:0] = Rs1.W[x] s* Rs2.W[x].H[1];
    if (`.u` form) {
        Mres[x][47:14] = Mres[x][47:14] + 1;
    }
    addop.W[x] = Mres[x][46:15]; // doubling
}
res[x] = Rd.W[x] + addop.W[x];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

#### Parameters

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMAWT2\_U(long t, unsigned long a, unsigned long b)**  
KMMAWT2.u (SIMD Saturating MSW Signed Multiply Word and Top Half & 2 and Add with Rounding)

**Type:** SIMD

**Syntax:**

```
KMMAWT2 Rd, Rs1, Rs2
KMMAWT2.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit elements of one register and the top 16-bit of the corresponding 32-bit elements of another register, double the multiplication results and add the saturated most significant 32-bit results with the corresponding signed 32-bit elements of a third register. The saturated addition result is written to the corresponding 32-bit elements of the third register. The `.u` form rounds up the multiplication results from the most significant discarded bit before the addition operations.

**Description:** This instruction multiplies the signed 32-bit Q31 elements of Rs1 with the signed top 16-bit Q15 content of the corresponding 32-bit elements of Rs2, doubles the Q46 results to Q47 numbers and adds the saturated most significant 32-bit Q31 multiplication results with the corresponding signed 32-bit elements of Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit Q47 multiplication results by adding a 1 to bit 15 (i.e., bit 14 before doubling) of the result before the addition operations.

#### Operations:

```
if ((Rs1.W[x] == 0x80000000) & (Rs2.W[x].H[1] == 0x8000)) {
    addop.W[x] = 0x7fffffff;
    OV = 1;
} else {
    Mres[x][47:0] = Rs1.W[x] s* Rs2.W[x].H[1];
    if (`.u` form) {
        Mres[x][47:14] = Mres[x][47:14] + 1;
    }
    addop.W[x] = Mres[x][46:15]; // doubling
}
res[x] = Rd.W[x] + addop.W[x];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

#### Parameters

- [in] `t`: long type of value stored in `t`
- [in] `a`: unsigned long type of value stored in `a`
- [in] `b`: unsigned long type of value stored in `b`

```
__STATIC_FORCEINLINE long __RV_KMMWB2(long a, unsigned long b)
KMMWB2 (SIMD Saturating MSW Signed Multiply Word and Bottom Half & 2)
```

**Type:** SIMD

**Syntax:**

```
KMMWB2 Rd, Rs1, Rs2
KMMWB2.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the bottom 16-bit of the corresponding 32-bit elements of another register, double the multiplication results and write the saturated most significant 32-bit results to the corresponding 32-bit elements of a register. The `.u` form rounds up the results from the most significant discarded bit.

**Description:** This instruction multiplies the signed 32-bit Q31 elements of Rs1 with the signed bottom 16-bit Q15 content of the corresponding 32-bit elements of Rs2, doubles the Q46 results to Q47 numbers and writes the saturated most significant 32-bit Q31 multiplication results to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit Q47 multiplication results by adding a 1 to bit 15 (i.e., bit 14 before doubling) of the results.

**Operations:**

```
if ((Rs1.W[x] == 0x80000000) & (Rs2.W[x].H[0] == 0x8000)) {
    Rd.W[x] = 0x7fffffff;
    OV = 1;
} else {
    Mres[x][47:0] = Rs1.W[x] s* Rs2.W[x].H[0];
    if (`.u` form) {
        Round[x][32:0] = Mres[x][46:14] + 1;
        Rd.W[x] = Round[x][32:1];
    } else {
        Rd.W[x] = Mres[x][46:15];
    }
}
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMWB2\_U(long a, unsigned long b)**

KMMWB2.u (SIMD Saturating MSW Signed Multiply Word and Bottom Half & 2 with Rounding)

**Type:** SIMD

**Syntax:**

```
KMMWB2 Rd, Rs1, Rs2
KMMWB2.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the bottom 16-bit of the corresponding 32-bit elements of another register, double the multiplication results and write the saturated most significant 32-bit results to the corresponding 32-bit elements of a register. The `.u` form rounds up the results from the most significant discarded bit.

**Description:** This instruction multiplies the signed 32-bit Q31 elements of Rs1 with the signed bottom 16-bit Q15 content of the corresponding 32-bit elements of Rs2, doubles the Q46 results to Q47 numbers and writes the saturated most significant 32-bit Q31 multiplication results to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit Q47 multiplication results by adding a 1 to bit 15 (i.e., bit 14 before doubling) of the results.

**Operations:**

```

if ((Rs1.W[x] == 0x80000000) & (Rs2.W[x].H[0] == 0x8000)) {
    Rd.W[x] = 0x7fffffff;
    OV = 1;
} else {
    Mres[x][47:0] = Rs1.W[x] s* Rs2.W[x].H[0];
    if (`.u` form) {
        Round[x][32:0] = Mres[x][46:14] + 1;
        Rd.W[x] = Round[x][32:1];
    } else {
        Rd.W[x] = Mres[x][46:15];
    }
}
}
for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMWT2(long a, unsigned long b)**  
 KMMWT2 (SIMD Saturating MSW Signed Multiply Word and Top Half & 2)

**Type:** SIMD

**Syntax:**

```

KMMWT2 Rd, Rs1, Rs2
KMMWT2.u Rd, Rs1, Rs2

```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the top 16-bit of the corresponding 32-bit elements of another register, double the multiplication results and write the saturated most significant 32-bit results to the corresponding 32-bit elements of a register. The `.u` form rounds up the results from the most significant discarded bit.

**Description:** This instruction multiplies the signed 32-bit Q31 elements of Rs1 with the signed top 16-bit Q15 content of the corresponding 32-bit elements of Rs2, doubles the Q46 results to Q47 numbers and writes the saturated most significant 32-bit Q31 multiplication results to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit Q47 multiplication results by adding a 1 to bit 15 (i.e., bit 14 before doubling) of the results.

**Operations:**

```

if ((Rs1.W[x] == 0x80000000) & (Rs2.W[x].H[1] == 0x8000)) {
    Rd.W[x] = 0x7fffffff;
    OV = 1;
} else {
    Mres[x][47:0] = Rs1.W[x] s* Rs2.W[x].H[1];
    if (`.u` form) {
        Round[x][32:0] = Mres[x][46:14] + 1;
        Rd.W[x] = Round[x][32:1];
    } else {
        Rd.W[x] = Mres[x][46:15];
    }
}
}

```

(continues on next page)

(continued from previous page)

```
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMMWT2 U(long a, unsigned long b)**  
KMMWT2.u (SIMD Saturating MSW Signed Multiply Word and Top Half & 2 with Rounding)

**Type:** SIMD

**Syntax:**

```
KMMWT2 Rd, Rs1, Rs2
KMMWT2.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the top 16-bit of the corresponding 32-bit elements of another register, double the multiplication results and write the saturated most significant 32-bit results to the corresponding 32-bit elements of a register. The .u form rounds up the results from the most significant discarded bit.

**Description:** This instruction multiplies the signed 32-bit Q31 elements of Rs1 with the signed top 16-bit Q15 content of the corresponding 32-bit elements of Rs2, doubles the Q46 results to Q47 numbers and writes the saturated most significant 32-bit Q31 multiplication results to the corresponding 32-bit elements of Rd. The .u form of the instruction rounds up the most significant 32-bit of the 48-bit Q47 multiplication results by adding a 1 to bit 15 (i.e., bit 14 before doubling) of the results.

**Operations:**

```
if ((Rs1.W[x] == 0x80000000) & (Rs2.W[x].H[1] == 0x8000)) {
    Rd.W[x] = 0x7fffffff;
    OV = 1;
} else {
    Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[1];
    if (`u` form) {
        Round[x][32:0] = Mres[x][46:14] + 1;
        Rd.W[x] = Round[x][32:1];
    } else {
        Rd.W[x] = Mres[x][46:15];
    }
}
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMMWB(long a, unsigned long b)**  
SMMWB (SIMD MSW Signed Multiply Word and Bottom Half)



**Type:** SIMD

**Syntax:**

```
SMMWB Rd, Rs1, Rs2
SMMWB.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the bottom 16-bit of the corresponding 32-bit elements of another register, and write the most significant 32-bit results to the corresponding 32-bit elements of a register. The `.u` form rounds up the results from the most significant discarded bit.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed bottom 16-bit content of the corresponding 32-bit elements of Rs2 and writes the most significant 32-bit multiplication results to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit multiplication results by adding a 1 to bit 15 of the results.

**Operations:**

```
Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[0];
if (`u` form) {
    Round[x][32:0] = Mres[x][47:15] + 1;
    Rd.W[x] = Round[x][32:1];
} else {
    Rd.W[x] = Mres[x][47:16];
}
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_SMMWB_U(long a, unsigned long b)
SMMWB.u (SIMD MSW Signed Multiply Word and Bottom Half with Rounding)
```

**Type:** SIMD

**Syntax:**

```
SMMWB Rd, Rs1, Rs2
SMMWB.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the bottom 16-bit of the corresponding 32-bit elements of another register, and write the most significant 32-bit results to the corresponding 32-bit elements of a register. The `.u` form rounds up the results from the most significant discarded bit.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the signed bottom 16-bit content of the corresponding 32-bit elements of Rs2 and writes the most significant 32-bit multiplication results to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit multiplication results by adding a 1 to bit 15 of the results.

**Operations:**

```

Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[0];
if (`.u` form) {
    Round[x][32:0] = Mres[x][47:15] + 1;
    Rd.W[x] = Round[x][32:1];
} else {
    Rd.W[x] = Mres[x][47:16];
}
for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMMWT(long a, unsigned long b)**  
 SMMWT (SIMD MSW Signed Multiply Word and Top Half)

**Type:** SIMD

**Syntax:**

```

SMMWT Rd, Rs1, Rs2
SMMWT.u Rd, Rs1, Rs2

```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the top 16-bit of the corresponding 32-bit elements of another register, and write the most significant 32-bit results to the corresponding 32-bit elements of a register. The `.u` form rounds up the results from the most significant discarded bit.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the top signed 16-bit content of the corresponding 32-bit elements of Rs2 and writes the most significant 32-bit multiplication results to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit multiplication results by adding a 1 to bit 15 of the results.

**Operations:**

```

Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[1];
if (`.u` form) {
    Round[x][32:0] = Mres[x][47:15] + 1;
    Rd.W[x] = Round[x][32:1];
} else {
    Rd.W[x] = Mres[x][47:16];
}
for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMMWT\_U(long a, unsigned long b)**  
 SMMWT.u (SIMD MSW Signed Multiply Word and Top Half with Rounding)

**Type:** SIMD

**Syntax:**

```
SMMWT Rd, Rs1, Rs2
SMMWT.u Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit integer elements of one register and the top 16-bit of the corresponding 32-bit elements of another register, and write the most significant 32-bit results to the corresponding 32-bit elements of a register. The `.u` form rounds up the results from the most significant discarded bit.

**Description:** This instruction multiplies the signed 32-bit elements of Rs1 with the top signed 16-bit content of the corresponding 32-bit elements of Rs2 and writes the most significant 32-bit multiplication results to the corresponding 32-bit elements of Rd. The `.u` form of the instruction rounds up the most significant 32-bit of the 48-bit multiplication results by adding a 1 to bit 15 of the results.

**Operations:**

```
Mres[x][47:0] = Rs1.W[x] * Rs2.W[x].H[1];
if (`.u` form) {
    Round[x][32:0] = Mres[x][47:15] + 1;
    Rd.W[x] = Round[x][32:1];
} else {
    Rd.W[x] = Mres[x][47:16];
}
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a
- [in] b: unsigned long type of value stored in b

**Signed 16-bit Multiply 32-bit Add/Subtract Instructions**

```
__STATIC_FORCEINLINE long __RV_KMABB(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMABT(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMATT(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMADA(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMAXDA(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMADS(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMADRS(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMAXDS(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMDA(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMXDA(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMSDA(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMSXDA(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMBB16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMBT16(unsigned long a, unsigned long b)
```

```

__STATIC_FORCEINLINE long __RV_SMTT16(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMDS(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMDRS(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMXDS(unsigned long a, unsigned long b)

```

group **NMSIS\_Core\_DSP\_Intrinsic\_SIGNED\_16B\_MULT\_32B\_ADDSUB**

Signed 16-bit Multiply 32-bit Add/Subtract Instructions.

there are 18 Signed 16-bit Multiply 32-bit Add/Subtract Instructions

## Functions

```

__STATIC_FORCEINLINE long __RV_KMABB(long t, unsigned long a, unsigned long b)
KMABB (SIMD Saturating Signed Multiply Bottom Halfs & Add)

```

Type: SIMD

Syntax:

```

KMABB Rd, Rs1, Rs2
KMABT Rd, Rs1, Rs2
KMATT Rd, Rs1, Rs2

```

**Purpose:** Multiply the signed 16-bit content of 32-bit elements in a register with the 16-bit content of 32-bit elements in another register and add the result to the content of 32-bit elements in the third register. The addition result may be saturated and is written to the third register.

- KMABB:  $rd.W[x] + bottom * bottom$  (per 32-bit element)
- KMABT  $rd.W[x] + bottom * top$  (per 32-bit element)
- KMATT  $rd.W[x] + top * top$  (per 32-bit element)

**Description:** For the KMABB instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. For the KMABT instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. For the KMATT instruction, it multiplies the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. The multiplication result is added to the content of 32-bit elements in Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to

1. The results after saturation are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

## Operations:

```

res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[0]); // KMABB
res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[1]); // KMABT
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[1]); // KMATT
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];

```

(continues on next page)

(continued from previous page)

```

for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMABT(long t, unsigned long a, unsigned long b)**  
KMABT (SIMD Saturating Signed Multiply Bottom & Top Halfs & Add)

**Type:** SIMD

**Syntax:**

```

KMABB Rd, Rs1, Rs2
KMABT Rd, Rs1, Rs2
KMATT Rd, Rs1, Rs2

```

**Purpose:** Multiply the signed 16-bit content of 32-bit elements in a register with the 16-bit content of 32-bit elements in another register and add the result to the content of 32-bit elements in the third register. The addition result may be saturated and is written to the third register.

- KMABB:  $rd.W[x] + bottom * bottom$  (per 32-bit element)
- KMABT:  $rd.W[x] + bottom * top$  (per 32-bit element)
- KMATT:  $rd.W[x] + top * top$  (per 32-bit element)

**Description:** For the KMABB instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. For the KMABT instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. For the KMATT instruction, it multiplies the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. The multiplication result is added to the content of 32-bit elements in Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to

1. The results after saturation are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```

res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[0]); // KMABB
res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[1]); // KMABT
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[1]); // KMATT
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_KMATT(long t, unsigned long a, unsigned long b)
KMATT (SIMD Saturating Signed Multiply Top Halves & Add)
```

**Type:** SIMD

**Syntax:**

```
KMABB Rd, Rs1, Rs2
KMABT Rd, Rs1, Rs2
KMATT Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 16-bit content of 32-bit elements in a register with the 16-bit content of 32-bit elements in another register and add the result to the content of 32-bit elements in the third register. The addition result may be saturated and is written to the third register.

- KMABB:  $rd.W[x] + bottom * bottom$  (per 32-bit element)
- KMABT  $rd.W[x] + bottom * top$  (per 32-bit element)
- KMATT  $rd.W[x] + top * top$  (per 32-bit element)

**Description:** For the KMABB instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. For the KMABT instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. For the KMATT instruction, it multiplies the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. The multiplication result is added to the content of 32-bit elements in Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to

1. The results after saturation are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[0]); // KMABB
res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[1]); // KMABT
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[1]); // KMATT
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMADA(long t, unsigned long a, unsigned long b)**  
KMADA (SIMD Saturating Signed Multiply Two Halfs and Two Adds)

**Type:** SIMD

**Syntax:**

```
KMADA Rd, Rs1, Rs2
KMAXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from 32-bit elements in two registers; and then adds the two 32-bit results and 32-bit elements in a third register together. The addition result may be saturated.

- KMADA:  $rd.W[x] + top * top + bottom * bottom$  (per 32-bit element)
- KMAXDA:  $rd.W[x] + top * bottom + bottom * top$  (per 32-bit element)

**Description:** For the ‘KMADA instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2 and then adds the result to the result of multiplying the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. For the KMAXDA instruction, it multiplies the top 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2 and then adds the result to the result of multiplying the bottom 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. The result is added to the content of 32-bit elements in Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The 32-bit results after saturation are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
// KMADA
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[1]) + (Rs1.W[x].H[0] * Rs2.
↪W[x].H[0]);
// KMAXDA
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[0]) + (Rs1.W[x].H[0] * Rs2.
↪W[x].H[1]);
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMAXDA(long t, unsigned long a, unsigned long b)**  
KMAXDA (SIMD Saturating Signed Crossed Multiply Two Halfs and Two Adds)

**Type:** SIMD**Syntax:**

```
KMADA Rd, Rs1, Rs2
KMAXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from 32-bit elements in two registers; and then adds the two 32-bit results and 32-bit elements in a third register together. The addition result may be saturated.

- KMADA:  $rd.W[x] + top*top + bottom*bottom$  (per 32-bit element)
- KMAXDA:  $rd.W[x] + top*bottom + bottom*top$  (per 32-bit element)

**Description:** For the ‘KMADA instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2 and then adds the result to the result of multiplying the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. For the KMAXDA instruction, it multiplies the top 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2 and then adds the result to the result of multiplying the bottom 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. The result is added to the content of 32-bit elements in Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The 32-bit results after saturation are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
// KMADA
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[1]) + (Rs1.W[x].H[0] * Rs2.
↪W[x].H[0]);
// KMAXDA
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[0]) + (Rs1.W[x].H[0] * Rs2.
↪W[x].H[1]);
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_KMADS(long t, unsigned long a, unsigned long b)
KMADS (SIMD Saturating Signed Multiply Two Halfs & Subtract & Add)
```

**Type:** SIMD**Syntax:**



```

KMADS Rd, Rs1, Rs2
KMADRS Rd, Rs1, Rs2
KMAXDS Rd, Rs1, Rs2

```

**Purpose:** Do two signed 16-bit multiplications from 32-bit elements in two registers; and then perform a subtraction operation between the two 32-bit results. Then add the subtraction result to the corresponding 32-bit elements in a third register. The addition result may be saturated.

- KMADS:  $rd.W[x] + (top * top - bottom * bottom)$  (per 32-bit element)
- KMADRS:  $rd.W[x] + (bottom * bottom - top * top)$  (per 32-bit element)
- KMAXDS:  $rd.W[x] + (top * bottom - bottom * top)$  (per 32-bit element)

**Description:** For the KMADS instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. For the KMADRS instruction, it multiplies the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. For the KMAXDS instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. The subtraction result is then added to the content of the corresponding 32-bit elements in Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The 32-bit results after saturation are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

#### Operations:

```

// KMADS
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[1]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[0]);
// KMADRS
res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[0]) - (Rs1.W[x].H[1] * Rs2.
↪W[x].H[1]);
// KMAXDS
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[0]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[1]);
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

#### Parameters

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

---

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMADRS(long t, unsigned long a, unsigned long b)**  
 KMADRS (SIMD Saturating Signed Multiply Two Halfs & Reverse Subtract & Add)

**Type:** SIMD

**Syntax:**

```
KMADS Rd, Rs1, Rs2
KMADRS Rd, Rs1, Rs2
KMAXDS Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from 32-bit elements in two registers; and then perform a subtraction operation between the two 32-bit results. Then add the subtraction result to the corresponding 32-bit elements in a third register. The addition result may be saturated.

- KMADS:  $rd.W[x] + (top * top - bottom * bottom)$  (per 32-bit element)
- KMADRS:  $rd.W[x] + (bottom * bottom - top * top)$  (per 32-bit element)
- KMAXDS:  $rd.W[x] + (top * bottom - bottom * top)$  (per 32-bit element)

**Description:** For the KMADS instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. For the KMADRS instruction, it multiplies the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. For the KMAXDS instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. The subtraction result is then added to the content of the corresponding 32-bit elements in Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The 32-bit results after saturation are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
// KMADS
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[1]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[0]);
// KMADRS
res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[0]) - (Rs1.W[x].H[1] * Rs2.
↪W[x].H[1]);
// KMAXDS
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[0]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[1]);
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMAXDS(long t, unsigned long a, unsigned long b)**  
KMAXDS (SIMD Saturating Signed Crossed Multiply Two Halfs & Subtract & Add)

**Type:** SIMD

**Syntax:**

```
KMADS Rd, Rs1, Rs2
KMADRS Rd, Rs1, Rs2
KMAXDS Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from 32-bit elements in two registers; and then perform a subtraction operation between the two 32-bit results. Then add the subtraction result to the corresponding 32-bit elements in a third register. The addition result may be saturated.

- KMADS:  $rd.W[x] + (top * top - bottom * bottom)$  (per 32-bit element)
- KMADRS:  $rd.W[x] + (bottom * bottom - top * top)$  (per 32-bit element)
- KMAXDS:  $rd.W[x] + (top * bottom - bottom * top)$  (per 32-bit element)

**Description:** For the KMADS instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2. For the KMADRS instruction, it multiplies the top 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. For the KMAXDS instruction, it multiplies the bottom 16-bit content of 32-bit elements in Rs1 with the top 16-bit content of 32-bit elements in Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of 32-bit elements in Rs1 with the bottom 16-bit content of 32-bit elements in Rs2. The subtraction result is then added to the content of the corresponding 32-bit elements in Rd. If the addition result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The 32-bit results after saturation are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
// KMADS
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[1]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[0]);
// KMADRS
res[x] = Rd.W[x] + (Rs1.W[x].H[0] * Rs2.W[x].H[0]) - (Rs1.W[x].H[1] * Rs2.
↪W[x].H[1]);
// KMAXDS
res[x] = Rd.W[x] + (Rs1.W[x].H[1] * Rs2.W[x].H[0]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[1]);
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
```

(continues on next page)

(continued from previous page)

```

for RV32: x=0
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMDA(unsigned long a, unsigned long b)**  
KMDA (SIMD Signed Multiply Two Halfs and Add)

**Type:** SIMD

**Syntax:**

```

KMDA Rd, Rs1, Rs2
KMXDA Rd, Rs1, Rs2

```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then adds the two 32-bit results together. The addition result may be saturated.

- KMDA: top\*top + bottom\*bottom (per 32-bit element)
- KMXDA: top\*bottom + bottom\*top (per 32-bit element)

**Description:** For the KMDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then adds the result to the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the KMXDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then adds the result to the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The addition result is checked for saturation. If saturation happens, the result is saturated to  $2^{31}-1$ . The final results are written to Rd. The 16-bit contents are treated as signed integers.

**Operations:**

```

if Rs1.W[x] != 0x80008000 or (Rs2.W[x] != 0x80008000 { // KMDA Rd.
    ↪W[x] = Rs1.W[x].H[1] *
Rs2.W[x].H[1] + (Rs1.W[x].H[0] * Rs2.W[x].H[0]; // KMXDA Rd.W[x] = Rs1.W[x].
    ↪H[1] * Rs2.W[x].H[0])
+ (Rs1.W[x].H[0] * Rs2.W[x].H[1]; } else { Rd.W[x] = 0x7fffffff; OV_
    ↪ = 1; } for RV32: x=0 for RV64:
x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMXDA(unsigned long a, unsigned long b)**  
KMXDA (SIMD Signed Crossed Multiply Two Halfs and Add)

**Type:** SIMD

**Syntax:**

```
KMDA Rd, Rs1, Rs2
KMXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then adds the two 32-bit results together. The addition result may be saturated.

- KMDA: top\*top + bottom\*bottom (per 32-bit element)
- KMXDA: top\*bottom + bottom\*top (per 32-bit element)

**Description:** For the KMDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then adds the result to the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the KMXDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then adds the result to the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The addition result is checked for saturation. If saturation happens, the result is saturated to  $2^{31}-1$ . The final results are written to Rd. The 16-bit contents are treated as signed integers.

**Operations:**

```
if Rs1.W[x] != 0x80008000) or (Rs2.W[x] != 0x80008000 { // KMDA Rd.
    W[x] = Rs1.W[x].H[1] *
Rs2.W[x].H[1]) + (Rs1.W[x].H[0] * Rs2.W[x].H[0]; // KMXDA Rd.W[x] = Rs1.W[x].
    H[1] * Rs2.W[x].H[0])
+ (Rs1.W[x].H[0] * Rs2.W[x].H[1]; } else { Rd.W[x] = 0x7fffffff; OV
    = 1; } for RV32: x=0 for RV64:
x=1...0
```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_KMSDA(long t, unsigned long a, unsigned long b)
KMSDA (SIMD Saturating Signed Multiply Two Halfs & Add & Subtract)
```

**Type:** SIMD

**Syntax:**

```
KMSDA Rd, Rs1, Rs2
KMSXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then subtracts the two 32-bit results from the corresponding 32-bit elements of a third register. The subtraction result may be saturated.

- KMSDA: rd.W[x] - top\*top - bottom\*bottom (per 32-bit element)
- KMSXDA: rd.W[x] - top\*bottom - bottom\*top (per 32-bit element)

**Description:** For the KMSDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the KMSXDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content

of the 32-bit elements of Rs2 and multiplies the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The two 32-bit multiplication results are then subtracted from the content of the corresponding 32-bit elements of Rd. If the subtraction result is beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to Rd. The 16-bit contents are treated as signed integers.

#### Operations:

```
// KMSDA
res[x] = Rd.W[x] - (Rs1.W[x].H[1] * Rs2.W[x].H[1]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[0]);
// KMSXDA
res[x] = Rd.W[x] - (Rs1.W[x].H[1] * Rs2.W[x].H[0]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[1]);
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

#### Parameters

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMSXDA(long t, unsigned long a, unsigned long b)**  
KMSXDA (SIMD Saturating Signed Crossed Multiply Two Halfs & Add & Subtract)

**Type:** SIMD

#### Syntax:

```
KMSDA Rd, Rs1, Rs2
KMSXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then subtracts the two 32-bit results from the corresponding 32-bit elements of a third register. The subtraction result may be saturated.

- KMSDA:  $rd.W[x] - top*top - bottom*bottom$  (per 32-bit element)
- KMSXDA:  $rd.W[x] - top*bottom - bottom*top$  (per 32-bit element)

**Description:** For the KMSDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the KMSXDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and multiplies the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The two 32-bit multiplication results are then subtracted from the content of the corresponding 32-bit elements of Rd. If the subtraction result is beyond

the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), it is saturated to the range and the OV bit is set to 1. The results after saturation are written to Rd. The 16-bit contents are treated as signed integers.

#### Operations:

```
// KMSDA
res[x] = Rd.W[x] - (Rs1.W[x].H[1] * Rs2.W[x].H[1]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[0]);
// KMSXDA
res[x] = Rd.W[x] - (Rs1.W[x].H[1] * Rs2.W[x].H[0]) - (Rs1.W[x].H[0] * Rs2.
↪W[x].H[1]);
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in long type

#### Parameters

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMBB16(unsigned long a, unsigned long b)**  
SMBB16 (SIMD Signed Multiply Bottom Half & Bottom Half)

**Type:** SIMD

#### Syntax:

```
SMBB16 Rd, Rs1, Rs2
SMBT16 Rd, Rs1, Rs2
SMTT16 Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 16-bit content of the 32-bit elements of a register with the signed 16-bit content of the 32-bit elements of another register and write the result to a third register.

- SMBB16:  $W[x].bottom * W[x].bottom$
- SMBT16:  $W[x].bottom * W[x].top$
- SMTT16:  $W[x].top * W[x].top$

**Description:** For the SMBB16 instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMBT16 instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMTT16 instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. The multiplication results are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

#### Operations:

```

Rd.W[x] = Rs1.W[x].H[0] * Rs2.W[x].H[0]; // SMBB16
Rd.W[x] = Rs1.W[x].H[0] * Rs2.W[x].H[1]; // SMBT16
Rd.W[x] = Rs1.W[x].H[1] * Rs2.W[x].H[1]; // SMTT16
for RV32: x=0,
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```

__STATIC_FORCEINLINE long __RV_SMBT16(unsigned long a, unsigned long b)
SMBT16 (SIMD Signed Multiply Bottom Half & Top Half)

```

**Type:** SIMD

**Syntax:**

```

SMBB16 Rd, Rs1, Rs2
SMBT16 Rd, Rs1, Rs2
SMTT16 Rd, Rs1, Rs2

```

**Purpose:** Multiply the signed 16-bit content of the 32-bit elements of a register with the signed 16-bit content of the 32-bit elements of another register and write the result to a third register.

- SMBB16:  $W[x].bottom * W[x].bottom$
- SMBT16:  $W[x].bottom * W[x].top$
- SMTT16:  $W[x].top * W[x].top$

**Description:** For the SMBB16 instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMBT16 instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMTT16 instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. The multiplication results are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```

Rd.W[x] = Rs1.W[x].H[0] * Rs2.W[x].H[0]; // SMBB16
Rd.W[x] = Rs1.W[x].H[0] * Rs2.W[x].H[1]; // SMBT16
Rd.W[x] = Rs1.W[x].H[1] * Rs2.W[x].H[1]; // SMTT16
for RV32: x=0,
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```

__STATIC_FORCEINLINE long __RV_SMTT16(unsigned long a, unsigned long b)
SMTT16 (SIMD Signed Multiply Top Half & Top Half)

```

**Type:** SIMD



**Syntax:**

```
SMBB16 Rd, Rs1, Rs2
SMBT16 Rd, Rs1, Rs2
SMTT16 Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 16-bit content of the 32-bit elements of a register with the signed 16-bit content of the 32-bit elements of another register and write the result to a third register.

- SMBB16:  $W[x].bottom * W[x].bottom$
- SMBT16:  $W[x].bottom * W[x].top$
- SMTT16:  $W[x].top * W[x].top$

**Description:** For the SMBB16 instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMBT16 instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMTT16 instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. The multiplication results are written to Rd. The 16-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
Rd.W[x] = Rs1.W[x].H[0] * Rs2.W[x].H[0]; // SMBB16
Rd.W[x] = Rs1.W[x].H[0] * Rs2.W[x].H[1]; // SMBT16
Rd.W[x] = Rs1.W[x].H[1] * Rs2.W[x].H[1]; // SMTT16
for RV32: x=0,
for RV64: x=1..0
```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMDS(unsigned long a, unsigned long b)**  
SMDS (SIMD Signed Multiply Two Halfs and Subtract)

**Type:** SIMD

**Syntax:**

```
SMDS Rd, Rs1, Rs2
SMDRS Rd, Rs1, Rs2
SMXDS Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then perform a subtraction operation between the two 32-bit results.

- SMDS:  $top * top - bottom * bottom$  (per 32-bit element)
- SMDRS:  $bottom * bottom - top * top$  (per 32-bit element)
- SMXDS:  $top * bottom - bottom * top$  (per 32-bit element)

**Description:** For the SMDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMDRS instruction, it multiplies the top 16-bit content of the 32-bit

elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMXDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The subtraction result is written to the corresponding 32-bit element of Rd. The 16-bit contents of multiplication are treated as signed integers.

#### Operations:

```
* SMDS:
Rd.W[x] = (Rs1.W[x].H[1] * Rs2.W[x].H[1]) - (Rs1.W[x].H[0] * Rs2.W[x].H[0]);
* SMDRS:
Rd.W[x] = (Rs1.W[x].H[0] * Rs2.W[x].H[0]) - (Rs1.W[x].H[1] * Rs2.W[x].H[1]);
* SMXDS:
Rd.W[x] = (Rs1.W[x].H[1] * Rs2.W[x].H[0]) - (Rs1.W[x].H[0] * Rs2.W[x].H[1]);
```

**Return** value stored in long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_SMDRS(unsigned long a, unsigned long b)
SMDRS (SIMD Signed Multiply Two Halfs and Reverse Subtract)
```

**Type:** SIMD

#### Syntax:

```
SMDS Rd, Rs1, Rs2
SMDRS Rd, Rs1, Rs2
SMXDS Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then perform a subtraction operation between the two 32-bit results.

- SMDS: top\*top - bottom\*bottom (per 32-bit element)
- SMDRS: bottom\*bottom - top\*top (per 32-bit element)
- SMXDS: top\*bottom - bottom\*top (per 32-bit element)

**Description:** For the SMDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMDRS instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMXDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The subtraction result is written to the corresponding 32-bit element of Rd. The 16-bit contents of multiplication are treated as signed integers.

#### Operations:

```

* SMDS:
Rd.W[x] = (Rs1.W[x].H[1] * Rs2.W[x].H[1]) - (Rs1.W[x].H[0] * Rs2.W[x].H[0]);
* SMDRS:
Rd.W[x] = (Rs1.W[x].H[0] * Rs2.W[x].H[0]) - (Rs1.W[x].H[1] * Rs2.W[x].H[1]);
* SMXDS:
Rd.W[x] = (Rs1.W[x].H[1] * Rs2.W[x].H[0]) - (Rs1.W[x].H[0] * Rs2.W[x].H[1]);

```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMXDS(unsigned long a, unsigned long b)**  
 SMXDS (SIMD Signed Crossed Multiply Two Halfs and Subtract)

**Type:** SIMD

**Syntax:**

```

SMDS Rd, Rs1, Rs2
SMDRS Rd, Rs1, Rs2
SMXDS Rd, Rs1, Rs2

```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then perform a subtraction operation between the two 32-bit results.

- SMDS: top\*top - bottom\*bottom (per 32-bit element)
- SMDRS: bottom\*bottom - top\*top (per 32-bit element)
- SMXDS: top\*bottom - bottom\*top (per 32-bit element)

**Description:** For the SMDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMDRS instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMXDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The subtraction result is written to the corresponding 32-bit element of Rd. The 16-bit contents of multiplication are treated as signed integers.

**Operations:**

```

* SMDS:
Rd.W[x] = (Rs1.W[x].H[1] * Rs2.W[x].H[1]) - (Rs1.W[x].H[0] * Rs2.W[x].H[0]);
* SMDRS:
Rd.W[x] = (Rs1.W[x].H[0] * Rs2.W[x].H[0]) - (Rs1.W[x].H[1] * Rs2.W[x].H[1]);
* SMXDS:
Rd.W[x] = (Rs1.W[x].H[1] * Rs2.W[x].H[0]) - (Rs1.W[x].H[0] * Rs2.W[x].H[1]);

```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## Partial-SIMD Miscellaneous Instructions

```

__STATIC_FORCEINLINE unsigned long __RV_CLRS32(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_CLO32(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_CLZ32(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_PBSAD(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_PBSADA(unsigned long t, unsigned long a, unsigned long b)
__RV_SCLIP32(a, b)
__RV_UCLIP32(a, b)

```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_PART\_SIMD\_MISC**

Partial-SIMD Miscellaneous Instructions.

there are 7 Partial-SIMD Miscellaneous Instructions

## Defines

**\_\_RV\_SCLIP32** (a, b)  
SCLIP32 (SIMD 32-bit Signed Clip Value)

**Type:** DSP

**Syntax:**

```
SCLIP32 Rd, Rs1, imm5u[4:0]
```

**Purpose:** Limit the 32-bit signed integer elements of a register into a signed range simultaneously.

**Description:** This instruction limits the 32-bit signed integer elements stored in Rs1 into a signed integer range between  $2^{\text{imm5u}}-1$  and  $-2^{\text{imm5u}}$ , and writes the limited results to Rd. For example, if imm5u is 3, the 32-bit input values should be saturated between 7 and -8. If saturation is performed, set OV bit to 1.

**Operations:**

```

src = Rs1.W[x];
if (src > (2^imm5u)-1) {
    src = (2^imm5u)-1;
    OV = 1;
} else if (src < -2^imm5u) {
    src = -2^imm5u;
    OV = 1;
}
Rd.W[x] = src
for RV32: x=0,
for RV64: x=1...0

```

**Return** value stored in long type

**Parameters**

- [in] a: long type of value stored in a

- [in] b: unsigned int type of value stored in b

**\_\_RV\_UCLIP32** (a, b)

UCLIP32 (SIMD 32-bit Unsigned Clip Value)

**Type:** SIMD

**Syntax:**

```
UCLIP32 Rd, Rs1, imm5u[4:0]
```

**Purpose:** Limit the 32-bit signed integer elements of a register into an unsigned range simultaneously.

**Description:** This instruction limits the 32-bit signed integer elements stored in Rs1 into an unsigned integer range between  $2^{\text{imm5u}}-1$  and 0, and writes the limited results to Rd. For example, if imm5u is 3, the 32-bit input values should be saturated between 7 and 0. If saturation is performed, set OV bit to 1.

**Operations:**

```
src = Rs1.W[x];
if (src > (2^imm5u)-1) {
    src = (2^imm5u)-1;
    OV = 1;
} else if (src < 0) {
    src = 0;
    OV = 1;
}
Rd.W[x] = src
for RV32: x=0,
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLRS32(unsigned long a)**

CLRS32 (SIMD 32-bit Count Leading Redundant Sign)

**Type:** SIMD

**Syntax:**

```
CLRS32 Rd, Rs1
```

**Purpose:** Count the number of redundant sign bits of the 32-bit elements of a general register.

**Description:** Starting from the bits next to the sign bits of the 32-bit elements of Rs1, this instruction counts the number of redundant sign bits and writes the result to the corresponding 32-bit elements of Rd.

**Operations:**

```
snum[x] = Rs1.W[x];
cnt[x] = 0;
for (i = 30 to 0) {
    if (snum[x](i) == snum[x](31)) {
        cnt[x] = cnt[x] + 1;
    } else {
        break;
    }
}
Rd.W[x] = cnt[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLO32(unsigned long a)**  
CLO32 (SIMD 32-bit Count Leading One)

**Type:** SIMD

**Syntax:**

```
CLO32 Rd, Rs1
```

**Purpose:** Count the number of leading one bits of the 32-bit elements of a general register.

**Description:** Starting from the most significant bits of the 32-bit elements of Rs1, this instruction counts the number of leading one bits and writes the results to the corresponding 32-bit elements of Rd.

**Operations:**

```
snum[x] = Rs1.W[x];
cnt[x] = 0;
for (i = 31 to 0) {
    if (snum[x](i) == 1) {
        cnt[x] = cnt[x] + 1;
    } else {
        break;
    }
}
Rd.W[x] = cnt[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CLZ32(unsigned long a)**  
CLZ32 (SIMD 32-bit Count Leading Zero)

**Type:** SIMD

**Syntax:**

```
CLZ32 Rd, Rs1
```

**Purpose:** Count the number of leading zero bits of the 32-bit elements of a general register.

**Description:** Starting from the most significant bits of the 32-bit elements of Rs1, this instruction counts the number of leading zero bits and writes the results to the corresponding 32-bit elements of Rd.

**Operations:**

```
snum[x] = Rs1.W[x];
cnt[x] = 0;
for (i = 31 to 0) {
    if (snum[x](i) == 0) {
        cnt[x] = cnt[x] + 1;
    } else {
        break;
    }
}
Rd.W[x] = cnt[x];
for RV32: x=0
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PBSAD(unsigned long a, unsigned long b)**  
PBSAD (Parallel Byte Sum of Absolute Difference)

**Type:** DSP

**Syntax:**

```
PBSAD Rd, Rs1, Rs2
```

**Purpose:** Calculate the sum of absolute difference of unsigned 8-bit data elements.

**Description:** This instruction subtracts the un-signed 8-bit elements of Rs2 from those of Rs1. Then it adds the absolute value of each difference together and writes the result to Rd.

**Operations:**

```
absdiff[x] = ABS(Rs1.B[x] - Rs2.B[x]);
Rd = SUM(absdiff[x]);
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PBSADA(unsigned long t, unsigned long a, unsigned long b)**  
PBSADA (Parallel Byte Sum of Absolute Difference Accum)

**Type:** DSP

**Syntax:**

```
PBSADA Rd, Rs1, Rs2
```

**Purpose:** Calculate the sum of absolute difference of four unsigned 8-bit data elements and accumulate it into a register.

**Description:** This instruction subtracts the un-signed 8-bit elements of Rs2 from those of Rs1. It then adds the absolute value of each difference together along with the content of Rd and writes the accumulated result back to Rd.

**Operations:**

```
absdiff[x] = ABS(Rs1.B[x] - Rs2.B[x]);
Rd = Rd + SUM(absdiff[x]);
for RV32: x=3...0,
for RV64: x=7...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] t: unsigned long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**8-bit Multiply with 32-bit Add Instructions**

```
__STATIC_FORCEINLINE long __RV_SMAQA(long t, unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE long __RV_SMAQA_SU(long t, unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_UMAQA(unsigned long t, unsigned long a, unsigned long b)
```

group NMSIS\_Core\_DSP\_Intrinsic\_8B\_MULT\_32B\_ADD

8-bit Multiply with 32-bit Add Instructions

there are 3 8-bit Multiply with 32-bit Add Instructions

**Functions**

```
__STATIC_FORCEINLINE long __RV_SMAQA(long t, unsigned long a, unsigned long b)
SMAQA (Signed Multiply Four Bytes with 32-bit Adds)
```

**Type:** Partial-SIMD (Reduction)

**Syntax:**

```
SMAQA Rd, Rs1, Rs2
```

**Purpose:** Do four signed 8-bit multiplications from 32-bit chunks of two registers; and then adds the four 16-bit results and the content of corresponding 32-bit chunks of a third register together.

**Description:** This instruction multiplies the four signed 8-bit elements of 32-bit chunks of Rs1 with the four signed 8-bit elements of 32-bit chunks of Rs2 and then adds the four results together with the signed content of the corresponding 32-bit chunks of Rd. The final results are written back to the corresponding 32-bit chunks in Rd.



**Operations:**

```

res[x] = Rd.W[x] +
    (Rs1.W[x].B[3] s* Rs2.W[x].B[3]) + (Rs1.W[x].B[2] s* Rs2.W[x].B[2]) +
    (Rs1.W[x].B[1] s* Rs2.W[x].B[1]) + (Rs1.W[x].B[0] s* Rs2.W[x].B[0]);
Rd.W[x] = res[x];
for RV32: x=0,
for RV64: x=1,0

```

**Return** value stored in long type**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMAQA\_SU(long t, unsigned long a, unsigned long b)**  
 SMAQA.SU (Signed and Unsigned Multiply Four Bytes with 32-bit Adds)

**Type:** Partial-SIMD (Reduction)**Syntax:**

```
SMAQA.SU Rd, Rs1, Rs2
```

**Purpose:** Do four signed  $\times$  unsigned 8-bit multiplications from 32-bit chunks of two registers; and then adds the four 16-bit results and the content of corresponding 32-bit chunks of a third register together.

**Description:** This instruction multiplies the four signed 8-bit elements of 32-bit chunks of Rs1 with the four unsigned 8-bit elements of 32-bit chunks of Rs2 and then adds the four results together with the signed content of the corresponding 32-bit chunks of Rd. The final results are written back to the corresponding 32-bit chunks in Rd.

**Operations:**

```

res[x] = Rd.W[x] +
    (Rs1.W[x].B[3] su* Rs2.W[x].B[3]) + (Rs1.W[x].B[2] su* Rs2.W[x].B[2]) +
    (Rs1.W[x].B[1] su* Rs2.W[x].B[1]) + (Rs1.W[x].B[0] su* Rs2.W[x].B[0]);
Rd.W[x] = res[x];
for RV32: x=0,
for RV64: x=1...0

```

**Return** value stored in long type**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UMAQA(unsigned long t, unsigned long a, unsigned long b)**  
 UMAQA (Unsigned Multiply Four Bytes with 32-bit Adds)

**Type:** DSP**Syntax:**

UMAQA Rd, Rs1, Rs2
--------------------

**Purpose:** Do four unsigned 8-bit multiplications from 32-bit chunks of two registers; and then adds the four 16-bit results and the content of corresponding 32-bit chunks of a third register together.

**Description:** This instruction multiplies the four unsigned 8-bit elements of 32-bit chunks of Rs1 with the four unsigned 8-bit elements of 32-bit chunks of Rs2 and then adds the four results together with the unsigned content of the corresponding 32-bit chunks of Rd. The final results are written back to the corresponding 32-bit chunks in Rd.

**Operations:**

<pre> res[x] = Rd.W[x] + (Rs1.W[x].B[3] u* Rs2.W[x].B[3]) +                   (Rs1.W[x].B[2] u* Rs2.W[x].B[2]) + (Rs1.W[x].B[1] u* Rs2.W[x].B[1]) +                   (Rs1.W[x].B[0] u* Rs2.W[x].B[0]); Rd.W[x] = res[x]; for RV32: x=0, for RV64: x=1...0 </pre>
---

**Return** value stored in unsigned long type

**Parameters**

- [in] t: unsigned long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

*group* **NMSIS\_Core\_DSP\_Intrinsic\_PART\_SIMD\_DATA\_PROCESS**  
Partial-SIMD Data Processing Instructions.

## 64-bit Profile Instructions

### 64-bit Addition & Subtraction Instructions

```

__STATIC_FORCEINLINE unsigned long long __RV_ADD64(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE long long __RV_KADD64(long long a, long long b)
__STATIC_FORCEINLINE long long __RV_KSUB64(long long a, long long b)
__STATIC_FORCEINLINE long long __RV_RADD64(long long a, long long b)
__STATIC_FORCEINLINE long long __RV_RSUB64(long long a, long long b)
__STATIC_FORCEINLINE unsigned long long __RV_SUB64(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_UKADD64(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_UKSUB64(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_URADD64(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_URSUB64(unsigned long long a, unsigned long long b)

```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_64B\_ADDSUB**  
64-bit Addition & Subtraction Instructions

there are 10 64-bit Addition & Subtraction Instructions.

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_ADD64(unsigned long long a, unsigned long long b)**  
ADD64 (64-bit Addition)

**Type:** 64-bit Profile

**Syntax:**

```
ADD64 Rd, Rs1, Rs2
```

**Purpose:** Add two 64-bit signed or unsigned integers.

**RV32 Description:** This instruction adds the 64-bit integer of an even/odd pair of registers specified by Rs1(4,1) with the 64-bit integer of an even/odd pair of registers specified by Rs2(4,1), and then writes the 64-bit result to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., value d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction has the same behavior as the ADD instruction in RV64I.

**Note:** This instruction can be used for either signed or unsigned addition.

**Operations:**

```
RV32:
t_L = CONCAT(Rd(4,1), 1'b0); t_H = CONCAT(Rd(4,1), 1'b1);
a_L = CONCAT(Rs1(4,1), 1'b0); a_H = CONCAT(Rs1(4,1), 1'b1);
b_L = CONCAT(Rs2(4,1), 1'b0); b_H = CONCAT(Rs2(4,1), 1'b1);
R[t_H].R[t_L] = R[a_H].R[a_L] + R[b_H].R[b_L];
RV64:
Rd = Rs1 + Rs2;
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_KADD64(long long a, long long b)**  
KADD64 (64-bit Signed Saturating Addition)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
KADD64 Rd, Rs1, Rs2
```

**Purpose:** Add two 64-bit signed integers. The result is saturated to the Q63 range.

**RV32 Description:** This instruction adds the 64-bit signed integer of an even/odd pair of registers specified by Rs1(4,1) with the 64-bit signed integer of an even/odd pair of registers specified by Rs2(4,1). If the 64-bit result is beyond the Q63 number range ( $-2^{63} \leq \text{Q63} \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., value d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction adds the 64-bit signed integer in Rs1 with the 64-bit signed integer in Rs2. If the result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written to Rd.

**Operations:**

```
RV32:
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
a_L = CONCAT(Rs1(4,1),1'b0); a_H = CONCAT(Rs1(4,1),1'b1);
b_L = CONCAT(Rs2(4,1),1'b0); b_H = CONCAT(Rs2(4,1),1'b1);
result = R[a_H].R[a_L] + R[b_H].R[b_L];
if (result > (2^63)-1) {
    result = (2^63)-1; OV = 1;
} else if (result < -2^63) {
    result = -2^63; OV = 1;
}
R[t_H].R[t_L] = result;
RV64:
result = Rs1 + Rs2;
if (result > (2^63)-1) {
    result = (2^63)-1; OV = 1;
} else if (result < -2^63) {
    result = -2^63; OV = 1;
}
Rd = result;
```

**Return** value stored in long long type

**Parameters**

- [in] a: long long type of value stored in a
- [in] b: long long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_KSUB64(long long a, long long b)**  
 KSUB64 (64-bit Signed Saturating Subtraction)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
KSUB64 Rd, Rs1, Rs2
```

**Purpose:** Perform a 64-bit signed integer subtraction. The result is saturated to the Q63 range.

**RV32 Description:** This instruction subtracts the 64-bit signed integer of an even/odd pair of registers specified by Rs2(4,1) from the 64-bit signed integer of an even/odd pair of registers specified by Rs1(4,1). If the 64-bit result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is then written to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the operand and the even 2d register of the pair contains the low 32-bit of the operand.

**RV64 Description:** This instruction subtracts the 64-bit signed integer of Rs2 from the 64-bit signed integer of Rs1. If the 64-bit result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is then written to Rd.

**Operations:**

```

RV32:
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
a_L = CONCAT(Rs1(4,1),1'b0); a_H = CONCAT(Rs1(4,1),1'b1);
b_L = CONCAT(Rs2(4,1),1'b0); b_H = CONCAT(Rs2(4,1),1'b1);
result = R[a_H].R[a_L] - R[b_H].R[b_L];
if (result > (2^63)-1) {
    result = (2^63)-1; OV = 1;
} else if (result < -2^63) {
    result = -2^63; OV = 1;
}
R[t_H].R[t_L] = result;
RV64:
result = Rs1 - Rs2;
if (result > (2^63)-1) {
    result = (2^63)-1; OV = 1;
} else if (result < -2^63) {
    result = -2^63; OV = 1;
}
Rd = result;

```

**Return** value stored in long long type

**Parameters**

- [in] a: long long type of value stored in a
- [in] b: long long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_RADD64(long long a, long long b)**  
 RADD64 (64-bit Signed Halving Addition)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
RADD64 Rd, Rs1, Rs2
```

**Purpose:** Add two 64-bit signed integers. The result is halved to avoid overflow or saturation.

**RV32 Description:** This instruction adds the 64-bit signed integer of an even/odd pair of registers specified by Rs1(4,1) with the 64-bit signed integer of an even/odd pair of registers specified by Rs2(4,1). The 64-bit addition result is first arithmetically right-shifted by 1 bit and then written to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., value d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction adds the 64-bit signed integer in Rs1 with the 64-bit signed integer in Rs2. The 64-bit addition result is first arithmetically right-shifted by 1 bit and then written to Rd.

**Operations:**

```

RV32:
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
a_L = CONCAT(Rs1(4,1),1'b0); a_H = CONCAT(Rs1(4,1),1'b1);
b_L = CONCAT(Rs2(4,1),1'b0); b_H = CONCAT(Rs2(4,1),1'b1);
R[t_H].R[t_L] = (R[a_H].R[a_L] + R[b_H].R[b_L]) s>> 1;
RV64:
Rd = (Rs1 + Rs2) s>> 1;

```

**Return** value stored in long long type

**Parameters**

- [in] a: long long type of value stored in a
- [in] b: long long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_RSUB64(long long a, long long b)**  
 RSUB64 (64-bit Signed Halving Subtraction)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
RSUB64 Rd, Rs1, Rs2
```

**Purpose:** Perform a 64-bit signed integer subtraction. The result is halved to avoid overflow or saturation.

**RV32 Description:** This instruction subtracts the 64-bit signed integer of an even/odd pair of registers specified by Rb(4,1) from the 64-bit signed integer of an even/odd pair of registers specified by Ra(4,1). The subtraction result is first arithmetically right-shifted by 1 bit and then written to an even/odd pair of registers specified by Rt(4,1). Rx(4,1), i.e., value d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction subtracts the 64-bit signed integer in Rs2 from the 64-bit signed integer in Rs1. The 64-bit subtraction result is first arithmetically right-shifted by 1 bit and then written to Rd.

**Operations:**

```
RV32:
t_L = CONCAT(Rd(4,1), 1'b0); t_H = CONCAT(Rd(4,1), 1'b1);
a_L = CONCAT(Rs1(4,1), 1'b0); a_H = CONCAT(Rs1(4,1), 1'b1);
b_L = CONCAT(Rs2(4,1), 1'b0); b_H = CONCAT(Rs2(4,1), 1'b1);
R[t_H].R[t_L] = (R[a_H].R[a_L] - R[b_H].R[b_L]) s>> 1;
RV64:
Rd = (Rs1 - Rs2) s>> 1;
```

**Return** value stored in long long type

**Parameters**

- [in] a: long long type of value stored in a
- [in] b: long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_SUB64(unsigned long long a, unsigned long long b)**  
 SUB64 (64-bit Subtraction)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
SUB64 Rd, Rs1, Rs2
```

**Purpose:** Perform a 64-bit signed or unsigned integer subtraction.

**RV32 Description:** This instruction subtracts the 64-bit integer of an even/odd pair of registers specified by Rs2(4,1) from the 64-bit integer of an even/odd pair of registers specified by Rs1(4,1), and then writes the 64-bit result to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The

odd  $2d+1$  register of the pair contains the high 32-bit of the operand and the even  $2d$  register of the pair contains the low 32-bit of the operand.

**RV64 Description:** This instruction subtracts the 64-bit integer of Rs2 from the 64-bit integer of Rs1, and then writes the 64-bit result to Rd.

**Note:** This instruction can be used for either signed or unsigned subtraction.

#### Operations:

```
* RV32:
t_L = CONCAT(Rd(4,1), 1'b0); t_H = CONCAT(Rd(4,1), 1'b1);
a_L = CONCAT(Rs1(4,1), 1'b0); a_H = CONCAT(Rs1(4,1), 1'b1);
b_L = CONCAT(Rs2(4,1), 1'b0); b_H = CONCAT(Rs2(4,1), 1'b1);
R[t_H].R[t_L] = R[a_H].R[a_L] - R[b_H].R[b_L];
* RV64:
Rd = Rs1 - Rs2;
```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UKADD64(unsigned long long a, unsigned long long b)**  
UKADD64 (64-bit Unsigned Saturating Addition)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
UKADD64 Rd, Rs1, Rs2
```

**Purpose:** Add two 64-bit unsigned integers. The result is saturated to the U64 range.

**RV32 Description:** This instruction adds the 64-bit unsigned integer of an even/odd pair of registers specified by Rs1(4,1) with the 64-bit unsigned integer of an even/odd pair of registers specified by Rs2(4,1). If the 64-bit result is beyond the U64 number range ( $0 \leq U64 \leq 2^{64}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register  $2d$  and  $2d+1$ . The odd  $2d+1$  register of the pair contains the high 32-bit of the result and the even  $2d$  register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction adds the 64-bit unsigned integer in Rs1 with the 64-bit unsigned integer in Rs2. If the 64-bit result is beyond the U64 number range ( $0 \leq U64 \leq 2^{64}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written to Rd.

#### Operations:

```
* RV32:
t_L = CONCAT(Rt(4,1), 1'b0); t_H = CONCAT(Rt(4,1), 1'b1);
a_L = CONCAT(Ra(4,1), 1'b0); a_H = CONCAT(Ra(4,1), 1'b1);
b_L = CONCAT(Rb(4,1), 1'b0); b_H = CONCAT(Rb(4,1), 1'b1);
result = R[a_H].R[a_L] + R[b_H].R[b_L];
if (result > (2^64)-1) {
    result = (2^64)-1; OV = 1;
}
R[t_H].R[t_L] = result;
* RV64:
```

(continues on next page)

(continued from previous page)

```

result = Rs1 + Rs2;
if (result > (2^64)-1) {
    result = (2^64)-1; OV = 1;
}
Rd = result;

```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UKSUB64(unsigned long long a, unsigned long long b)**  
 UKSUB64 (64-bit Unsigned Saturating Subtraction)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
UKSUB64 Rd, Rs1, Rs2
```

**Purpose:** Perform a 64-bit signed integer subtraction. The result is saturated to the U64 range.

**RV32 Description:** This instruction subtracts the 64-bit unsigned integer of an even/odd pair of registers specified by Rs2(4,1) from the 64-bit unsigned integer of an even/odd pair of registers specified by Rs1(4,1). If the 64-bit result is beyond the U64 number range ( $0 \leq U64 \leq 2^{64}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is then written to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the operand and the even 2d register of the pair contains the low 32-bit of the operand.

**RV64 Description:** This instruction subtracts the 64-bit unsigned integer of Rs2 from the 64-bit unsigned integer of an even/odd pair of Rs1. If the 64-bit result is beyond the U64 number range ( $0 \leq U64 \leq 2^{64}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is then written to Rd.

#### Operations:

```

* RV32:
t_L = CONCAT(Rd(4,1), 1'b0); t_H = CONCAT(Rd(4,1), 1'b1);
a_L = CONCAT(Rs1(4,1), 1'b0); a_H = CONCAT(Rs1(4,1), 1'b1);
b_L = CONCAT(Rs2(4,1), 1'b0); b_H = CONCAT(Rs2(4,1), 1'b1);
result = R[a_H].R[a_L] - R[b_H].R[b_L];
if (result < 0) {
    result = 0; OV = 1;
}
R[t_H].R[t_L] = result;
* RV64
result = Rs1 - Rs2;
if (result < 0) {
    result = 0; OV = 1;
}
Rd = result;

```

**Return** value stored in unsigned long long type

#### Parameters



- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_URADD64(unsigned long long a, unsigned long long b)**  
URADD64 (64-bit Unsigned Halving Addition)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
URADD64 Rd, Rs1, Rs2
```

**Purpose:** Add two 64-bit unsigned integers. The result is halved to avoid overflow or saturation.

**RV32 Description:** This instruction adds the 64-bit unsigned integer of an even/odd pair of registers specified by Rs1(4,1) with the 64-bit unsigned integer of an even/odd pair of registers specified by Rs2(4,1). The 64-bit addition result is first logically right-shifted by 1 bit and then written to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction adds the 64-bit unsigned integer in Rs1 with the 64-bit unsigned integer Rs2. The 64-bit addition result is first logically right-shifted by 1 bit and then written to Rd.

**Operations:**

```
* RV32:
t_L = CONCAT(Rt(4,1), 1'b0); t_H = CONCAT(Rt(4,1), 1'b1);
a_L = CONCAT(Ra(4,1), 1'b0); a_H = CONCAT(Ra(4,1), 1'b1);
b_L = CONCAT(Rb(4,1), 1'b0); b_H = CONCAT(Rb(4,1), 1'b1);
R[t_H].R[t_L] = (R[a_H].R[a_L] + R[b_H].R[b_L]) u>> 1;
* RV64:
Rd = (Rs1 + Rs2) u>> 1;
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_URSUB64(unsigned long long a, unsigned long long b)**  
URSUB64 (64-bit Unsigned Halving Subtraction)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
URSUB64 Rd, Rs1, Rs2
```

**Purpose:** Perform a 64-bit unsigned integer subtraction. The result is halved to avoid overflow or saturation.

**RV32 Description:** This instruction subtracts the 64-bit unsigned integer of an even/odd pair of registers specified by Rs2(4,1) from the 64-bit unsigned integer of an even/odd pair of registers specified by Rs1(4,1). The subtraction result is first logically right-shifted by 1 bit and then written to an even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction subtracts the 64-bit unsigned integer in Rs2 from the 64-bit unsigned integer in Rs1. The subtraction result is first logically right-shifted by 1 bit and then written to Rd.

**Operations:**

```
* RV32:
t_L = CONCAT(Rt(4,1),1'b0); t_H = CONCAT(Rt(4,1),1'b1);
a_L = CONCAT(Ra(4,1),1'b0); a_H = CONCAT(Ra(4,1),1'b1);
b_L = CONCAT(Rb(4,1),1'b0); b_H = CONCAT(Rb(4,1),1'b1);
R[t_H].R[t_L] = (R[a_H].R[a_L] - R[b_H].R[b_L]) u>> 1;
* RV64:
Rd = (Rs1 - Rs2) u>> 1;
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

### 32-bit Multiply with 64-bit Add/Subtract Instructions

```
__STATIC_FORCEINLINE long long __RV_KMAR64(long long t, long a, long b)
__STATIC_FORCEINLINE long long __RV_KMSR64(long long t, long a, long b)
__STATIC_FORCEINLINE long long __RV_SMAR64(long long t, long a, long b)
__STATIC_FORCEINLINE long long __RV_SMSR64(long long t, long a, long b)
__STATIC_FORCEINLINE unsigned long long __RV_UKMAR64(unsigned long long t, unsigned long a,
__STATIC_FORCEINLINE unsigned long long __RV_UKMSR64(unsigned long long t, unsigned long a,
__STATIC_FORCEINLINE unsigned long long __RV_UMAR64(unsigned long long t, unsigned long a,
__STATIC_FORCEINLINE unsigned long long __RV_UMSR64(unsigned long long t, unsigned long a,
```

group **NMSIS\_Core\_DSP\_Intrinsic\_32B\_MULT\_64B\_ADDSUB**

32-bit Multiply with 64-bit Add/Subtract Instructions

there are 32-bit Multiply 64-bit Add/Subtract Instructions

### Functions

```
__STATIC_FORCEINLINE long long __RV_KMAR64(long long t, long a, long b)
```

KMAR64 (Signed Multiply and Saturating Add to 64-Bit Data)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
KMAR64 Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit signed elements in two registers and add the 64-bit multiplication results to the 64-bit signed data of a pair of registers (RV32) or a register (RV64). The result is saturated to the Q63 range and written back to the pair of registers (RV32) or the register (RV64).

**RV32 Description:** This instruction multiplies the 32-bit signed data of Rs1 with that of Rs2. It adds the 64-bit multiplication result to the 64-bit signed data of an even/odd pair of registers specified by Rd(4,1)

with unlimited precision. If the 64-bit addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written back to the even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., value d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction multiplies the 32-bit signed elements of Rs1 with that of Rs2. It adds the 64-bit multiplication results to the 64-bit signed data of Rd with unlimited precision. If the 64-bit addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written back to Rd.

#### Operations:

```
RV32:
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
result = R[t_H].R[t_L] + (Rs1 * Rs2);
if (result > (2^63)-1) {
    result = (2^63)-1; OV = 1;
} else if (result < -2^63) {
    result = -2^63; OV = 1;
}
R[t_H].R[t_L] = result;
RV64:
// `result` has unlimited precision
result = Rd + (Rs1.W[0] * Rs2.W[0]) + (Rs1.W[1] * Rs2.W[1]);
if (result > (2^63)-1) {
    result = (2^63)-1; OV = 1;
} else if (result < -2^63) {
    result = -2^63; OV = 1;
}
Rd = result;
```

**Return** value stored in long long type

#### Parameters

- [in] t: long long type of value stored in t
- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_KMSR64(long long t, long a, long b)**  
KMSR64 (Signed Multiply and Saturating Subtract from 64-Bit Data)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
KMSR64 Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit signed elements in two registers and subtract the 64-bit multiplication results from the 64-bit signed data of a pair of registers (RV32) or a register (RV64). The result is saturated to the Q63 range and written back to the pair of registers (RV32) or the register (RV64).

**RV32 Description:** This instruction multiplies the 32-bit signed data of Rs1 with that of Rs2. It subtracts the 64-bit multiplication result from the 64-bit signed data of an even/odd pair of registers specified by Rd(4,1) with unlimited precision. If the 64-bit subtraction result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written

back to the even/odd pair of registers specified by  $Rd(4,1)$ .  $Rx(4,1)$ , i.e.,  $d$ , determines the even/odd pair group of two registers. Specifically, the register pair includes register  $2d$  and  $2d+1$ . The odd  $2d+1$  register of the pair contains the high 32-bit of the result and the even  $2d$  register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction multiplies the 32-bit signed elements of  $Rs1$  with that of  $Rs2$ . It subtracts the 64-bit multiplication results from the 64-bit signed data in  $Rd$  with unlimited precision. If the 64-bit subtraction result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written back to  $Rd$ .

#### Operations:

```
RV32:
t_L = CONCAT(Rd(4,1), 1'b0); t_H = CONCAT(Rd(4,1), 1'b1);
result = R[t_H].R[t_L] - (Rs1 * Rs2);
if (result > (2^63)-1) {
    result = (2^63)-1; OV = 1;
} else if (result < -2^63) {
    result = -2^63; OV = 1;
}
R[t_H].R[t_L] = result;
RV64:
// `result` has unlimited precision
result = Rd - (Rs1.W[0] * Rs2.W[0]) - (Rs1.W[1] * Rs2.W[1]);
if (result > (2^63)-1) {
    result = (2^63)-1; OV = 1;
} else if (result < -2^63) {
    result = -2^63; OV = 1;
}
Rd = result;
```

**Return** value stored in long long type

#### Parameters

- [in]  $t$ : long long type of value stored in  $t$
- [in]  $a$ : long type of value stored in  $a$
- [in]  $b$ : long type of value stored in  $b$

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMAR64(long long t, long a, long b)**  
 SMAR64 (Signed Multiply and Add to 64-Bit Data)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
SMAR64 Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit signed elements in two registers and add the 64-bit multiplication result to the 64-bit signed data of a pair of registers (RV32) or a register (RV64). The result is written back to the pair of registers (RV32) or a register (RV64).

**RV32 Description:** This instruction multiplies the 32-bit signed data of  $Rs1$  with that of  $Rs2$ . It adds the 64-bit multiplication result to the 64-bit signed data of an even/odd pair of registers specified by  $Rd(4,1)$ . The addition result is written back to the even/odd pair of registers specified by  $Rd(4,1)$ .  $Rx(4,1)$ , i.e.,  $d$ , determines the even/odd pair group of two registers. Specifically, the register pair includes register  $2d$  and  $2d+1$ . The odd  $2d+1$  register of the pair contains the high 32-bit of the result and the even  $2d$  register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction multiplies the 32-bit signed elements of Rs1 with that of Rs2. It adds the 64-bit multiplication results to the 64-bit signed data of Rd. The addition result is written back to Rd.

**Operations:**

```
* RV32:
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H].R[t_L] = R[t_H].R[t_L] + (Rs1 * Rs2);
* RV64:
Rd = Rd + (Rs1.W[0] * Rs2.W[0]) + (Rs1.W[1] * Rs2.W[1]);
```

**Return** value stored in long long type

**Parameters**

- [in] t: long long type of value stored in t
- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMSR64(long long t, long a, long b)**  
SMSR64 (Signed Multiply and Subtract from 64- Bit Data)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
SMSR64 Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit signed elements in two registers and subtract the 64-bit multiplication results from the 64-bit signed data of a pair of registers (RV32) or a register (RV64). The result is written back to the pair of registers (RV32) or a register (RV64).

**RV32 Description:** This instruction multiplies the 32-bit signed data of Rs1 with that of Rs2. It subtracts the 64-bit multiplication result from the 64-bit signed data of an even/odd pair of registers specified by Rd(4,1). The subtraction result is written back to the even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction multiplies the 32-bit signed elements of Rs1 with that of Rs2. It subtracts the 64-bit multiplication results from the 64-bit signed data of Rd. The subtraction result is written back to Rd.

**Operations:**

```
* RV32:
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
R[t_H].R[t_L] = R[t_H].R[t_L] - (Rs1 * Rs2);
* RV64:
Rd = Rd - (Rs1.W[0] * Rs2.W[0]) - (Rs1.W[1] * Rs2.W[1]);
```

**Return** value stored in long long type

**Parameters**

- [in] t: long long type of value stored in t
- [in] a: long type of value stored in a
- [in] b: long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UKMAR64(unsigned long long t, unsigned long long a, unsigned long long b)**  
 UKMAR64 (Unsigned Multiply and Saturating Add to 64-Bit Data)

**Type:** DSP (64-bit Profile)

**Syntax:**

UKMAR64 Rd, Rs1, Rs2
----------------------

**Purpose:** Multiply the 32-bit unsigned elements in two registers and add the 64-bit multiplication results to the 64-bit unsigned data of a pair of registers (RV32) or a register (RV64). The result is saturated to the U64 range and written back to the pair of registers (RV32) or the register (RV64).

**RV32 Description:** This instruction multiplies the 32-bit unsigned data of Rs1 with that of Rs2. It adds the 64-bit multiplication result to the 64-bit unsigned data of an even/odd pair of registers specified by Rd(4,1) with unlimited precision. If the 64-bit addition result is beyond the U64 number range ( $0 \leq U64 \leq 2^{64}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written back to the even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction multiplies the 32-bit unsigned elements of Rs1 with that of Rs2. It adds the 64-bit multiplication results to the 64-bit unsigned data in Rd with unlimited precision. If the 64-bit addition result is beyond the U64 number range ( $0 \leq U64 \leq 2^{64}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written back to Rd.

**Operations:**

```
* RV32:
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
result = R[t_H].R[t_L] + (Rs1 * Rs2);
if (result > (2^64)-1) {
    result = (2^64)-1; OV = 1;
}
R[t_H].R[t_L] = result;
* RV64:
// `result` has unlimited precision
result = Rd + (Rs1.W[0] u* Rs2.W[0]) + (Rs1.W[1] u* Rs2.W[1]);
if (result > (2^64)-1) {
    result = (2^64)-1; OV = 1;
}
Rd = result;
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] t: unsigned long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UKMSR64(unsigned long long t, unsigned long long a, unsigned long long b)**  
 UKMSR64 (Unsigned Multiply and Saturating Subtract from 64-Bit Data)

**Type:** DSP (64-bit Profile)

**Syntax:**

UKMSR64 Rd, Rs1, Rs2
----------------------

**Purpose:** Multiply the 32-bit unsigned elements in two registers and subtract the 64-bit multiplication results from the 64-bit unsigned data of a pair of registers (RV32) or a register (RV64). The result is saturated to the U64 range and written back to the pair of registers (RV32) or a register (RV64).

**RV32 Description:** This instruction multiplies the 32-bit unsigned data of Rs1 with that of Rs2. It subtracts the 64-bit multiplication result from the 64-bit unsigned data of an even/odd pair of registers specified by Rd(4,1) with unlimited precision. If the 64-bit subtraction result is beyond the U64 number range ( $0 \leq \text{U64} \leq 2^{64}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written back to the even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction multiplies the 32-bit unsigned elements of Rs1 with that of Rs2. It subtracts the 64-bit multiplication results from the 64-bit unsigned data of Rd with unlimited precision. If the 64-bit subtraction result is beyond the U64 number range ( $0 \leq \text{U64} \leq 2^{64}-1$ ), it is saturated to the range and the OV bit is set to 1. The saturated result is written back to Rd.

#### Operations:

```
* RV32:
t_L = CONCAT(Rd(4,1),1'b0); t_H = CONCAT(Rd(4,1),1'b1);
result = R[t_H].R[t_L] - (Rs1 u* Rs2);
if (result < 0) {
    result = 0; OV = 1;
}
R[t_H].R[t_L] = result;
* RV64:
// `result` has unlimited precision
result = Rd - (Rs1.W[0] u* Rs2.W[0]) - (Rs1.W[1] u* Rs2.W[1]);
if (result < 0) {
    result = 0; OV = 1;
}
Rd = result;
```

**Return** value stored in unsigned long long type

#### Parameters

- [in] t: unsigned long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UMAR64(unsigned long long t, unsigned long long a, unsigned long long b)**  
UMAR64 (Unsigned Multiply and Add to 64-Bit Data)

**Type:** DSP (64-bit Profile)

#### Syntax:

UMAR64 Rd, Rs1, Rs2
---------------------

**Purpose:** Multiply the 32-bit unsigned elements in two registers and add the 64-bit multiplication results to the 64-bit unsigned data of a pair of registers (RV32) or a register (RV64). The result is written back to the pair of registers (RV32) or a register (RV64).

**RV32 Description:** This instruction multiplies the 32-bit unsigned data of Rs1 with that of Rs2. It adds the 64-bit multiplication result to the 64-bit unsigned data of an even/odd pair of registers specified by Rd(4,1). The addition result is written back to the even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction multiplies the 32-bit unsigned elements of Rs1 with that of Rs2. It adds the 64-bit multiplication results to the 64-bit unsigned data of Rd. The addition result is written back to Rd.

**Operations:**

```
* RV32:
t_L = CONCAT(Rd(4,1), 1'b0); t_H = CONCAT(Rd(4,1), 1'b1);
R[t_H].R[t_L] = R[t_H].R[t_L] + (Rs1 * Rs2);
* RV64:
Rd = Rd + (Rs1.W[0] u* Rs2.W[0]) + (Rs1.W[1] u* Rs2.W[1]);
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] t: unsigned long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_UMSR64(unsigned long long t, unsigned long long a, unsigned long long b)**  
UMSR64 (Unsigned Multiply and Subtract from 64-Bit Data)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
UMSR64 Rd, Rs1, Rs2
```

**Purpose:** Multiply the 32-bit unsigned elements in two registers and subtract the 64-bit multiplication results from the 64-bit unsigned data of a pair of registers (RV32) or a register (RV64). The result is written back to the pair of registers (RV32) or a register (RV64).

**RV32 Description:** This instruction multiplies the 32-bit unsigned data of Rs1 with that of Rs2. It subtracts the 64-bit multiplication result from the 64-bit unsigned data of an even/odd pair of registers specified by Rd(4,1). The subtraction result is written back to the even/odd pair of registers specified by Rd(4,1). Rx(4,1), i.e., d, determines the even/odd pair group of two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** This instruction multiplies the 32-bit unsigned elements of Rs1 with that of Rs2. It subtracts the 64-bit multiplication results from the 64-bit unsigned data of Rd. The subtraction result is written back to Rd.

**Operations:**

```
* RV32:
t_L = CONCAT(Rd(4,1), 1'b0); t_H = CONCAT(Rd(4,1), 1'b1);
R[t_H].R[t_L] = R[t_H].R[t_L] - (Rs1 * Rs2);
* RV64:
Rd = Rd - (Rs1.W[0] u* Rs2.W[0]) - (Rs1.W[1] u* Rs2.W[1]);
```



**Return** value stored in unsigned long long type

**Parameters**

- [in] t: unsigned long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## Signed 16-bit Multiply 64-bit Add/Subtract Instructions

```
__STATIC_FORCEINLINE long long __RV_SMAL(long long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMALBB(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMALBT(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMALTT(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMALDA(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMALXDA(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMALDS(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMALDRS(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMALXDS(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMSLDA(long long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long long __RV_SMSLXDA(long long t, unsigned long a, unsigned long b)
```

group NMSIS\_Core\_DSP\_Intrinsic\_SIGNED\_16B\_MULT\_64B\_ADDSUB

Signed 16-bit Multiply 64-bit Add/Subtract Instructions.

Signed 16-bit Multiply with 64-bit Add/Subtract Instructions.

there is Signed 16-bit Multiply 64-bit Add/Subtract Instructions

there are 10 Signed 16-bit Multiply with 64-bit Add/Subtract Instructions

## Functions

```
__STATIC_FORCEINLINE long long __RV_SMAL(long long a, unsigned long b)
```

SMAL (Signed Multiply Halfs & Add 64-bit)

**Type:** Partial-SIMD

**Syntax:**

SMAL Rd, Rs1, Rs2
-------------------

**Purpose:** Multiply the signed bottom 16-bit content of the 32-bit elements of a register with the top 16-bit content of the same 32-bit elements of the same register, and add the results with a 64-bit value of an even/odd pair of registers (RV32) or a register (RV64). The addition result is written back to another even/odd pair of registers (RV32) or a register (RV64).

**RV32 Description:** This instruction multiplies the bottom 16-bit content of the lower 32-bit of Rs2 with the top 16-bit content of the lower 32-bit of Rs2 and adds the result with the 64-bit value of an even/odd pair of registers specified by Rs1(4,1). The 64-bit addition result is written back to an even/odd pair of registers specified by Rd(4,1). The 16-bit values of Rs2, and the 64-bit value of the Rs1(4,1) register- pair

are treated as signed integers.  $Rx(4,1)$ , i.e.,  $d$ , determines the even/odd pair group of the two registers. Specifically, the register pair includes register  $2d$  and  $2d+1$ . The odd  $2d+1$  register of the pair contains the high 32-bit of the operand and the even  $2d$  register of the pair contains the low 32-bit of the operand.

**RV64 Description:** This instruction multiplies the bottom 16-bit content of the 32-bit elements of  $Rs2$  with the top 16-bit content of the same 32-bit elements of  $Rs2$  and adds the results with the 64-bit value of  $Rs1$ . The 64-bit addition result is written back to  $Rd$ . The 16-bit values of  $Rs2$ , and the 64-bit value of  $Rs1$  are treated as signed integers.

#### Operations:

```
RV32:
Mres[31:0] = Rs2.H[1] * Rs2.H[0];
Idx0 = CONCAT(Rs1(4,1), 1'b0); Idx1 = CONCAT(Rs1(4,1), 1'b1); +
Idx2 = CONCAT(Rd(4,1), 1'b0); Idx3 = CONCAT(Rd(4,1), 1'b1);
R[Idx3].R[Idx2] = R[Idx1].R[Idx0] + SE64(Mres[31:0]);
RV64:
Mres[0][31:0] = Rs2.W[0].H[1] * Rs2.W[0].H[0];
Mres[1][31:0] = Rs2.W[1].H[1] * Rs2.W[1].H[0];
Rd = Rs1 + SE64(Mres[1][31:0]) + SE64(Mres[0][31:0]);
```

**Return** value stored in long long type

#### Parameters

- [in] a: long long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMALBB(long long t, unsigned long a, unsigned long b)**  
SMALBB (Signed Multiply Bottom Halfs & Add 64-bit)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
SMALBB Rd, Rs1, Rs2
SMALBT Rd, Rs1, Rs2
SMALTT Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 16-bit content of the 32-bit elements of a register with the 16-bit content of the corresponding 32-bit elements of another register and add the results with a 64-bit value of an even/odd pair of registers (RV32) or a register (RV64). The addition result is written back to the register-pair (RV32) or the register (RV64).

- SMALBB:  $rt\ pair + bottom * bottom$  (all 32-bit elements)
- SMALBT  $rt\ pair + bottom * top$  (all 32-bit elements)
- SMALTT  $rt\ pair + top * top$  (all 32-bit elements)

**RV32 Description:** For the SMALBB instruction, it multiplies the bottom 16-bit content of  $Rs1$  with the bottom 16-bit content of  $Rs2$ . For the SMALBT instruction, it multiplies the bottom 16-bit content of  $Rs1$  with the top 16-bit content of  $Rs2$ . For the SMALTT instruction, it multiplies the top 16-bit content of  $Rs1$  with the top 16-bit content of  $Rs2$ . The multiplication result is added with the 64-bit value of an even/odd pair of registers specified by  $Rd(4,1)$ . The 64-bit addition result is written back to the register-pair. The 16-bit values of  $Rs1$  and  $Rs2$ , and the 64-bit value of the register-pair are treated as signed integers.  $Rd(4,1)$ , i.e.,  $d$ , determines the even/odd pair group of the two registers. Specifically, the register pair includes register  $2d$  and  $2d+1$ . The odd  $2d+1$  register of the pair contains the high 32-bit of the operand and the even  $2d$  register of the pair contains the low 32-bit of the operand.

**RV64 Description:** For the `SMALBB` instruction, it multiplies the bottom 16-bit content of the 32-bit elements of `Rs1` with the bottom 16-bit content of the 32-bit elements of `Rs2`. For the `SMALBT` instruction, it multiplies the bottom 16-bit content of the 32-bit elements of `Rs1` with the top 16-bit content of the 32-bit elements of `Rs2`. For the `SMALTT` instruction, it multiplies the top 16-bit content of the 32-bit elements of `Rs1` with the top 16-bit content of the 32-bit elements of `Rs2`. The multiplication results are added with the 64-bit value of `Rd`. The 64-bit addition result is written back to `Rd`. The 16-bit values of `Rs1` and `Rs2`, and the 64-bit value of `Rd` are treated as signed integers.

#### Operations:

```
RV32:
Mres[31:0] = Rs1.H[0] * Rs2.H[0]; // SMALBB
Mres[31:0] = Rs1.H[0] * Rs2.H[1]; // SMALBT
Mres[31:0] = Rs1.H[1] * Rs2.H[1]; // SMALTT
Idx0 = CONCAT(Rd(4,1),1'b0); Idx1 = CONCAT(Rd(4,1),1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] + SE64(Mres[31:0]);
RV64:
// SMALBB
Mres[0][31:0] = Rs1.W[0].H[0] * Rs2.W[0].H[0];
Mres[1][31:0] = Rs1.W[1].H[0] * Rs2.W[1].H[0];
// SMALBT
Mres[0][31:0] = Rs1.W[0].H[0] * Rs2.W[0].H[1];
Mres[1][31:0] = Rs1.W[1].H[0] * Rs2.W[1].H[1];
// SMALTT
Mres[0][31:0] = Rs1.W[0].H[1] * Rs2.W[0].H[1];
Mres[1][31:0] = Rs1.W[1].H[1] * Rs2.W[1].H[1];
Rd = Rd + SE64(Mres[0][31:0]) + SE64(Mres[1][31:0]);
```

**Return** value stored in long long type

#### Parameters

- [in] `t`: long long type of value stored in `t`
- [in] `a`: unsigned long type of value stored in `a`
- [in] `b`: unsigned long type of value stored in `b`

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMALBT(long long t, unsigned long a, unsigned long b)**  
SMALBT (Signed Multiply Bottom Half & Top Half & Add 64-bit)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
SMALBB Rd, Rs1, Rs2
SMALBT Rd, Rs1, Rs2
SMALTT Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 16-bit content of the 32-bit elements of a register with the 16-bit content of the corresponding 32-bit elements of another register and add the results with a 64-bit value of an even/odd pair of registers (RV32) or a register (RV64). The addition result is written back to the register-pair (RV32) or the register (RV64).

- SMALBB: `rt pair + bottom*bottom` (all 32-bit elements)
- SMALBT: `rt pair + bottom*top` (all 32-bit elements)
- SMALTT: `rt pair + top*top` (all 32-bit elements)

**RV32 Description:** For the `SMALBB` instruction, it multiplies the bottom 16-bit content of `Rs1` with the bottom 16-bit content of `Rs2`. For the `SMALBT` instruction, it multiplies the bottom 16-bit content of `Rs1`

with the top 16-bit content of Rs2. For the SMALTT instruction, it multiplies the top 16-bit content of Rs1 with the top 16-bit content of Rs2. The multiplication result is added with the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit addition result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the operand and the even 2d register of the pair contains the low 32-bit of the operand.

**RV64 Description:** For the SMALBB instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMALBT instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMALTT instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. The multiplication results are added with the 64-bit value of Rd. The 64-bit addition result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

#### Operations:

```
RV32:
Mres[31:0] = Rs1.H[0] * Rs2.H[0]; // SMALBB
Mres[31:0] = Rs1.H[0] * Rs2.H[1]; // SMALBT
Mres[31:0] = Rs1.H[1] * Rs2.H[1]; // SMALTT
Idx0 = CONCAT(Rd(4,1), 1'b0); Idx1 = CONCAT(Rd(4,1), 1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] + SE64(Mres[31:0]);
RV64:
// SMALBB
Mres[0][31:0] = Rs1.W[0].H[0] * Rs2.W[0].H[0];
Mres[1][31:0] = Rs1.W[1].H[0] * Rs2.W[1].H[0];
// SMALBT
Mres[0][31:0] = Rs1.W[0].H[0] * Rs2.W[0].H[1];
Mres[1][31:0] = Rs1.W[1].H[0] * Rs2.W[1].H[1];
// SMALTT
Mres[0][31:0] = Rs1.W[0].H[1] * Rs2.W[0].H[1];
Mres[1][31:0] = Rs1.W[1].H[1] * Rs2.W[1].H[1];
Rd = Rd + SE64(Mres[0][31:0]) + SE64(Mres[1][31:0]);
```

**Return** value stored in long long type

#### Parameters

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMALTT(long long t, unsigned long a, unsigned long b)**  
SMALTT (Signed Multiply Top Halfs & Add 64-bit)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
SMALBB Rd, Rs1, Rs2
SMALBT Rd, Rs1, Rs2
SMALTT Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 16-bit content of the 32-bit elements of a register with the 16-bit content of the corresponding 32-bit elements of another register and add the results with a 64-bit value of an even/odd

pair of registers (RV32) or a register (RV64). The addition result is written back to the register-pair (RV32) or the register (RV64).

- **SMALBB:**  $rt \text{ pair} + \text{bottom} * \text{bottom}$  (all 32-bit elements)
- **SMALBT**  $rt \text{ pair} + \text{bottom} * \text{top}$  (all 32-bit elements)
- **SMALTT**  $rt \text{ pair} + \text{top} * \text{top}$  (all 32-bit elements)

**RV32 Description:** For the **SMALBB** instruction, it multiplies the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2. For the **SMALBT** instruction, it multiplies the bottom 16-bit content of Rs1 with the top 16-bit content of Rs2. For the **SMALTT** instruction, it multiplies the top 16-bit content of Rs1 with the top 16-bit content of Rs2. The multiplication result is added with the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit addition result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register  $2d$  and  $2d+1$ . The odd  $2d+1$  register of the pair contains the high 32-bit of the operand and the even  $2d$  register of the pair contains the low 32-bit of the operand.

**RV64 Description:** For the **SMALBB** instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the **SMALBT** instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the **SMALTT** instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. The multiplication results are added with the 64-bit value of Rd. The 64-bit addition result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

#### Operations:

```
RV32:
Mres[31:0] = Rs1.H[0] * Rs2.H[0]; // SMALBB
Mres[31:0] = Rs1.H[0] * Rs2.H[1]; // SMALBT
Mres[31:0] = Rs1.H[1] * Rs2.H[1]; // SMALTT
Idx0 = CONCAT(Rd(4,1), 1'b0); Idx1 = CONCAT(Rd(4,1), 1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] + SE64(Mres[31:0]);
RV64:
// SMALBB
Mres[0][31:0] = Rs1.W[0].H[0] * Rs2.W[0].H[0];
Mres[1][31:0] = Rs1.W[1].H[0] * Rs2.W[1].H[0];
// SMALBT
Mres[0][31:0] = Rs1.W[0].H[0] * Rs2.W[0].H[1];
Mres[1][31:0] = Rs1.W[1].H[0] * Rs2.W[1].H[1];
// SMALTT
Mres[0][31:0] = Rs1.W[0].H[1] * Rs2.W[0].H[1];
Mres[1][31:0] = Rs1.W[1].H[1] * Rs2.W[1].H[1];
Rd = Rd + SE64(Mres[0][31:0]) + SE64(Mres[1][31:0]);
```

**Return** value stored in long long type

#### Parameters

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMALDA(long long t, unsigned long a, unsigned long b)**  
SMALDA (Signed Multiply Two Halfs and Two Adds 64-bit)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
SMALDA Rd, Rs1, Rs2
SMALXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then adds the two 32-bit results and the 64-bit value of an even/odd pair of registers together.

- SMALDA:  $rt\ pair + top * top + bottom * bottom$  (all 32-bit elements)
- SMALXDA:  $rt\ pair + top * bottom + bottom * top$  (all 32-bit elements)

**RV32 Description:** For the SMALDA instruction, it multiplies the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and then adds the result to the result of multiplying the top 16-bit content of Rs1 with the top 16-bit content of Rs2 with unlimited precision. For the SMALXDA instruction, it multiplies the top 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and then adds the result to the result of multiplying the bottom 16-bit content of Rs1 with the top 16-bit content of Rs2 with unlimited precision. The result is added to the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit addition result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the operand and the even 2d register of the pair contains the low 32-bit of the operand.

**RV64 Description:** For the SMALDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then adds the result to the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 with unlimited precision. For the SMALXDA instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then adds the result to the result of multiplying the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 with unlimited precision. The results are added to the 64-bit value of Rd. The 64-bit addition result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

**Operations:**

```
RV32:
// SMALDA
Mres0[31:0] = (Rs1.H[0] * Rs2.H[0]);
Mres1[31:0] = (Rs1.H[1] * Rs2.H[1]);
// SMALXDA
Mres0[31:0] = (Rs1.H[0] * Rs2.H[1]);
Mres1[31:0] = (Rs1.H[1] * Rs2.H[0]);
Idx0 = CONCAT(Rd(4,1), 1'b0); Idx1 = CONCAT(Rd(4,1), 1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] + SE64(Mres0[31:0]) + SE64(Mres1[31:0]);
RV64:
// SMALDA
Mres0[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[0]);
Mres1[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[1]);
Mres0[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[1].H[0]);
Mres1[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[1].H[1]);
// SMALXDA
Mres0[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[1]);
Mres1[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[0]);
Mres0[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[1].H[1]);
Mres1[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[1].H[0]);
Rd = Rd + SE64(Mres0[0][31:0]) + SE64(Mres1[0][31:0]) + SE64(Mres0[1][31:0]) +
SE64(Mres1[1][31:0]);
```

**Return** value stored in long long type

**Parameters**

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMALXDA(long long t, unsigned long a, unsigned long b)**  
SMALXDA (Signed Crossed Multiply Two Halfs and Two Adds 64-bit)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
SMALDA Rd, Rs1, Rs2
SMALXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then adds the two 32-bit results and the 64-bit value of an even/odd pair of registers together.

- SMALDA: rt pair+ top\*top + bottom\*bottom (all 32-bit elements)
- SMALXDA: rt pair+ top\*bottom + bottom\*top (all 32-bit elements)

**RV32 Description:** For the SMALDA instruction, it multiplies the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and then adds the result to the result of multiplying the top 16-bit content of Rs1 with the top 16-bit content of Rs2 with unlimited precision. For the SMALXDA instruction, it multiplies the top 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and then adds the result to the result of multiplying the bottom 16-bit content of Rs1 with the top 16-bit content of Rs2 with unlimited precision. The result is added to the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit addition result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the operand and the even 2d register of the pair contains the low 32-bit of the operand.

**RV64 Description:** For the SMALDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then adds the result to the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 with unlimited precision. For the SMALXDA instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then adds the result to the result of multiplying the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 with unlimited precision. The results are added to the 64-bit value of Rd. The 64-bit addition result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

**Operations:**

```
RV32:
// SMALDA
Mres0[31:0] = (Rs1.H[0] * Rs2.H[0]);
Mres1[31:0] = (Rs1.H[1] * Rs2.H[1]);
// SMALXDA
Mres0[31:0] = (Rs1.H[0] * Rs2.H[1]);
Mres1[31:0] = (Rs1.H[1] * Rs2.H[0]);
Idx0 = CONCAT(Rd(4,1), 1'b0); Idx1 = CONCAT(Rd(4,1), 1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] + SE64(Mres0[31:0]) + SE64(Mres1[31:0]);
```

(continues on next page)

(continued from previous page)

```

RV64:
// SMALDA
Mres0[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[0]);
Mres1[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[1]);
Mres0[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[1].H[0]);
Mres1[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[1].H[1]);
// SMALXDA
Mres0[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[1]);
Mres1[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[0]);
Mres0[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[1].H[1]);
Mres1[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[1].H[0]);
Rd = Rd + SE64(Mres0[0][31:0]) + SE64(Mres1[0][31:0]) + SE64(Mres0[1][31:0]) +
SE64(Mres1[1][31:0]);

```

**Return** value stored in long long type

#### Parameters

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMALDS(long long t, unsigned long a, unsigned long b)**  
SMALDS (Signed Multiply Two Halfs & Subtract & Add 64-bit)

**Type:** DSP (64-bit Profile)

#### Syntax:

```

SMALDS Rd, Rs1, Rs2
SMALDRS Rd, Rs1, Rs2
SMALXDS Rd, Rs1, Rs2

```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then perform a subtraction operation between the two 32-bit results. Then add the subtraction result to the 64-bit value of an even/odd pair of registers (RV32) or a register (RV64). The addition result is written back to the register-pair.

- SMALDS: rt pair + (top\*top - bottom\*bottom) (all 32-bit elements)
- SMALDRS: rt pair + (bottom\*bottom - top\*top) (all 32-bit elements)
- SMALXDS: rt pair + (top\*bottom - bottom\*top) (all 32-bit elements)

**RV32 Description:** For the SMALDS instruction, it multiplies the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of Rs1 with the top 16-bit content of Rs2. For the SMALDRS instruction, it multiplies the top 16-bit content of Rs1 with the top 16-bit content of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2. For the SMALXDS instruction, it multiplies the bottom 16-bit content of Rs1 with the top 16-bit content of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of Rs1 with the bottom 16-bit content of Rs2. The subtraction result is then added to the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit addition result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the operand and the even 2d register of the pair contains the low 32-bit of the operand.



**RV64 Description:** For the SMALDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMALDRS instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMALXDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The subtraction results are then added to the 64-bit value of Rd. The 64-bit addition result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

#### Operations:

```
* RV32:
Mres[31:0] = (Rs1.H[1] * Rs2.H[1]) - (Rs1.H[0] * Rs2.H[0]); // SMALDS
Mres[31:0] = (Rs1.H[0] * Rs2.H[0]) - (Rs1.H[1] * Rs2.H[1]); // SMALDRS
Mres[31:0] = (Rs1.H[1] * Rs2.H[0]) - (Rs1.H[0] * Rs2.H[1]); // SMALXDS
Idx0 = CONCAT(Rd(4,1),1'b0); Idx1 = CONCAT(Rd(4,1),1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] + SE64(Mres[31:0]);

* RV64:
// SMALDS
Mres[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[1]) - (Rs1.W[0].H[0] * Rs2.W[0].
↪H[0]);
Mres[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[0].H[1]) - (Rs1.W[1].H[0] * Rs2.W[1].
↪H[0]);
// SMALDRS
Mres[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[0]) - (Rs1.W[0].H[1] * Rs2.W[0].
↪H[1]);
Mres[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[0].H[0]) - (Rs1.W[1].H[1] * Rs2.W[1].
↪H[1]);
// SMALXDS
Mres[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[0]) - (Rs1.W[0].H[0] * Rs2.W[0].
↪H[1]);
Mres[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[0].H[0]) - (Rs1.W[1].H[0] * Rs2.W[1].
↪H[1]);
Rd = Rd + SE64(Mres[0][31:0]) + SE64(Mres[1][31:0]);
```

**Return** value stored in long long type

#### Parameters

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMALDRS(long long t, unsigned long a, unsigned long b)**  
SMALDRS (Signed Multiply Two Halves & Reverse Subtract & Add 64- bit)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
SMALDS Rd, Rs1, Rs2
SMALDRS Rd, Rs1, Rs2
SMALXDS Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then perform a subtraction operation between the two 32-bit results. Then add the subtraction result to the 64-bit value of an even/odd pair of registers (RV32) or a register (RV64). The addition result is written back to the register-pair.

- SMALDS:  $rt \text{ pair} + (top * top - bottom * bottom)$  (all 32-bit elements)
- SMALDRS:  $rt \text{ pair} + (bottom * bottom - top * top)$  (all 32-bit elements)
- SMALXDS:  $rt \text{ pair} + (top * bottom - bottom * top)$  (all 32-bit elements)

**RV32 Description:** For the SMALDS instruction, it multiplies the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of Rs1 with the top 16-bit content of Rs2. For the SMALDRS instruction, it multiplies the top 16-bit content of Rs1 with the top 16-bit content of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2. For the SMALXDS instruction, it multiplies the bottom 16-bit content of Rs1 with the top 16-bit content of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of Rs1 with the bottom 16-bit content of Rs2. The subtraction result is then added to the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit addition result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the operand and the even 2d register of the pair contains the low 32-bit of the operand.

**RV64 Description:** For the SMALDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMALDRS instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMALXDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The subtraction results are then added to the 64-bit value of Rd. The 64-bit addition result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

#### Operations:

```
* RV32:
Mres[31:0] = (Rs1.H[1] * Rs2.H[1]) - (Rs1.H[0] * Rs2.H[0]); // SMALDS
Mres[31:0] = (Rs1.H[0] * Rs2.H[0]) - (Rs1.H[1] * Rs2.H[1]); // SMALDRS
Mres[31:0] = (Rs1.H[1] * Rs2.H[0]) - (Rs1.H[0] * Rs2.H[1]); // SMALXDS
Idx0 = CONCAT(Rd(4,1),1'b0); Idx1 = CONCAT(Rd(4,1),1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] + SE64(Mres[31:0]);

* RV64:
// SMALDS
Mres[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[1]) - (Rs1.W[0].H[0] * Rs2.W[0].
↪H[0]);
Mres[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[0].H[1]) - (Rs1.W[1].H[0] * Rs2.W[1].
↪H[0]);
// SMALDRS
Mres[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[0]) - (Rs1.W[0].H[1] * Rs2.W[0].
↪H[1]);
Mres[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[0].H[0]) - (Rs1.W[1].H[1] * Rs2.W[1].
↪H[1]);
// SMALXDS
```

(continues on next page)

(continued from previous page)

```

Mres[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[0]) - (Rs1.W[0].H[0] * Rs2.W[0].
↪H[1]);
Mres[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[0].H[0]) - (Rs1.W[1].H[0] * Rs2.W[1].
↪H[1]);
Rd = Rd + SE64(Mres[0][31:0]) + SE64(Mres[1][31:0]);

```

**Return** value stored in long long type

**Parameters**

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMALXDS(long long t, unsigned long a, unsigned long b)**  
 SMALXDS (Signed Crossed Multiply Two Halfs & Subtract & Add 64-bit)

**Type:** DSP (64-bit Profile)

**Syntax:**

```

SMALDS Rd, Rs1, Rs2
SMALDRS Rd, Rs1, Rs2
SMALXDS Rd, Rs1, Rs2

```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then perform a subtraction operation between the two 32-bit results. Then add the subtraction result to the 64-bit value of an even/odd pair of registers (RV32) or a register (RV64). The addition result is written back to the register-pair.

- SMALDS: rt pair + (top\*top - bottom\*bottom) (all 32-bit elements)
- SMALDRS: rt pair + (bottom\*bottom - top\*top) (all 32-bit elements)
- SMALXDS: rt pair + (top\*bottom - bottom\*top) (all 32-bit elements)

**RV32 Description:** For the SMALDS instruction, it multiplies the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of Rs1 with the top 16-bit content of Rs2. For the SMALDRS instruction, it multiplies the top 16-bit content of Rs1 with the top 16-bit content of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2. For the SMALXDS instruction, it multiplies the bottom 16-bit content of Rs1 with the top 16-bit content of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of Rs1 with the bottom 16-bit content of Rs2. The subtraction result is then added to the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit addition result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the operand and the even 2d register of the pair contains the low 32-bit of the operand.

**RV64 Description:** For the SMALDS instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMALDRS instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. For the SMALXDS instruction, it multiplies the bottom

16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2 and then subtracts the result from the result of multiplying the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2. The subtraction results are then added to the 64-bit value of Rd. The 64-bit addition result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

#### Operations:

```
* RV32:
Mres[31:0] = (Rs1.H[1] * Rs2.H[1]) - (Rs1.H[0] * Rs2.H[0]); // SMALDS
Mres[31:0] = (Rs1.H[0] * Rs2.H[0]) - (Rs1.H[1] * Rs2.H[1]); // SMALDRS
Mres[31:0] = (Rs1.H[1] * Rs2.H[0]) - (Rs1.H[0] * Rs2.H[1]); // SMALXDS
Idx0 = CONCAT(Rd(4,1),1'b0); Idx1 = CONCAT(Rd(4,1),1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] + SE64(Mres[31:0]);

* RV64:
// SMALDS
Mres[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[1]) - (Rs1.W[0].H[0] * Rs2.W[0].
↪H[0]);
Mres[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[0].H[1]) - (Rs1.W[1].H[0] * Rs2.W[1].
↪H[0]);
// SMALDRS
Mres[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[0]) - (Rs1.W[0].H[1] * Rs2.W[0].
↪H[1]);
Mres[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[0].H[0]) - (Rs1.W[1].H[1] * Rs2.W[1].
↪H[1]);
// SMALXDS
Mres[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[0]) - (Rs1.W[0].H[0] * Rs2.W[0].
↪H[1]);
Mres[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[0].H[0]) - (Rs1.W[1].H[0] * Rs2.W[1].
↪H[1]);
Rd = Rd + SE64(Mres[0][31:0]) + SE64(Mres[1][31:0]);
```

**Return** value stored in long long type

#### Parameters

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMSLDA(long long t, unsigned long a, unsigned long b)**  
SMSLDA (Signed Multiply Two Halfs & Add & Subtract 64-bit)

**Type:** DSP (64-bit Profile)

#### Syntax:

```
SMSLDA Rd, Rs1, Rs2
SMSLXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then subtracts the two 32-bit results from the 64-bit value of an even/odd pair of registers (RV32) or a register (RV64). The subtraction result is written back to the register-pair.

- SMSLDA: rd pair - top\*top - bottom\*bottom (all 32-bit elements)
- SMSLXDA: rd pair - top\*bottom - bottom\*top (all 32-bit elements)

**RV32 Description:** For the SMSLDA instruction, it multiplies the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and multiplies the top 16-bit content of Rs1 with the top 16-bit content of Rs2. For the SMSLXDA instruction, it multiplies the top 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and multiplies the bottom 16-bit content of Rs1 with the top 16-bit content of Rs2. The two multiplication results are subtracted from the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit subtraction result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** For the SMSLDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMSLXDA instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. The four multiplication results are subtracted from the 64-bit value of Rd. The 64-bit subtraction result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

#### Operations:

```
* RV32:
// SMSLDA
Mres0[31:0] = (Rs1.H[0] * Rs2.H[0]);
Mres1[31:0] = (Rs1.H[1] * Rs2.H[1]);
// SMSLXDA
Mres0[31:0] = (Rs1.H[0] * Rs2.H[1]);
Mres1[31:0] = (Rs1.H[1] * Rs2.H[0]);
Idx0 = CONCAT(Rd(4,1), 1'b0); Idx1 = CONCAT(Rd(4,1), 1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] - SE64(Mres0[31:0]) - SE64(Mres1[31:0]);
* RV64:
// SMSLDA
Mres0[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[0]);
Mres1[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[1]);
Mres0[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[1].H[0]);
Mres1[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[1].H[1]);
// SMSLXDA
Mres0[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[1]);
Mres1[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[0]);
Mres0[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[1].H[1]);
Mres1[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[1].H[0]);
Rd = Rd - SE64(Mres0[0][31:0]) - SE64(Mres1[0][31:0]) - SE64(Mres0[1][31:0]) - SE64(Mres1[1][31:0]);
```

**Return** value stored in long long type

#### Parameters

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long long \_\_RV\_SMSLXDA(long long t, unsigned long a, unsigned long b)**  
SMSLXDA (Signed Crossed Multiply Two Halves & Add & Subtract 64-bit)

**Type:** DSP (64-bit Profile)

**Syntax:**

```
SMSLDA Rd, Rs1, Rs2
SMSLXDA Rd, Rs1, Rs2
```

**Purpose:** Do two signed 16-bit multiplications from the 32-bit elements of two registers; and then subtracts the two 32-bit results from the 64-bit value of an even/odd pair of registers (RV32) or a register (RV64). The subtraction result is written back to the register-pair.

- SMSLDA: rd pair - top\*top - bottom\*bottom (all 32-bit elements)
- SMSLXDA: rd pair - top\*bottom - bottom\*top (all 32-bit elements)

**RV32 Description:** For the SMSLDA instruction, it multiplies the bottom 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and multiplies the top 16-bit content of Rs1 with the top 16-bit content of Rs2. For the SMSLXDA instruction, it multiplies the top 16-bit content of Rs1 with the bottom 16-bit content of Rs2 and multiplies the bottom 16-bit content of Rs1 with the top 16-bit content of Rs2. The two multiplication results are subtracted from the 64-bit value of an even/odd pair of registers specified by Rd(4,1). The 64-bit subtraction result is written back to the register-pair. The 16-bit values of Rs1 and Rs2, and the 64-bit value of the register-pair are treated as signed integers. Rd(4,1), i.e., d, determines the even/odd pair group of the two registers. Specifically, the register pair includes register 2d and 2d+1. The odd 2d+1 register of the pair contains the high 32-bit of the result and the even 2d register of the pair contains the low 32-bit of the result.

**RV64 Description:** For the SMSLDA instruction, it multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and multiplies the top 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. For the SMSLXDA instruction, it multiplies the top 16-bit content of the 32-bit elements of Rs1 with the bottom 16-bit content of the 32-bit elements of Rs2 and multiplies the bottom 16-bit content of the 32-bit elements of Rs1 with the top 16-bit content of the 32-bit elements of Rs2. The four multiplication results are subtracted from the 64-bit value of Rd. The 64-bit subtraction result is written back to Rd. The 16-bit values of Rs1 and Rs2, and the 64-bit value of Rd are treated as signed integers.

**Operations:**

```
* RV32:
// SMSLDA
Mres0[31:0] = (Rs1.H[0] * Rs2.H[0]);
Mres1[31:0] = (Rs1.H[1] * Rs2.H[1]);
// SMSLXDA
Mres0[31:0] = (Rs1.H[0] * Rs2.H[1]);
Mres1[31:0] = (Rs1.H[1] * Rs2.H[0]);
Idx0 = CONCAT(Rd(4,1),1'b0); Idx1 = CONCAT(Rd(4,1),1'b1);
R[Idx1].R[Idx0] = R[Idx1].R[Idx0] - SE64(Mres0[31:0]) - SE64(Mres1[31:0]);
* RV64:
// SMSLDA
Mres0[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[0]);
Mres1[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[1]);
Mres0[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[1].H[0]);
Mres1[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[1].H[1]);
// SMSLXDA
Mres0[0][31:0] = (Rs1.W[0].H[0] * Rs2.W[0].H[1]);
Mres1[0][31:0] = (Rs1.W[0].H[1] * Rs2.W[0].H[0]);
Mres0[1][31:0] = (Rs1.W[1].H[0] * Rs2.W[1].H[1]);
Mres1[1][31:0] = (Rs1.W[1].H[1] * Rs2.W[1].H[0]);
Rd = Rd - SE64(Mres0[0][31:0]) - SE64(Mres1[0][31:0]) - SE64(Mres0[1][31:0]) -
SE64(Mres1[1][31:0]);
```

**Return** value stored in long long type

**Parameters**

- [in] t: long long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

*group* **NMSIS\_Core\_DSP\_Intrinsic\_64B\_PROFILE**  
64-bit Profile Instructions

## RV64 Only Instructions

### (RV64 Only) SIMD 32-bit Add/Subtract Instructions

```

__STATIC_FORCEINLINE unsigned long __RV_ADD32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_CRAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_CRSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KADD32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KCRAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KCRSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KSTAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KSTSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_KSUB32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RADD32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RCRAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RCRSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RSTAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RSTSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_RSUB32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_STAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_STSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_SUB32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKADD32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKCRAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKCRSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKSTAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKSTSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_UKSUB32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URADD32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URCRAS32(unsigned long a, unsigned long b)

```

```

__STATIC_FORCEINLINE unsigned long __RV_URCRSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URSTAS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URSTSA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE unsigned long __RV_URSUB32(unsigned long a, unsigned long b)

```

group **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_SIMD\_32B\_ADDSUB**  
(RV64 Only) SIMD 32-bit Add/Subtract Instructions

The following tables list instructions that are only present in RV64. There are 30 SIMD 32-bit addition or subtraction instructions. There are 4 SIMD16-bit Packing Instructions.

## Functions

```

__STATIC_FORCEINLINE unsigned long __RV_ADD32(unsigned long a, unsigned long b)
ADD32 (SIMD 32-bit Addition)

```

**Type:** SIMD (RV64 Only)

**Syntax:**

```
ADD32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit integer element additions simultaneously.

**Description:** This instruction adds the 32-bit integer elements in Rs1 with the 32-bit integer elements in Rs2, and then writes the 32-bit element results to Rd.

**Note:** This instruction can be used for either signed or unsigned addition.

**Operations:**

```

Rd.W[x] = Rs1.W[x] + Rs2.W[x];
for RV64: x=1..0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```

__STATIC_FORCEINLINE unsigned long __RV_CRAS32(unsigned long a, unsigned long b)
CRAS32 (SIMD 32-bit Cross Addition & Subtraction)

```

**Type:** SIMD (RV64 Only)

**Syntax:**

```
CRAS32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit integer element addition and 32-bit integer element subtraction in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements.

**Description:** This instruction adds the 32-bit integer element in [63:32] of Rs1 with the 32-bit integer element in [31:0] of Rs2, and writes the result to [63:32] of Rd; at the same time, it subtracts the 32-bit integer element in [63:32] of Rs2 from the 32-bit integer element in [31:0] of Rs1, and writes the result to [31:0] of Rd.

**Note:** This instruction can be used for either signed or unsigned operations.



**Operations:**

```
Rd.W[1] = Rs1.W[1] + Rs2.W[0];
Rd.W[0] = Rs1.W[0] - Rs2.W[1];
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_CRSA32(unsigned long a, unsigned long b)**  
CRSA32 (SIMD 32-bit Cross Subtraction & Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
CRSA32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit integer element subtraction and 32-bit integer element addition in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements. \*Description: \* This instruction subtracts the 32-bit integer element in [31:0] of Rs2 from the 32-bit integer element in [63:32] of Rs1, and writes the result to [63:32] of Rd; at the same time, it adds the 32-bit integer element in [31:0] of Rs1 with the 32-bit integer element in [63:32] of Rs2, and writes the result to [31:0] of Rd

**Note:** This instruction can be used for either signed or unsigned operations.

**Operations:**

```
Rd.W[1] = Rs1.W[1] - Rs2.W[0];
Rd.W[0] = Rs1.W[0] + Rs2.W[1];
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KADD32(unsigned long a, unsigned long b)**  
KADD32 (SIMD 32-bit Signed Saturating Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KADD32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element saturating additions simultaneously.

**Description:** This instruction adds the 32-bit signed integer elements in Rs1 with the 32-bit signed integer elements in Rs2. If any of the results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```

res[x] = Rs1.W[x] + Rs2.W[x];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV64: x=1...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KCRAS32(unsigned long a, unsigned long b)**  
 KCRAS32 (SIMD 32-bit Signed Saturating Cross Addition & Subtraction)

**Type:** SIM (RV64 Only)

#### Syntax:

```
KCRAS32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element saturating addition and 32-bit signed integer element saturating subtraction in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements.

**Description:** This instruction adds the 32-bit integer element in [63:32] of Rs1 with the 32-bit integer element in [31:0] of Rs2; at the same time, it subtracts the 32-bit integer element in [63:32] of Rs2 from the 32-bit integer element in [31:0] of Rs1. If any of the results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [63:32] of Rd for addition and [31:0] of Rd for subtraction.

#### Operations:

```

res[1] = Rs1.W[1] + Rs2.W[0];
res[0] = Rs1.W[0] - Rs2.W[1];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[1] = res[1];
Rd.W[0] = res[0];
for RV64, x=1...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KCRSA32(unsigned long a, unsigned long b)**  
 KCRSA32 (SIMD 32-bit Signed Saturating Cross Subtraction & Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KCRSA32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element saturating subtraction and 32-bit signed integer element saturating addition in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements. \*Description: \* This instruction subtracts the 32-bit integer element in [31:0] of Rs2 from the 32-bit integer element in [63:32] of Rs1; at the same time, it adds the 32-bit integer element in [31:0] of Rs1 with the 32-bit integer element in [63:32] of Rs2. If any of the results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [63:32] of Rd for subtraction and [31:0] of Rd for addition.

**Operations:**

```
res[1] = Rs1.W[1] - Rs2.W[0];
res[0] = Rs1.W[0] + Rs2.W[1];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[1] = res[1];
Rd.W[0] = res[0];
for RV64, x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSTAS32(unsigned long a, unsigned long b)**  
 KSTAS32 (SIMD 32-bit Signed Saturating Straight Addition & Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KSTAS32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element saturating addition and 32-bit signed integer element saturating subtraction in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements.

**Description:** This instruction adds the 32-bit integer element in [63:32] of Rs1 with the 32-bit integer element in [63:32] of Rs2; at the same time, it subtracts the 32-bit integer element in [31:0] of Rs2 from the 32-bit integer element in [31:0] of Rs1. If any of the results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [63:32] of Rd for addition and [31:0] of Rd for subtraction.

**Operations:**

```

res[1] = Rs1.W[1] + Rs2.W[1];
res[0] = Rs1.W[0] - Rs2.W[0];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[1] = res[1];
Rd.W[0] = res[0];
for RV64, x=1...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KTSA32(unsigned long a, unsigned long b)**  
 KTSA32 (SIMD 32-bit Signed Saturating Straight Subtraction & Addition)

**Type:** SIM (RV64 Only)

#### Syntax:

```
KTSA32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element saturating subtraction and 32-bit signed integer element saturating addition in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements. \*Description: \* This instruction subtracts the 32-bit integer element in [63:32] of Rs2 from the 32-bit integer element in [63:32] of Rs1; at the same time, it adds the 32-bit integer element in [31:0] of Rs1 with the 32-bit integer element in [31:0] of Rs2. If any of the results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [63:32] of Rd for subtraction and [31:0] of Rd for addition.

#### Operations:

```

res[1] = Rs1.W[1] - Rs2.W[1];
res[0] = Rs1.W[0] + Rs2.W[0];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[1] = res[1];
Rd.W[0] = res[0];
for RV64, x=1...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSUB32(unsigned long a, unsigned long b)**  
 KSUB32 (SIMD 32-bit Signed Saturating Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KSUB32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer elements saturating subtractions simultaneously.

**Description:** This instruction subtracts the 32-bit signed integer elements in Rs2 from the 32-bit signed integer elements in Rs1. If any of the results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```
res[x] = Rs1.W[x] - Rs2.W[x];
if (res[x] > (2^31)-1) {
    res[x] = (2^31)-1;
    OV = 1;
} else if (res[x] < -2^31) {
    res[x] = -2^31;
    OV = 1;
}
Rd.W[x] = res[x];
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RADD32(unsigned long a, unsigned long b)**  
 RADD32 (SIMD 32-bit Signed Halving Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
RADD32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element additions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 32-bit signed integer elements in Rs1 with the 32-bit signed integer elements in Rs2. The results are first arithmetically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Rs1 = 0x7FFFFFFF, Rs2 = 0x7FFFFFFF Rd = 0x7FFFFFFF
* Rs1 = 0x80000000, Rs2 = 0x80000000 Rd = 0x80000000
* Rs1 = 0x40000000, Rs2 = 0x80000000 Rd = 0xE0000000
```

**Operations:**

```
Rd.W[x] = (Rs1.W[x] + Rs2.W[x]) s>> 1;
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RCRAS32(unsigned long a, unsigned long b)**  
RCRAS32 (SIMD 32-bit Signed Halving Cross Addition & Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
RCRAS32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element addition and 32-bit signed integer element subtraction in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 32-bit signed integer element in [63:32] of Rs1 with the 32-bit signed integer element in [31:0] of Rs2, and subtracts the 32-bit signed integer element in [63:32] of Rs2 from the 32-bit signed integer element in [31:0] of Rs1. The element results are first arithmetically right-shifted by 1 bit and then written to [63:32] of Rd for addition and [31:0] of Rd for subtraction.

**Examples:**

```
Please see `RADD32` and `RSUB32` instructions.
```

**Operations:**

```
Rd.W[1] = (Rs1.W[1] + Rs2.W[0]) s>> 1;  
Rd.W[0] = (Rs1.W[0] - Rs2.W[1]) s>> 1;
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RCRSA32(unsigned long a, unsigned long b)**  
RCRSA32 (SIMD 32-bit Signed Halving Cross Subtraction & Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
RCRSA32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element subtraction and 32-bit signed integer element addition in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 32-bit signed integer element in [31:0] of Rs2 from the 32-bit signed integer element in [63:32] of Rs1, and adds the 32-bit signed element integer in [31:0] of Rs1 with the 32-bit signed integer element in [63:32] of Rs2. The two results are first arithmetically right-shifted by 1 bit and then written to [63:32] of Rd for subtraction and [31:0] of Rd for addition.

**Examples:**

Please see `RADD32` and `RSUB32` instructions.

#### Operations:

```
Rd.W[1] = (Rs1.W[1] - Rs2.W[0]) s>> 1;
Rd.W[0] = (Rs1.W[0] + Rs2.W[1]) s>> 1;
```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RSTAS32(unsigned long a, unsigned long b)**  
RSTAS32 (SIMD 32-bit Signed Halving Straight Addition & Subtraction)

**Type:** SIMD (RV64 Only)

#### Syntax:

```
RSTAS32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element addition and 32-bit signed integer element subtraction in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 32-bit signed integer element in [63:32] of Rs1 with the 32-bit signed integer element in [63:32] of Rs2, and subtracts the 32-bit signed integer element in [31:0] of Rs2 from the 32-bit signed integer element in [31:0] of Rs1. The element results are first arithmetically right-shifted by 1 bit and then written to [63:32] of Rd for addition and [31:0] of Rd for subtraction.

#### Examples:

Please see `RADD32` and `RSUB32` instructions.

#### Operations:

```
Rd.W[1] = (Rs1.W[1] + Rs2.W[1]) s>> 1;
Rd.W[0] = (Rs1.W[0] - Rs2.W[0]) s>> 1;
```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RSTSA32(unsigned long a, unsigned long b)**  
RSTSA32 (SIMD 32-bit Signed Halving Straight Subtraction & Addition)

**Type:** SIMD (RV64 Only)

#### Syntax:

```
RSTSA32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element subtraction and 32-bit signed integer element addition in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 32-bit signed integer element in [63:32] of Rs2 from the 32-bit signed integer element in [63:32] of Rs1, and adds the 32-bit signed element integer in [31:0] of Rs1 with the 32-bit signed integer element in [31:0] of Rs2. The two results are first arithmetically right-shifted by 1 bit and then written to [63:32] of Rd for subtraction and [31:0] of Rd for addition.

**Examples:**

Please see `RADD32` and `RSUB32` instructions.

**Operations:**

```
Rd.W[1] = (Rs1.W[1] - Rs2.W[1]) s>> 1;
Rd.W[0] = (Rs1.W[0] + Rs2.W[0]) s>> 1;
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_RSUB32(unsigned long a, unsigned long b)**  
 RSUB32 (SIMD 32-bit Signed Halving Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
RSUB32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer element subtractions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 32-bit signed integer elements in Rs2 from the 32-bit signed integer elements in Rs1. The results are first arithmetically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Ra = 0x7FFFFFFF, Rb = 0x80000000 Rt = 0x7FFFFFFF
* Ra = 0x80000000, Rb = 0x7FFFFFFF Rt = 0x80000000
* Ra = 0x80000000, Rb = 0x40000000 Rt = 0xA0000000
```

**Operations:**

```
Rd.W[x] = (Rs1.W[x] - Rs2.W[x]) s>> 1;
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b



**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_STAS32(unsigned long a, unsigned long b)**  
STAS32 (SIMD 32-bit Straight Addition & Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
STAS32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit integer element addition and 32-bit integer element subtraction in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements.

**Description:** This instruction adds the 32-bit integer element in [63:32] of Rs1 with the 32-bit integer element in [63:32] of Rs2, and writes the result to [63:32] of Rd; at the same time, it subtracts the 32-bit integer element in [31:0] of Rs2 from the 32-bit integer element in [31:0] of Rs1, and writes the result to [31:0] of Rd.

**Note:** This instruction can be used for either signed or unsigned operations.

**Operations:**

```
Rd.W[1] = Rs1.W[1] + Rs2.W[1];
Rd.W[0] = Rs1.W[0] - Rs2.W[0];
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_STSA32(unsigned long a, unsigned long b)**  
STSA32 (SIMD 32-bit Straight Subtraction & Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
STSA32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit integer element subtraction and 32-bit integer element addition in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements. *\*Description:* \* This instruction subtracts the 32-bit integer element in [63:32] of Rs2 from the 32-bit integer element in [63:32] of Rs1, and writes the result to [63:32] of Rd; at the same time, it adds the 32-bit integer element in [31:0] of Rs1 with the 32-bit integer element in [31:0] of Rs2, and writes the result to [31:0] of Rd

**Note:** This instruction can be used for either signed or unsigned operations.

**Operations:**

```
Rd.W[1] = Rs1.W[1] - Rs2.W[1];
Rd.W[0] = Rs1.W[0] + Rs2.W[0];
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SUB32(unsigned long a, unsigned long b)**  
SUB32 (SIMD 32-bit Subtraction)

**Type:** DSP (RV64 Only)

**Syntax:**

```
SUB32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit integer element subtractions simultaneously.

**Description:** This instruction subtracts the 32-bit integer elements in Rs2 from the 32-bit integer elements in Rs1, and then writes the results to Rd.

**Note:** This instruction can be used for either signed or unsigned subtraction.

**Operations:**

```
Rd.W[x] = Rs1.W[x] - Rs2.W[x];  
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKADD32(unsigned long a, unsigned long b)**  
UKADD32 (SIMD 32-bit Unsigned Saturating Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
UKADD32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer element saturating additions simultaneously.

**Description:** This instruction adds the 32-bit unsigned integer elements in Rs1 with the 32-bit unsigned integer elements in Rs2. If any of the results are beyond the 32-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{32}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```
res[x] = Rs1.W[x] + Rs2.W[x];  
if (res[x] > (2^32)-1) {  
    res[x] = (2^32)-1;  
    OV = 1;  
}  
Rd.W[x] = res[x];  
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKCRAS32(unsigned long a, unsigned long b)**  
 UKCRAS32 (SIMD 32-bit Unsigned Saturating Cross Addition & Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
UKCRAS32 Rd, Rs1, Rs2
```

**Purpose:** Do one 32-bit unsigned integer element saturating addition and one 32-bit unsigned integer element saturating subtraction in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements.

**Description:** This instruction adds the 32-bit unsigned integer element in [63:32] of Rs1 with the 32-bit unsigned integer element in [31:0] of Rs2; at the same time, it subtracts the 32-bit unsigned integer element in [63:32] of Rs2 from the 32-bit unsigned integer element in [31:0] of Rs1. If any of the results are beyond the 32-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{32}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [63:32] of Rd for addition and [31:0] of Rd for subtraction.

**Operations:**

```
res1 = Rs1.W[1] + Rs2.W[0];
res2 = Rs1.W[0] - Rs2.W[1];
if (res1 > (2^32)-1) {
    res1 = (2^32)-1;
    OV = 1;
}
if (res2 < 0) {
    res2 = 0;
    OV = 1;
}
Rd.W[1] = res1;
Rd.W[0] = res2;
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKCRSA32(unsigned long a, unsigned long b)**  
 UKCRSA32 (SIMD 32-bit Unsigned Saturating Cross Subtraction & Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
UKCRSA32 Rd, Rs1, Rs2
```

**Purpose:** Do one 32-bit unsigned integer element saturating subtraction and one 32-bit unsigned integer element saturating addition in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements.

**Description:** This instruction subtracts the 32-bit unsigned integer element in [31:0] of Rs2 from the 32-bit unsigned integer element in [63:32] of Rs1; at the same time, it adds the 32-bit unsigned integer element in [63:32] of Rs2 with the 32-bit unsigned integer element in [31:0] of Rs1. If any of the results are beyond the 32-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{32}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [63:32] of Rd for subtraction and [31:0] of Rd for addition.

**Operations:**

```

res1 = Rs1.W[1] - Rs2.W[0];
res2 = Rs1.W[0] + Rs2.W[1];
if (res1 < 0) {
    res1 = 0;
    OV = 1;
} else if (res2 > (2^32)-1) {
    res2 = (2^32)-1;
    OV = 1;
}
Rd.W[1] = res1;
Rd.W[0] = res2;

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSTAS32(unsigned long a, unsigned long b)**  
 UKSTAS32 (SIMD 32-bit Unsigned Saturating Straight Addition & Subtraction)

**Type:** SIMD (RV64 Only)

#### Syntax:

```
UKSTAS32 Rd, Rs1, Rs2
```

**Purpose:** Do one 32-bit unsigned integer element saturating addition and one 32-bit unsigned integer element saturating subtraction in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements.

**Description:** This instruction adds the 32-bit unsigned integer element in [63:32] of Rs1 with the 32-bit unsigned integer element in [63:32] of Rs2; at the same time, it subtracts the 32-bit unsigned integer element in [31:0] of Rs2 from the 32-bit unsigned integer element in [31:0] of Rs1. If any of the results are beyond the 32-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{32}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [63:32] of Rd for addition and [31:0] of Rd for subtraction.

#### Operations:

```

res1 = Rs1.W[1] + Rs2.W[1];
res2 = Rs1.W[0] - Rs2.W[0];
if (res1 > (2^32)-1) {
    res1 = (2^32)-1;
    OV = 1;
}
if (res2 < 0) {
    res2 = 0;
    OV = 1;
}
Rd.W[1] = res1;
Rd.W[0] = res2;

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a

- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSTSA32(unsigned long a, unsigned long b)**  
UKSTSA32 (SIMD 32-bit Unsigned Saturating Straight Subtraction & Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
UKSTSA32 Rd, Rs1, Rs2
```

**Purpose:** Do one 32-bit unsigned integer element saturating subtraction and one 32-bit unsigned integer element saturating addition in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements.

**Description:** This instruction subtracts the 32-bit unsigned integer element in [63:32] of Rs2 from the 32-bit unsigned integer element in [63:32] of Rs1; at the same time, it adds the 32-bit unsigned integer element in [31:0] of Rs2 with the 32-bit unsigned integer element in [31:0] Rs1. If any of the results are beyond the 32-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{32}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to [63:32] of Rd for subtraction and [31:0] of Rd for addition.

**Operations:**

```
res1 = Rs1.W[1] - Rs2.W[1];
res2 = Rs1.W[0] + Rs2.W[0];
if (res1 < 0) {
    res1 = 0;
    OV = 1;
} else if (res2 > (2^32)-1) {
    res2 = (2^32)-1;
    OV = 1;
}
Rd.W[1] = res1;
Rd.W[0] = res2;
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UKSUB32(unsigned long a, unsigned long b)**  
UKSUB32 (SIMD 32-bit Unsigned Saturating Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
UKSUB32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer elements saturating subtractions simultaneously.

**Description:** This instruction subtracts the 32-bit unsigned integer elements in Rs2 from the 32-bit unsigned integer elements in Rs1. If any of the results are beyond the 32-bit unsigned number range ( $0 \leq \text{RES} \leq 2^{32}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```

res[x] = Rs1.W[x] - Rs2.W[x];
if (res[x] < 0) {
    res[x] = 0;
    OV = 1;
}
Rd.W[x] = res[x];
for RV64: x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URADD32(unsigned long a, unsigned long b)**  
URADD32 (SIMD 32-bit Unsigned Halving Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
URADD32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer element additions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 32-bit unsigned integer elements in Rs1 with the 32-bit unsigned integer elements in Rs2. The results are first logically right-shifted by 1 bit and then written to Rd.

**Examples:**

```

* Ra = 0x7FFFFFFF, Rb = 0x7FFFFFFF Rt = 0x7FFFFFFF
* Ra = 0x80000000, Rb = 0x80000000 Rt = 0x80000000
* Ra = 0x40000000, Rb = 0x80000000 Rt = 0x60000000

```

**Operations:**

```

Rd.W[x] = (Rs1.W[x] + Rs2.W[x]) u>> 1;
for RV64: x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URCRAS32(unsigned long a, unsigned long b)**  
URCRAS32 (SIMD 32-bit Unsigned Halving Cross Addition & Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
URCRAS32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer element addition and 32-bit unsigned integer element subtraction in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 32-bit unsigned integer element in [63:32] of Rs1 with the 32-bit unsigned integer element in [31:0] of Rs2, and subtracts the 32-bit unsigned integer element in [63:32] of Rs2 from the 32-bit unsigned integer element in [31:0] of Rs1. The element results are first logically right-shifted by 1 bit and then written to [63:32] of Rd for addition and [31:0] of Rd for subtraction.

**Examples:**

Please see `URADD32` and `URSUB32` instructions.

**Operations:**

```
Rd.W[1] = (Rs1.W[1] + Rs2.W[0]) u>> 1;
Rd.W[0] = (Rs1.W[0] - Rs2.W[1]) u>> 1;
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URCRSA32(unsigned long a, unsigned long b)**  
URCRSA32 (SIMD 32-bit Unsigned Halving Cross Subtraction & Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

URCRSA32 Rd, Rs1, Rs2

**Purpose:** Do 32-bit unsigned integer element subtraction and 32-bit unsigned integer element addition in a 64-bit chunk simultaneously. Operands are from crossed 32-bit elements. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 32-bit unsigned integer element in [31:0] of Rs2 from the 32-bit unsigned integer element in [63:32] of Rs1, and adds the 32-bit unsigned element integer in [31:0] of Rs1 with the 32-bit unsigned integer element in [63:32] of Rs2. The two results are first logically right-shifted by 1 bit and then written to [63:32] of Rd for subtraction and [31:0] of Rd for addition.

**Examples:**

Please see `URADD32` and `URSUB32` instructions.

**Operations:**

```
Rd.W[1] = (Rs1.W[1] - Rs2.W[0]) u>> 1;
Rd.W[0] = (Rs1.W[0] + Rs2.W[1]) u>> 1;
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URSTAS32(unsigned long a, unsigned long b)**  
 URSTAS32 (SIMD 32-bit Unsigned Halving Straight Addition & Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
URSTAS32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer element addition and 32-bit unsigned integer element subtraction in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements. The results are halved to avoid overflow or saturation.

**Description:** This instruction adds the 32-bit unsigned integer element in [63:32] of Rs1 with the 32-bit unsigned integer element in [63:32] of Rs2, and subtracts the 32-bit unsigned integer element in [31:0] of Rs2 from the 32-bit unsigned integer element in [31:0] of Rs1. The element results are first logically right-shifted by 1 bit and then written to [63:32] of Rd for addition and [31:0] of Rd for subtraction.

**Examples:**

```
Please see `URADD32` and `URSUB32` instructions.
```

**Operations:**

```
Rd.W[1] = (Rs1.W[1] + Rs2.W[1]) u>> 1;
Rd.W[0] = (Rs1.W[0] - Rs2.W[0]) u>> 1;
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URSTSA32(unsigned long a, unsigned long b)**  
 URSTSA32 (SIMD 32-bit Unsigned Halving Straight Subtraction & Addition)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
URSTSA32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer element subtraction and 32-bit unsigned integer element addition in a 64-bit chunk simultaneously. Operands are from corresponding 32-bit elements. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 32-bit unsigned integer element in [63:32] of Rs2 from the 32-bit unsigned integer element in [63:32] of Rs1, and adds the 32-bit unsigned element integer in [31:0] of Rs1 with the 32-bit unsigned integer element in [31:0] of Rs2. The two results are first logically right-shifted by 1 bit and then written to [63:32] of Rd for subtraction and [31:0] of Rd for addition.

**Examples:**

```
Please see `URADD32` and `URSUB32` instructions.
```

**Operations:**

```
Rd.W[1] = (Rs1.W[1] - Rs2.W[1]) u>> 1;
Rd.W[0] = (Rs1.W[0] + Rs2.W[0]) u>> 1;
```



**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_URSUB32(unsigned long a, unsigned long b)**  
 URSUB32 (SIMD 32-bit Unsigned Halving Subtraction)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
URSUB32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer element subtractions simultaneously. The results are halved to avoid overflow or saturation.

**Description:** This instruction subtracts the 32-bit unsigned integer elements in Rs2 from the 32-bit unsigned integer elements in Rs1. The results are first logically right-shifted by 1 bit and then written to Rd.

**Examples:**

```
* Ra = 0x7FFFFFFF, Rb = 0x80000000, Rt = 0xFFFFFFFF
* Ra = 0x80000000, Rb = 0x7FFFFFFF, Rt = 0x00000000
* Ra = 0x80000000, Rb = 0x40000000, Rt = 0x20000000
```

**Operations:**

```
Rd.W[x] = (Rs1.W[x] - Rs2.W[x]) u>> 1;
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## (RV64 Only) SIMD 32-bit Shift Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_KSLL32(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_KSLLI32(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_KSLRA32(unsigned long a, int b)
__STATIC_FORCEINLINE unsigned long __RV_KSLRA32_U(unsigned long a, int b)
__STATIC_FORCEINLINE unsigned long __RV_SLL32(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SLLI32(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRA32(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRA32_U(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRAI32(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRAI32_U(unsigned long a, unsigned int b)
```

```

__STATIC_FORCEINLINE unsigned long __RV_SRL32(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRL32_U(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRLI32(unsigned long a, unsigned int b)
__STATIC_FORCEINLINE unsigned long __RV_SRLI32_U(unsigned long a, unsigned int b)

```

group **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_SIMD\_32B\_SHIFT**

(RV64 Only) SIMD 32-bit Shift Instructions

there are 14 (RV64 Only) SIMD 32-bit Shift Instructions

## Functions

```

__STATIC_FORCEINLINE unsigned long __RV_KSLL32(unsigned long a, unsigned int b)
KSLL32 (SIMD 32-bit Saturating Shift Left Logical)

```

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KSLL32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit elements logical left shift operations with saturation simultaneously. The shift amount is a variable from a GPR.

**Description:** The 32-bit data elements in Rs1 are left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the low-order 5-bits of the value in the Rs2 register. Any shifted value greater than  $2^{31}-1$  is saturated to  $2^{31}-1$ . Any shifted value smaller than  $-2^{31}$  is saturated to  $-2^{31}$ . And the saturated results are written to Rd. If any saturation is performed, set OV bit to 1.

**Operations:**

```

sa = Rs2[4:0];
if (sa != 0) {
    res[(31+sa):0] = Rs1.W[x] << sa;
    if (res > (2^31)-1) {
        res = 0x7fffffff; OV = 1;
    } else if (res < -2^31) {
        res = 0x80000000; OV = 1;
    }
    Rd.W[x] = res[31:0];
} else {
    Rd = Rs1;
}
for RV64: x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

```

__STATIC_FORCEINLINE unsigned long __RV_KSLLI32(unsigned long a, unsigned int b)
KSLLI32 (SIMD 32-bit Saturating Shift Left Logical Immediate)

```

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KSLLI32 Rd, Rs1, imm5u
```

**Purpose:** Do 32-bit elements logical left shift operations with saturation simultaneously. The shift amount is an immediate value.

**Description:** The 32-bit data elements in Rs1 are left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the imm5u constant. Any shifted value greater than  $2^{31}-1$  is saturated to  $2^{31}-1$ . Any shifted value smaller than  $-2^{31}$  is saturated to  $-2^{31}$ . And the saturated results are written to Rd. If any saturation is performed, set OV bit to 1.

**Operations:**

```
sa = imm5u[4:0];
if (sa != 0) {
    res[(31+sa):0] = Rs1.W[x] << sa;
    if (res > (2^31)-1) {
        res = 0x7fffffff; OV = 1;
    } else if (res < -2^31) {
        res = 0x80000000; OV = 1;
    }
    Rd.W[x] = res[31:0];
} else {
    Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSLRA32(unsigned long a, int b)**  
KSLRA32 (SIMD 32-bit Shift Left Logical with Saturation or Shift Right Arithmetic)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KSLRA32 Rd, Rs1, Rs2
KSLRA32.u Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit elements logical left (positive) or arithmetic right (negative) shift operation with Q31 saturation for the left shift. The .u form performs additional rounding up operations for the right shift.

**Description:** The 32-bit data elements of Rs1 are left-shifted logically or right-shifted arithmetically based on the value of Rs2[5:0]. Rs2[5:0] is in the signed range of [-25, 25-1]. A positive Rs2[5:0] means logical left shift and a negative Rs2[5:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[5:0]. However, the behavior of Rs2[5:0]==-25 (0x20) is defined to be equivalent to the behavior of Rs2[5:0]==-(25-1) (0x21). The left-shifted results are saturated to the 32-bit signed integer range of  $[-2^{31}, 2^{31}-1]$ . For the .u form of the instruction, the right-shifted results are added a 1 to the most significant discarded bit position for rounding effect. After the shift, saturation, or rounding, the final results are written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:6] will not affect this instruction.

**Operations:**

```

if (Rs2[5:0] < 0) {
    sa = -Rs2[5:0];
    sa = (sa == 32)? 31 : sa;
    if (`.u` form) {
        res[31:-1] = SE33(Rs1.W[x][31:sa-1]) + 1;
        Rd.W[x] = res[31:0];
    } else {
        Rd.W[x] = SE32(Rs1.W[x][31:sa]);
    }
} else {
    sa = Rs2[4:0];
    res[(31+sa):0] = Rs1.W[x] <<(logic) sa;
    if (res > (2^31)-1) {
        res[31:0] = 0x7fffffff; OV = 1;
    } else if (res < -2^31) {
        res[31:0] = 0x80000000; OV = 1;
    }
    Rd.W[x] = res[31:0];
}
for RV64: x=1...0

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KSLRA32\_U(unsigned long a, int b)**  
 KSLRA32.u (SIMD 32-bit Shift Left Logical with Saturation or Rounding Shift Right Arithmetic)

**Type:** SIMD (RV64 Only)

#### Syntax:

```

KSLRA32 Rd, Rs1, Rs2
KSLRA32.u Rd, Rs1, Rs2

```

**Purpose:** Do 32-bit elements logical left (positive) or arithmetic right (negative) shift operation with Q31 saturation for the left shift. The .u form performs additional rounding up operations for the right shift.

**Description:** The 32-bit data elements of Rs1 are left-shifted logically or right-shifted arithmetically based on the value of Rs2[5:0]. Rs2[5:0] is in the signed range of [-25, 25-1]. A positive Rs2[5:0] means logical left shift and a negative Rs2[5:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[5:0]. However, the behavior of Rs2[5:0]==-25 (0x20) is defined to be equivalent to the behavior of Rs2[5:0]==-(25-1) (0x21). The left-shifted results are saturated to the 32-bit signed integer range of [-2^31, 2^31-1]. For the .u form of the instruction, the right-shifted results are added a 1 to the most significant discarded bit position for rounding effect. After the shift, saturation, or rounding, the final results are written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:6] will not affect this instruction.

#### Operations:

```

if (Rs2[5:0] < 0) {
    sa = -Rs2[5:0];
    sa = (sa == 32)? 31 : sa;
    if (`.u` form) {
        res[31:-1] = SE33(Rs1.W[x][31:sa-1]) + 1;
    }
}

```

(continues on next page)

(continued from previous page)

```

    Rd.W[x] = res[31:0];
} else {
    Rd.W[x] = SE32(Rs1.W[x][31:sa]);
}
} else {
    sa = Rs2[4:0];
    res[(31+sa):0] = Rs1.W[x] <<(logic) sa;
    if (res > (2^31)-1) {
        res[31:0] = 0x7fffffff; OV = 1;
    } else if (res < -2^31) {
        res[31:0] = 0x80000000; OV = 1;
    }
    Rd.W[x] = res[31:0];
}
for RV64: x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SLL32(unsigned long a, unsigned int b)**  
SLL32 (SIMD 32-bit Shift Left Logical)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SLL32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit elements logical left shift operations simultaneously. The shift amount is a variable from a GPR.

**Description:** The 32-bit elements in Rs1 are left-shifted logically. And the results are written to Rd. The shifted out bits are filled with zero and the shift amount is specified by the low-order 5-bits of the value in the Rs2 register.

**Operations:**

```

sa = Rs2[4:0];
Rd.W[x] = Rs1.W[x] << sa;
for RV64: x=1...0

```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SLLI32(unsigned long a, unsigned int b)**  
SLLI32 (SIMD 32-bit Shift Left Logical Immediate)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SLLI32 Rd, Rs1, imm5u[4:0]
```

**Purpose:** Do 32-bit element logical left shift operations simultaneously. The shift amount is an immediate value.

**Description:** The 32-bit elements in Rs1 are left-shifted logically. The shifted out bits are filled with zero and the shift amount is specified by the imm5u[4:0] constant. And the results are written to Rd.

**Operations:**

```
sa = imm5u[4:0];
Rd.W[x] = Rs1.W[x] << sa;
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRA32(unsigned long a, unsigned int b)**  
SRA32 (SIMD 32-bit Shift Right Arithmetic)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SRA32 Rd, Rs1, Rs2
SRA32.u Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit element arithmetic right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 32-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the data elements. The shift amount is specified by the low-order 5-bits of the value in the Rs2 register. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 32-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = Rs2[4:0];
if (sa > 0) {
    if (`.u` form) { // SRA32.u
        res[31:-1] = SE33(Rs1.W[x][31:sa-1]) + 1;
        Rd.W[x] = res[31:0];
    } else { // SRA32
        Rd.W[x] = SE32(Rs1.W[x][31:sa])
    }
} else {
    Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRA32\_U(unsigned long a, unsigned int b)**  
SRA32.u (SIMD 32-bit Rounding Shift Right Arithmetic)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SRA32 Rd, Rs1, Rs2
SRA32.u Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit element arithmetic right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 32-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the data elements. The shift amount is specified by the low-order 5-bits of the value in the Rs2 register. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 32-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = Rs2[4:0];
if (sa > 0) {
    if (`.u` form) { // SRA32.u
        res[31:-1] = SE33(Rs1.W[x][31:sa-1]) + 1;
        Rd.W[x] = res[31:0];
    } else { // SRA32
        Rd.W[x] = SE32(Rs1.W[x][31:sa])
    }
} else {
    Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRAI32(unsigned long a, unsigned int b)**  
SRAI32 (SIMD 32-bit Shift Right Arithmetic Immediate)

**Type:** DSP (RV64 Only)

**Syntax:**

```
SRAI32 Rd, Rs1, imm5u
SRAI32.u Rd, Rs1, imm5u
```

**Purpose:** Do 32-bit elements arithmetic right shift operations simultaneously. The shift amount is an immediate value. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 32-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the 32-bit data elements. The shift amount is specified by the imm5u constant.

For the rounding operation of the `.u` form, a value of 1 is added to the most significant discarded bit of each 32-bit data to calculate the final results. And the results are written to Rd.

#### Operations:

```
sa = imm5u[4:0];
if (sa > 0) {
  if (`.u` form) { // SRAI32.u
    res[31:-1] = SE33(Rs1.W[x][31:sa-1]) + 1;
    Rd.W[x] = res[31:0];
  } else { // SRAI32
    Rd.W[x] = SE32(Rs1.W[x][31:sa]);
  }
} else {
  Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRAI32\_U(unsigned long a, unsigned int b)**  
SRAI32.u (SIMD 32-bit Rounding Shift Right Arithmetic Immediate)

**Type:** DSP (RV64 Only)

#### Syntax:

```
SRAI32 Rd, Rs1, imm5u
SRAI32.u Rd, Rs1, imm5u
```

**Purpose:** Do 32-bit elements arithmetic right shift operations simultaneously. The shift amount is an immediate value. The `.u` form performs additional rounding up operations on the shifted results.

**Description:** The 32-bit data elements in Rs1 are right-shifted arithmetically, that is, the shifted out bits are filled with the sign-bit of the 32-bit data elements. The shift amount is specified by the imm5u constant. For the rounding operation of the `.u` form, a value of 1 is added to the most significant discarded bit of each 32-bit data to calculate the final results. And the results are written to Rd.

#### Operations:

```
sa = imm5u[4:0];
if (sa > 0) {
  if (`.u` form) { // SRAI32.u
    res[31:-1] = SE33(Rs1.W[x][31:sa-1]) + 1;
    Rd.W[x] = res[31:0];
  } else { // SRAI32
    Rd.W[x] = SE32(Rs1.W[x][31:sa]);
  }
} else {
  Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type



**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRL32(unsigned long a, unsigned int b)**  
SRL32 (SIMD 32-bit Shift Right Logical)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SRL32 Rd, Rs1, Rs2
SRL32.u Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit element logical right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 32-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the low-order 5-bits of the value in the Rs2 register. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 32-bit data element to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = Rs2[4:0];
if (sa > 0) {
    if (`.u` form) { // SRA32.u
        res[31:-1] = ZE33(Rs1.W[x][31:sa-1]) + 1;
        Rd.W[x] = res[31:0];
    } else { // SRA32
        Rd.W[x] = ZE32(Rs1.W[x][31:sa])
    }
} else {
    Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRL32\_U(unsigned long a, unsigned int b)**  
SRL32.u (SIMD 32-bit Rounding Shift Right Logical)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SRL32 Rd, Rs1, Rs2
SRL32.u Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit element logical right shift operations simultaneously. The shift amount is a variable from a GPR. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 32-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the low-order 5-bits of the value in the Rs2 register. For

the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 32-bit data element to calculate the final results. And the results are written to Rd.

#### Operations:

```
sa = Rs2[4:0];
if (sa > 0) {
    if (`.u` form) { // SRA32.u
        res[31:-1] = ZE33(Rs1.W[x][31:sa-1]) + 1;
        Rd.W[x] = res[31:0];
    } else { // SRA32
        Rd.W[x] = ZE32(Rs1.W[x][31:sa])
    }
} else {
    Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRLI32(unsigned long a, unsigned int b)**  
SRLI32 (SIMD 32-bit Shift Right Logical Immediate)

**Type:** SIMD (RV64 Only)

#### Syntax:

```
SRLI32 Rd, Rs1, imm5u
SRLI32.u Rd, Rs1, imm5u
```

**Purpose:** Do 32-bit elements logical right shift operations simultaneously. The shift amount is an immediate value. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 32-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the imm5u constant. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 32-bit data to calculate the final results. And the results are written to Rd.

#### Operations:

```
sa = imm5u[4:0];
if (sa > 0) {
    if (`.u` form) { // SRLI32.u
        res[31:-1] = ZE33(Rs1.W[x][31:sa-1]) + 1;
        Rd.W[x] = res[31:0];
    } else { // SRLI32
        Rd.W[x] = ZE32(Rs1.W[x][31:sa]);
    }
} else {
    Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SRLI32\_U(unsigned long a, unsigned int b)**  
SRLI32.u (SIMD 32-bit Rounding Shift Right Logical Immediate)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SRLI32 Rd, Rs1, imm5u
SRLI32.u Rd, Rs1, imm5u
```

**Purpose:** Do 32-bit elements logical right shift operations simultaneously. The shift amount is an immediate value. The .u form performs additional rounding up operations on the shifted results.

**Description:** The 32-bit data elements in Rs1 are right-shifted logically, that is, the shifted out bits are filled with zero. The shift amount is specified by the imm5u constant. For the rounding operation of the .u form, a value of 1 is added to the most significant discarded bit of each 32-bit data to calculate the final results. And the results are written to Rd.

**Operations:**

```
sa = imm5u[4:0];
if (sa > 0) {
    if (`.u` form) { // SRLI32.u
        res[31:-1] = ZE33(Rs1.W[x][31:sa-1]) + 1;
        Rd.W[x] = res[31:0];
    } else { // SRLI32
        Rd.W[x] = ZE32(Rs1.W[x][31:sa]);
    }
} else {
    Rd = Rs1;
}
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned int type of value stored in b

**(RV64 Only) SIMD 32-bit Miscellaneous Instructions**

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KABS32(unsigned long a)**

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SMAX32(unsigned long a, unsigned long b)**

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SMIN32(unsigned long a, unsigned long b)**

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UMAX32(unsigned long a, unsigned long b)**

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UMIN32(unsigned long a, unsigned long b)**

*group* **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_SIMD\_32B\_MISC**  
(RV64 Only) SIMD 32-bit Miscellaneous Instructions

there are 5 (RV64 Only) SIMD 32-bit Miscellaneous Instructions

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KABS32(unsigned long a)**

KABS32 (Scalar 32-bit Absolute Value with Saturation)

**Type:** DSP (RV64 Only) 24 20 19 15 14 12 11 7 KABS32 10010 Rs1 000 Rd 6 0 GE80B 1111111

**Syntax:**

```
KABS32 Rd, Rs1
```

**Purpose:** Get the absolute value of signed 32-bit integer elements in a general register.

**Description:** This instruction calculates the absolute value of signed 32-bit integer elements stored in Rs1. The results are written to Rd. This instruction with the minimum negative integer input of 0x80000000 will produce a saturated output of maximum positive integer of 0x7fffffff and the OV flag will be set to 1.

**Operations:**

```
if (Rs1.W[x] >= 0) {
    res[x] = Rs1.W[x];
} else {
    If (Rs1.W[x] == 0x80000000) {
        res[x] = 0x7fffffff;
        OV = 1;
    } else {
        res[x] = -Rs1.W[x];
    }
}
Rd.W[x] = res[x];
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SMAX32(unsigned long a, unsigned long b)**

SMAX32 (SIMD 32-bit Signed Maximum)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SMAX32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer elements finding maximum operations simultaneously.

**Description:** This instruction compares the 32-bit signed integer elements in Rs1 with the 32-bit signed integer elements in Rs2 and selects the numbers that is greater than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.W[x] = (Rs1.W[x] > Rs2.W[x]) ? Rs1.W[x] : Rs2.W[x];
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_SMIN32(unsigned long a, unsigned long b)**  
SMIN32 (SIMD 32-bit Signed Minimum)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
SMIN32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit signed integer elements finding minimum operations simultaneously.

**Description:** This instruction compares the 32-bit signed integer elements in Rs1 with the 32-bit signed integer elements in Rs2 and selects the numbers that is less than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.W[x] = (Rs1.W[x] < Rs2.W[x]) ? Rs1.W[x] : Rs2.W[x];  
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_UMAX32(unsigned long a, unsigned long b)**  
UMAX32 (SIMD 32-bit Unsigned Maximum)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
UMAX32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer elements finding maximum operations simultaneously.

**Description:** This instruction compares the 32-bit unsigned integer elements in Rs1 with the 32-bit unsigned integer elements in Rs2 and selects the numbers that is greater than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.W[x] = (Rs1.W[x] u> Rs2.W[x]) ? Rs1.W[x] : Rs2.W[x];  
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE unsigned long __RV_UMIN32(unsigned long a, unsigned long b)
    UMIN32 (SIMD 32-bit Unsigned Minimum)
```

**Type:** SIMD (RV64 Only)

**Syntax:**

```
UMIN32 Rd, Rs1, Rs2
```

**Purpose:** Do 32-bit unsigned integer elements finding minimum operations simultaneously.

**Description:** This instruction compares the 32-bit unsigned integer elements in Rs1 with the 32-bit unsigned integer elements in Rs2 and selects the numbers that is less than the other one. The selected results are written to Rd.

**Operations:**

```
Rd.W[x] = (Rs1.W[x] <u Rs2.W[x]) ? Rs1.W[x] : Rs2.W[x];
for RV64: x=1...0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## (RV64 Only) SIMD Q15 Saturating Multiply Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_KDMBB16(unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_KDMBT16(unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_KDMTT16(unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_KDMABB16(unsigned long t, unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_KDMABT16(unsigned long t, unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_KDMATT16(unsigned long t, unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_KHMBB16(unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_KHMBT16(unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_KHMTT16(unsigned long a, unsigned long b)
```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_SIMD\_Q15\_SAT\_MULT**

(RV64 Only) SIMD Q15 Saturating Multiply Instructions

there are 9 (RV64 Only) SIMD Q15 saturating Multiply Instructions

## Functions

```
__STATIC_FORCEINLINE unsigned long __RV_KDMBB16(unsigned long a, unsigned long b)
    KDMBB16 (SIMD Signed Saturating Double Multiply B16 x B16)
```

**Type:** SIMD (RV64 only)

**Syntax:**

```
KDMxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then double and saturate the Q31 results into the 32-bit chunks in the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the 32-bit portions in Rs2. The Q30 results are then doubled and saturated into Q31 values. The Q31 values are then written into the 32-bit chunks in Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFFFFFF and the overflow flag OV will be set.

#### Operations:

```
// KDMBB16: (x,y,z)=(0,0,0), (2,2,1)
// KDMBT16: (x,y,z)=(0,1,0), (2,3,1)
// KDMTT16: (x,y,z)=(1,1,0), (3,3,1)
aop[z] = Rs1.H[x]; bop[z] = Rs2.H[y];
If (0x8000 != aop[z] | 0x8000 != bop[z]) {
    Mresult[z] = aop[z] * bop[z];
    resQ31[z] = Mresult[z] << 1;
} else {
    resQ31[z] = 0x7FFFFFFF;
    OV = 1;
}
Rd.W[z] = resQ31[z];
```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE unsigned long __RV_KDMBT16(unsigned long a, unsigned long b)
KDMBT16 (SIMD Signed Saturating Double Multiply B16 x T16)
```

**Type:** SIMD (RV64 only)

#### Syntax:

```
KDMxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then double and saturate the Q31 results into the 32-bit chunks in the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the 32-bit portions in Rs2. The Q30 results are then doubled and saturated into Q31 values. The Q31 values are then written into the 32-bit chunks in Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFFFFFF and the overflow flag OV will be set.

#### Operations:

```
// KDMBB16: (x,y,z)=(0,0,0), (2,2,1)
// KDMBT16: (x,y,z)=(0,1,0), (2,3,1)
// KDMTT16: (x,y,z)=(1,1,0), (3,3,1)
```

(continues on next page)

(continued from previous page)

```

aop[z] = Rs1.H[x]; bop[z] = Rs2.H[y];
If (0x8000 != aop[z] | 0x8000 != bop[z]) {
    Mresult[z] = aop[z] * bop[z];
    resQ31[z] = Mresult[z] << 1;
} else {
    resQ31[z] = 0x7FFFFFFF;
    OV = 1;
}
Rd.W[z] = resQ31[z];

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KDMTT16(unsigned long a, unsigned long b)**  
 KDMTT16 (SIMD Signed Saturating Double Multiply T16 x T16)

**Type:** SIMD (RV64 only)

#### Syntax:

```
KDMxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then double and saturate the Q31 results into the 32-bit chunks in the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the 32-bit portions in Rs2. The Q30 results are then doubled and saturated into Q31 values. The Q31 values are then written into the 32-bit chunks in Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFFFFFF and the overflow flag OV will be set.

#### Operations:

```

// KDMBB16: (x,y,z)=(0,0,0), (2,2,1)
// KDMBT16: (x,y,z)=(0,1,0), (2,3,1)
// KDMTT16: (x,y,z)=(1,1,0), (3,3,1)
aop[z] = Rs1.H[x]; bop[z] = Rs2.H[y];
If (0x8000 != aop[z] | 0x8000 != bop[z]) {
    Mresult[z] = aop[z] * bop[z];
    resQ31[z] = Mresult[z] << 1;
} else {
    resQ31[z] = 0x7FFFFFFF;
    OV = 1;
}
Rd.W[z] = resQ31[z];

```

**Return** value stored in unsigned long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b



**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KDMABB16(unsigned long t, unsigned long a, unsigned long b)**  
 KDMABB16 (SIMD Signed Saturating Double Multiply Addition B16 x B16)

**Type:** SIMD (RV64 only)

**Syntax:**

KDMAxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
---

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then double and saturate the Q31 results, add the results with the values of the corresponding 32-bit chunks from the destination register and write the saturated addition results back into the corresponding 32-bit chunks of the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the corresponding 32-bit portions in Rs2. The Q30 results are then doubled and saturated into Q31 values. The Q31 values are then added with the content of the corresponding 32-bit portions of Rd. If the addition results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV flag is set to 1. The results after saturation are written back to Rd. When both the two Q15 inputs are 0x8000, saturation will happen and the overflow flag OV will be set.

**Operations:**

```
// KDMABB16: (x,y,z)=(0,0,0), (2,2,1)
// KDMABT16: (x,y,z)=(0,1,0), (2,3,1)
// KDMATT16: (x,y,z)=(1,1,0), (3,3,1)
aop[z] = Rs1.H[x]; bop[z] = Rs2.H[y];
If (0x8000 != aop[z] | 0x8000 != bop[z]) {
    Mresult[z] = aop[z] * bop[z];
    resQ31[z] = Mresult[z] << 1;
} else {
    resQ31[z] = 0x7FFFFFFF;
    OV = 1;
}
resadd[z] = Rd.W[z] + resQ31[z];
if (resadd[z] > (2^31)-1) {
    resadd[z] = (2^31)-1;
    OV = 1;
} else if (resadd[z] < -2^31) {
    resadd[z] = -2^31;
    OV = 1;
}
Rd.W[z] = resadd[z];
```

**Return** value stored in unsigned long type

**Parameters**

- [in] t: unsigned long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KDMABT16(unsigned long t, unsigned long a, unsigned long b)**  
 KDMABT16 (SIMD Signed Saturating Double Multiply Addition B16 x T16)

**Type:** SIMD (RV64 only)

**Syntax:**

```
KDMAxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then double and saturate the Q31 results, add the results with the values of the corresponding 32-bit chunks from the destination register and write the saturated addition results back into the corresponding 32-bit chunks of the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the corresponding 32-bit portions in Rs2. The Q30 results are then doubled and saturated into Q31 values. The Q31 values are then added with the content of the corresponding 32-bit portions of Rd. If the addition results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV flag is set to 1. The results after saturation are written back to Rd. When both the two Q15 inputs are 0x8000, saturation will happen and the overflow flag OV will be set.

#### Operations:

```
// KDMABB16: (x,y,z)=(0,0,0), (2,2,1)
// KDMABT16: (x,y,z)=(0,1,0), (2,3,1)
// KDMATT16: (x,y,z)=(1,1,0), (3,3,1)
aop[z] = Rs1.H[x]; bop[z] = Rs2.H[y];
If (0x8000 != aop[z] | 0x8000 != bop[z]) {
    Mresult[z] = aop[z] * bop[z];
    resQ31[z] = Mresult[z] << 1;
} else {
    resQ31[z] = 0x7FFFFFFF;
    OV = 1;
}
resadd[z] = Rd.W[z] + resQ31[z];
if (resadd[z] > (2^31)-1) {
    resadd[z] = (2^31)-1;
    OV = 1;
} else if (resadd[z] < -2^31) {
    resadd[z] = -2^31;
    OV = 1;
}
Rd.W[z] = resadd[z];
```

**Return** value stored in unsigned long type

#### Parameters

- [in] t: unsigned long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KDMATT16(unsigned long t, unsigned long a, unsigned long b)**  
KDMATT16 (SIMD Signed Saturating Double Multiply Addition T16 x T16)

**Type:** SIMD (RV64 only)

#### Syntax:

```
KDMAxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then double and saturate the Q31 results, add the results with the values of the corresponding 32-bit chunks from the destination register and write the saturated addition results back

into the corresponding 32-bit chunks of the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the corresponding 32-bit portions in Rs2. The Q30 results are then doubled and saturated into Q31 values. The Q31 values are then added with the content of the corresponding 32-bit portions of Rd. If the addition results are beyond the Q31 number range ( $-2^{31} \leq Q31 \leq 2^{31}-1$ ), they are saturated to the range and the OV flag is set to 1. The results after saturation are written back to Rd. When both the two Q15 inputs are 0x8000, saturation will happen and the overflow flag OV will be set.

#### Operations:

```
// KDMABB16: (x,y,z)=(0,0,0), (2,2,1)
// KDMABT16: (x,y,z)=(0,1,0), (2,3,1)
// KDMATT16: (x,y,z)=(1,1,0), (3,3,1)
aop[z] = Rs1.H[x]; bop[z] = Rs2.H[y];
If (0x8000 != aop[z] | 0x8000 != bop[z]) {
    Mresult[z] = aop[z] * bop[z];
    resQ31[z] = Mresult[z] << 1;
} else {
    resQ31[z] = 0x7FFFFFFF;
    OV = 1;
}
resadd[z] = Rd.W[z] + resQ31[z];
if (resadd[z] > (2^31)-1) {
    resadd[z] = (2^31)-1;
    OV = 1;
} else if (resadd[z] < -2^31) {
    resadd[z] = -2^31;
    OV = 1;
}
Rd.W[z] = resadd[z];
```

**Return** value stored in unsigned long type

#### Parameters

- [in] t: unsigned long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KHMBB16(unsigned long a, unsigned long b)**  
KHMBB16 (SIMD Signed Saturating Half Multiply B16 x B16)

**Type:** SIMD (RV64 Only)

#### Syntax:

```
KHMxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then right-shift 15 bits to turn the Q30 results into Q15 numbers again and saturate the Q15 results into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the 32-bit portion in Rs2. The Q30 results are then right-shifted 15- bits and saturated into Q15 values. The 32-bit Q15 values are then written into the 32-bit chunks in Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

**Operations:**

```
// KHMBB16: (x,y,z)=(0,0,0), (2,2,1)
// KHMBT16: (x,y,z)=(0,1,0), (2,3,1)
// KHMTT16: (x,y,z)=(1,1,0), (3,3,1)
aop = Rs1.H[x]; bop = Rs2.H[y];
If (0x8000 != aop | 0x8000 != bop) {
    Mresult[31:0] = aop * bop;
    res[15:0] = Mresult[30:15];
} else {
    res[15:0] = 0x7FFF;
    OV = 1;
}
Rd.W[z] = SE32(res[15:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KHMBT16(unsigned long a, unsigned long b)**  
KHMBT16 (SIMD Signed Saturating Half Multiply B16 x T16)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KHMxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then right-shift 15 bits to turn the Q30 results into Q15 numbers again and saturate the Q15 results into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the 32-bit portion in Rs2. The Q30 results are then right-shifted 15- bits and saturated into Q15 values. The 32-bit Q15 values are then written into the 32-bit chunks in Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

**Operations:**

```
// KHMBB16: (x,y,z)=(0,0,0), (2,2,1)
// KHMBT16: (x,y,z)=(0,1,0), (2,3,1)
// KHMTT16: (x,y,z)=(1,1,0), (3,3,1)
aop = Rs1.H[x]; bop = Rs2.H[y];
If (0x8000 != aop | 0x8000 != bop) {
    Mresult[31:0] = aop * bop;
    res[15:0] = Mresult[30:15];
} else {
    res[15:0] = 0x7FFF;
    OV = 1;
}
Rd.W[z] = SE32(res[15:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_KHMTT16(unsigned long a, unsigned long b)**  
KHMTT16 (SIMD Signed Saturating Half Multiply T16 x T16)

**Type:** SIMD (RV64 Only)

**Syntax:**

```
KHMxy16 Rd, Rs1, Rs2 (xy = BB, BT, TT)
```

**Purpose:** Multiply the signed Q15 integer contents of two 16-bit data in the corresponding portion of the 32-bit chunks in registers and then right-shift 15 bits to turn the Q30 results into Q15 numbers again and saturate the Q15 results into the destination register. If saturation happens, an overflow flag OV will be set.

**Description:** Multiply the top or bottom 16-bit Q15 content of the 32-bit portions in Rs1 with the top or bottom 16-bit Q15 content of the 32-bit portion in Rs2. The Q30 results are then right-shifted 15- bits and saturated into Q15 values. The 32-bit Q15 values are then written into the 32-bit chunks in Rd. When both the two Q15 inputs are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

**Operations:**

```
// KHMBB16: (x,y,z)=(0,0,0),(2,2,1)
// KHMBT16: (x,y,z)=(0,1,0),(2,3,1)
// KHMTT16: (x,y,z)=(1,1,0),(3,3,1)
aop = Rs1.H[x]; bop = Rs2.H[y];
If (0x8000 != aop | 0x8000 != bop) {
    Mresult[31:0] = aop * bop;
    res[15:0] = Mresult[30:15];
} else {
    res[15:0] = 0x7FFF;
    OV = 1;
}
Rd.W[z] = SE32(res[15:0]);
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

### (RV64 Only) 32-bit Multiply Instructions

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMBB32(unsigned long a, unsigned long b)**

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMBT32(unsigned long a, unsigned long b)**

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMTT32(unsigned long a, unsigned long b)**

*group* **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_32B\_MULT**

(RV64 Only) 32-bit Multiply Instructions

there is 3 RV64 Only) 32-bit Multiply Instructions

## Functions

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMBB32(unsigned long a, unsigned long b)**  
SMBB32 (Signed Multiply Bottom Word & Bottom Word)

**Type:** DSP (RV64 Only)

**Syntax:**

```
SMBB32 Rd, Rs1, Rs2
SMBT32 Rd, Rs1, Rs2
SMTT32 Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit element of a register with the signed 32-bit element of another register and write the 64-bit result to a third register.

- SMBB32: bottom\*bottom
- SMBT32: bottom\*top
- SMTT32: top\*top

**Description:** For the SMBB32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2. It is actually an alias of MULSR64 instruction. For the SMBT32 instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2. For the SMTT32 instruction, it multiplies the top 32-bit element of Rs1 with the top 32-bit element of Rs2. The 64-bit multiplication result is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
res = Rs1.W[0] * Rs2.W[0]; // SMBB32 res = Rs1.W[0] * Rs2.w[1]; // SMBT32 res_
↪ = Rs1.W[1] * Rs2.W[1];
// SMTT32 Rd = res;
```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMBT32(unsigned long a, unsigned long b)**  
SMBT32 (Signed Multiply Bottom Word & Top Word)

**Type:** DSP (RV64 Only)

**Syntax:**

```
SMBB32 Rd, Rs1, Rs2
SMBT32 Rd, Rs1, Rs2
SMTT32 Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit element of a register with the signed 32-bit element of another register and write the 64-bit result to a third register.

- SMBB32: bottom\*bottom
- SMBT32: bottom\*top
- SMTT32: top\*top

**Description:** For the SMBB32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2. It is actually an alias of MULSR64 instruction. For the SMBT32 instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2. For the SMTT32 instruction, it multiplies the top 32-bit element of Rs1 with the top 32-bit element of Rs2. The 64-bit multiplication result is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
res = Rs1.W[0] * Rs2.W[0]; // SMBB32 res = Rs1.W[0] * Rs2.w[1]; // SMBT32 res_
↔= Rs1.W[1] * Rs2.W[1];
// SMTT32 Rd = res;
```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMTT32(unsigned long a, unsigned long b)**  
SMTT32 (Signed Multiply Top Word & Top Word)

**Type:** DSP (RV64 Only)

**Syntax:**

```
SMBB32 Rd, Rs1, Rs2
SMBT32 Rd, Rs1, Rs2
SMTT32 Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit element of a register with the signed 32-bit element of another register and write the 64-bit result to a third register.

- SMBB32: bottom\*bottom
- SMBT32: bottom\*top
- SMTT32: top\*top

**Description:** For the SMBB32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2. It is actually an alias of MULSR64 instruction. For the SMBT32 instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2. For the SMTT32 instruction, it multiplies the top 32-bit element of Rs1 with the top 32-bit element of Rs2. The 64-bit multiplication result is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
res = Rs1.W[0] * Rs2.W[0]; // SMBB32 res = Rs1.W[0] * Rs2.w[1]; // SMBT32 res_
↔= Rs1.W[1] * Rs2.W[1];
// SMTT32 Rd = res;
```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**(RV64 Only) 32-bit Multiply & Add Instructions**

```

__STATIC_FORCEINLINE long __RV_KMABB32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMABT32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMATT32(long t, unsigned long a, unsigned long b)

```

group **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_32B\_MULT\_ADD**  
 (RV64 Only) 32-bit Multiply & Add Instructions

there are 3 (RV64 Only) 32-bit Multiply & Add Instructions

**Functions**

```

__STATIC_FORCEINLINE long __RV_KMABB32(long t, unsigned long a, unsigned long b)
KMABB32 (Saturating Signed Multiply Bottom Words & Add)

```

**Type:** DSP (RV64 Only)

**Syntax:**

```

KMABB32 Rd, Rs1, Rs2
KMABT32 Rd, Rs1, Rs2
KMATT32 Rd, Rs1, Rs2

```

**Purpose:** Multiply the signed 32-bit element in a register with the 32-bit element in another register and add the result to the content of 64-bit data in the third register. The addition result may be saturated and is written to the third register.

- KMABB32:  $rd + \text{bottom} * \text{bottom}$
- KMABT32:  $rd + \text{bottom} * \text{top}$
- KMATT32:  $rd + \text{top} * \text{top}$

**Description:** For the KMABB32 instruction, it multiplies the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2. For the KMABT32 instruction, it multiplies the bottom 32-bit element in Rs1 with the top 32-bit element in Rs2. For the KMATT32 instruction, it multiplies the top 32-bit element in Rs1 with the top 32-bit element in Rs2. The multiplication result is added to the content of 64-bit data in Rd. If the addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The result after saturation is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```

res = Rd + (Rs1.W[0] * Rs2.W[0]); // KMABB32
res = Rd + (Rs1.W[0] * Rs2.W[1]); // KMABT32
res = Rd + (Rs1.W[1] * Rs2.W[1]); // KMATT32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;
*Exceptions:* None

```

**Return** value stored in long type



**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMABT32(long t, unsigned long a, unsigned long b)**  
 KMABT32 (Saturating Signed Multiply Bottom & Top Words & Add)

**Type:** DSP (RV64 Only)

**Syntax:**

```
KMABB32 Rd, Rs1, Rs2
KMABT32 Rd, Rs1, Rs2
KMATT32 Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit element in a register with the 32-bit element in another register and add the result to the content of 64-bit data in the third register. The addition result may be saturated and is written to the third register.

- KMABB32:  $rd + \text{bottom} * \text{bottom}$
- KMABT32:  $rd + \text{bottom} * \text{top}$
- KMATT32:  $rd + \text{top} * \text{top}$

**Description:** For the KMABB32 instruction, it multiplies the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2. For the KMABT32 instruction, it multiplies the bottom 32-bit element in Rs1 with the top 32-bit element in Rs2. For the KMATT32 instruction, it multiplies the top 32-bit element in Rs1 with the top 32-bit element in Rs2. The multiplication result is added to the content of 64-bit data in Rd. If the addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The result after saturation is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
res = Rd + (Rs1.W[0] * Rs2.W[0]); // KMABB32
res = Rd + (Rs1.W[0] * Rs2.W[1]); // KMABT32
res = Rd + (Rs1.W[1] * Rs2.W[1]); // KMATT32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;
*Exceptions:* None
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_KMATT32(long t, unsigned long a, unsigned long b)
KMATT32 (Saturating Signed Multiply Top Words & Add)
```

**Type:** DSP (RV64 Only)

**Syntax:**

```
KMABB32 Rd, Rs1, Rs2
KMABT32 Rd, Rs1, Rs2
KMATT32 Rd, Rs1, Rs2
```

**Purpose:** Multiply the signed 32-bit element in a register with the 32-bit element in another register and add the result to the content of 64-bit data in the third register. The addition result may be saturated and is written to the third register.

- KMABB32:  $rd + \text{bottom} * \text{bottom}$
- KMABT32:  $rd + \text{bottom} * \text{top}$
- KMATT32:  $rd + \text{top} * \text{top}$

**Description:** For the KMABB32 instruction, it multiplies the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2. For the KMABT32 instruction, it multiplies the bottom 32-bit element in Rs1 with the top 32-bit element in Rs2. For the KMATT32 instruction, it multiplies the top 32-bit element in Rs1 with the top 32-bit element in Rs2. The multiplication result is added to the content of 64-bit data in Rd. If the addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The result after saturation is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```
res = Rd + (Rs1.W[0] * Rs2.W[0]); // KMABB32
res = Rd + (Rs1.W[0] * Rs2.W[1]); // KMABT32
res = Rd + (Rs1.W[1] * Rs2.W[1]); // KMATT32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;
*Exceptions: * None
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

### (RV64 Only) 32-bit Parallel Multiply & Add Instructions

```
__STATIC_FORCEINLINE long __RV_KMADA32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMAXDA32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMDA32(unsigned long a, unsigned long b)
```

```

__STATIC_FORCEINLINE long __RV_KMXDA32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMADS32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMADRS32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMAXDS32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMSDA32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_KMSXDA32(long t, unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMDS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMDRS32(unsigned long a, unsigned long b)
__STATIC_FORCEINLINE long __RV_SMXDS32(unsigned long a, unsigned long b)

```

*group* **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_32B\_PARALLEL\_MAC**

(RV64 Only) 32-bit Parallel Multiply & Add Instructions

there are 12 (RV64 Only) 32-bit Parallel Multiply & Add Instructions

## Functions

```

__STATIC_FORCEINLINE long __RV_KMADA32(long t, unsigned long a, unsigned long b)
KMADA32 (Saturating Signed Multiply Two Words and Two Adds)

```

**Type:** DSP (RV64 Only)

**Syntax:**

```

KMADA32 Rd, Rs1, Rs2
KMAXDA32 Rd, Rs1, Rs2

```

**Purpose:** Do two signed 32-bit multiplications from 32-bit data in two registers; and then adds the two 64-bit results and 64-bit data in a third register together. The addition result may be saturated.

- KMADA32:  $rd + top * top + bottom * bottom$
- KMAXDA32:  $rd + top * bottom + bottom * top$

**Description:** For the KMADA32 instruction, it multiplies the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2 and then adds the result to the result of multiplying the top 32-bit element in Rs1 with the top 32-bit element in Rs2. It is actually an alias of the KMAR64 instruction. For the KMAXDA32 instruction, it multiplies the top 32-bit element in Rs1 with the bottom 32-bit element in Rs2 and then adds the result to the result of multiplying the bottom 32-bit element in Rs1 with the top 32-bit element in Rs2. The result is added to the content of 64-bit data in Rd. If the addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The 64-bit result is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```

res = Rd + (Rs1.W[1] * Rs2.w[1]) + (Rs1.W[0] * Rs2.W[0]); // KMADA32
res = Rd + (Rs1.W[1] * Rs2.W[0]) + (Rs1.W[0] * Rs2.W[1]); // KMAXDA32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}

```

(continues on next page)

(continued from previous page)

```

}
Rd = res;

```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMAXDA32(long t, unsigned long a, unsigned long b)**  
 KMAXDA32 (Saturating Signed Crossed Multiply Two Words and Two Adds)

**Type:** DSP (RV64 Only)

**Syntax:**

```

KMADA32 Rd, Rs1, Rs2
KMAXDA32 Rd, Rs1, Rs2

```

**Purpose:** Do two signed 32-bit multiplications from 32-bit data in two registers; and then adds the two 64-bit results and 64-bit data in a third register together. The addition result may be saturated.

- KMADA32:  $rd + top * top + bottom * bottom$
- KMAXDA32:  $rd + top * bottom + bottom * top$

**Description:** For the KMADA32 instruction, it multiplies the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2 and then adds the result to the result of multiplying the top 32-bit element in Rs1 with the top 32-bit element in Rs2. It is actually an alias of the KMAR64 instruction. For the KMAXDA32 instruction, it multiplies the top 32-bit element in Rs1 with the bottom 32-bit element in Rs2 and then adds the result to the result of multiplying the bottom 32-bit element in Rs1 with the top 32-bit element in Rs2. The result is added to the content of 64-bit data in Rd. If the addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The 64-bit result is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```

res = Rd + (Rs1.W[1] * Rs2.w[1]) + (Rs1.W[0] * Rs2.W[0]); // KMADA32
res = Rd + (Rs1.W[1] * Rs2.W[0]) + (Rs1.W[0] * Rs2.W[1]); // KMAXDA32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;

```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

---

```
__STATIC_FORCEINLINE long __RV_KMDA32(unsigned long a, unsigned long b)
    KMDA32 (Signed Multiply Two Words and Add)
```

**Type:** DSP (RV64 Only)

**Syntax:**

```
KMDA32 Rd, Rs1, Rs2
KMXDA32 Rd, Rs1, Rs2
```

**Purpose:** Do two signed 32-bit multiplications from the 32-bit element of two registers; and then adds the two 64-bit results together. The addition result may be saturated.

- KMDA32: top\*top + bottom\*bottom
- KMXDA32: top\*bottom + bottom\*top

**Description:** For the KMDA32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2 and then adds the result to the result of multiplying the top 32-bit element of Rs1 with the top 32-bit element of Rs2. For the KMXDA32 instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2 and then adds the result to the result of multiplying the top 32-bit element of Rs1 with the bottom 32-bit element of Rs2. The addition result is checked for saturation. If saturation happens, the result is saturated to  $2^{63}-1$ . The final result is written to Rd. The 32-bit contents are treated as signed integers.

**Operations:**

```
if ((Rs1 != 0x8000000080000000) or (Rs2 != 0x8000000080000000)) {
    Rd = (Rs1.W[1] * Rs2.W[1]) + (Rs1.W[0] * Rs2.W[0]); // KMDA32
    Rd = (Rs1.W[1] * Rs2.W[0]) + (Rs1.W[0] * Rs2.W[1]); // KMXDA32
} else {
    Rd = 0x7fffffffffffffff;
    OV = 1;
}
```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_KMXDA32(unsigned long a, unsigned long b)
    KMXDA32 (Signed Crossed Multiply Two Words and Add)
```

**Type:** DSP (RV64 Only)

**Syntax:**

```
KMDA32 Rd, Rs1, Rs2
KMXDA32 Rd, Rs1, Rs2
```

**Purpose:** Do two signed 32-bit multiplications from the 32-bit element of two registers; and then adds the two 64-bit results together. The addition result may be saturated.

- KMDA32: top\*top + bottom\*bottom
- KMXDA32: top\*bottom + bottom\*top

**Description:** For the KMDA32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2 and then adds the result to the result of multiplying the top 32-bit element of Rs1

with the top 32-bit element of Rs2. For the `KMXDA32` instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2 and then adds the result to the result of multiplying the top 32-bit element of Rs1 with the bottom 32-bit element of Rs2. The addition result is checked for saturation. If saturation happens, the result is saturated to  $2^{63}-1$ . The final result is written to Rd. The 32-bit contents are treated as signed integers.

#### Operations:

```
if ((Rs1 != 0x8000000080000000) or (Rs2 != 0x8000000080000000)) {
    Rd = (Rs1.W[1] * Rs2.W[1]) + (Rs1.W[0] * Rs2.W[0]); // KMDA32
    Rd = (Rs1.W[1] * Rs2.W[0]) + (Rs1.W[0] * Rs2.W[1]); // KMXDA32
} else {
    Rd = 0x7fffffffffffffff;
    OV = 1;
}
```

**Return** value stored in long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_KMADS32(long t, unsigned long a, unsigned long b)
KMADS32 (Saturating Signed Multiply Two Words & Subtract & Add)
```

**Type:** DSP (RV64 Only)

#### Syntax:

```
KMADS32 Rd, Rs1, Rs2
KMADRS32 Rd, Rs1, Rs2
KMAXDS32 Rd, Rs1, Rs2
```

**Purpose:** Do two signed 32-bit multiplications from 32-bit elements in two registers; and then perform a subtraction operation between the two 64-bit results. Then add the subtraction result to 64-bit data in a third register. The addition result may be saturated.

- `KMADS32`:  $rd + (top * top - bottom * bottom)$
- `KMADRS32`:  $rd + (bottom * bottom - top * top)$
- `KMAXDS32`:  $rd + (top * bottom - bottom * top)$

**Description:** For the `KMADS32` instruction, it multiplies the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2 and then subtracts the result from the result of multiplying the top 32-bit element in Rs1 with the top 32-bit element in Rs2. For the `KMADRS32` instruction, it multiplies the top 32-bit element in Rs1 with the top 32-bit element in Rs2 and then subtracts the result from the result of multiplying the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2. For the `KMAXDS32` instruction, it multiplies the bottom 32-bit element in Rs1 with the top 32-bit element in Rs2 and then subtracts the result from the result of multiplying the top 32-bit element in Rs1 with the bottom 32-bit element in Rs2. The subtraction result is then added to the content of 64-bit data in Rd. If the addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to

1. The 64-bit result after saturation is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

#### Operations:

```

res = Rd + (Rs1.W[1] * Rs2.W[1]) - (Rs1.W[0] * Rs2.W[0]); // KMADS32
res = Rd + (Rs1.W[0] * Rs2.W[0]) - (Rs1.W[1] * Rs2.W[1]); // KMADRS32
res = Rd + (Rs1.W[1] * Rs2.W[0]) - (Rs1.W[0] * Rs2.W[1]); // KMAXDS32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;

```

**Return** value stored in long type

#### Parameters

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMADRS32(long t, unsigned long a, unsigned long b)**  
KMADRS32 (Saturating Signed Multiply Two Words & Reverse Subtract & Add)

**Type:** DSP (RV64 Only)

#### Syntax:

```

KMADS32 Rd, Rs1, Rs2
KMADRS32 Rd, Rs1, Rs2
KMAXDS32 Rd, Rs1, Rs2

```

**Purpose:** Do two signed 32-bit multiplications from 32-bit elements in two registers; and then perform a subtraction operation between the two 64-bit results. Then add the subtraction result to 64-bit data in a third register. The addition result may be saturated.

- KMADS32:  $rd + (top * top - bottom * bottom)$
- KMADRS32:  $rd + (bottom * bottom - top * top)$
- KMAXDS32:  $rd + (top * bottom - bottom * top)$

**Description:** For the KMADS32 instruction, it multiplies the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2 and then subtracts the result from the result of multiplying the top 32-bit element in Rs1 with the top 32-bit element in Rs2. For the KMADRS32 instruction, it multiplies the top 32-bit element in Rs1 with the top 32-bit element in Rs2 and then subtracts the result from the result of multiplying the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2. For the KMAXDS32 instruction, it multiplies the bottom 32-bit element in Rs1 with the top 32-bit element in Rs2 and then subtracts the result from the result of multiplying the top 32-bit element in Rs1 with the bottom 32-bit element in Rs2. The subtraction result is then added to the content of 64-bit data in Rd. If the addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to

1. The 64-bit result after saturation is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

#### Operations:

```

res = Rd + (Rs1.W[1] * Rs2.W[1]) - (Rs1.W[0] * Rs2.W[0]); // KMADS32
res = Rd + (Rs1.W[0] * Rs2.W[0]) - (Rs1.W[1] * Rs2.W[1]); // KMADRS32

```

(continues on next page)

(continued from previous page)

```

res = Rd + (Rs1.W[1] * Rs2.W[0]) - (Rs1.W[0] * Rs2.W[1]); // KMAXDS32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;

```

**Return** value stored in long type

#### Parameters

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMAXDS32(long t, unsigned long a, unsigned long b)**  
KMAXDS32 (Saturating Signed Crossed Multiply Two Words & Subtract & Add)

**Type:** DSP (RV64 Only)

#### Syntax:

```

KMADS32 Rd, Rs1, Rs2
KMADRS32 Rd, Rs1, Rs2
KMAXDS32 Rd, Rs1, Rs2

```

**Purpose:** Do two signed 32-bit multiplications from 32-bit elements in two registers; and then perform a subtraction operation between the two 64-bit results. Then add the subtraction result to 64-bit data in a third register. The addition result may be saturated.

- KMADS32:  $rd + (top * top - bottom * bottom)$
- KMADRS32:  $rd + (bottom * bottom - top * top)$
- KMAXDS32:  $rd + (top * bottom - bottom * top)$

**Description:** For the KMADS32 instruction, it multiplies the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2 and then subtracts the result from the result of multiplying the top 32-bit element in Rs1 with the top 32-bit element in Rs2. For the KMADRS32 instruction, it multiplies the top 32-bit element in Rs1 with the top 32-bit element in Rs2 and then subtracts the result from the result of multiplying the bottom 32-bit element in Rs1 with the bottom 32-bit element in Rs2. For the KMAXDS32 instruction, it multiplies the bottom 32-bit element in Rs1 with the top 32-bit element in Rs2 and then subtracts the result from the result of multiplying the top 32-bit element in Rs1 with the bottom 32-bit element in Rs2. The subtraction result is then added to the content of 64-bit data in Rd. If the addition result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to

1. The 64-bit result after saturation is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

#### Operations:

```

res = Rd + (Rs1.W[1] * Rs2.W[1]) - (Rs1.W[0] * Rs2.W[0]); // KMADS32
res = Rd + (Rs1.W[0] * Rs2.W[0]) - (Rs1.W[1] * Rs2.W[1]); // KMADRS32
res = Rd + (Rs1.W[1] * Rs2.W[0]) - (Rs1.W[0] * Rs2.W[1]); // KMAXDS32
if (res > (2^63)-1) {

```

(continues on next page)



(continued from previous page)

```

    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;

```

**Return** value stored in long type

#### Parameters

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMSDA32(long t, unsigned long a, unsigned long b)**  
KMSDA32 (Saturating Signed Multiply Two Words & Add & Subtract)

**Type:** DSP (RV64 Only)

#### Syntax:

```

KMSDA32 Rd, Rs1, Rs2
KMSXDA32 Rd, Rs1, Rs2

```

**Purpose:** Do two signed 32-bit multiplications from the 32-bit element of two registers; and then subtracts the two 64-bit results from a third register. The subtraction result may be saturated.

- KMSDA: rd - top\*top - bottom\*bottom
- KMSXDA: rd - top\*bottom - bottom\*top

**Description:** For the KMSDA32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2 and multiplies the top 32-bit element of Rs1 with the top 32-bit element of Rs2. For the KMSXDA32 instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2 and multiplies the top 32-bit element of Rs1 with the bottom 32-bit element of Rs2. The two 64-bit multiplication results are then subtracted from the content of Rd. If the subtraction result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The result after saturation is written to Rd. The 32-bit contents are treated as signed integers.

#### Operations:

```

res = Rd - (Rs1.W[1] * Rs2.W[1]) - (Rs1.W[0] * Rs2.W[0]); // KMSDA32
res = Rd - (Rs1.W[1] * Rs2.W[0]) - (Rs1.W[0] * Rs2.W[1]); // KMSXDA32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;

```

**Return** value stored in long type

#### Parameters

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_KMSXDA32(long t, unsigned long a, unsigned long b)**  
KMSXDA32 (Saturating Signed Crossed Multiply Two Words & Add & Subtract)

**Type:** DSP (RV64 Only)

**Syntax:**

```
KMSDA32 Rd, Rs1, Rs2
KMSXDA32 Rd, Rs1, Rs2
```

**Purpose:** Do two signed 32-bit multiplications from the 32-bit element of two registers; and then subtracts the two 64-bit results from a third register. The subtraction result may be saturated.

- KMSDA: rd - top\*top - bottom\*bottom
- KMSXDA: rd - top\*bottom - bottom\*top

**Description:** For the KMSDA32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2 and multiplies the top 32-bit element of Rs1 with the top 32-bit element of Rs2. For the KMSXDA32 instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2 and multiplies the top 32-bit element of Rs1 with the bottom 32-bit element of Rs2. The two 64-bit multiplication results are then subtracted from the content of Rd. If the subtraction result is beyond the Q63 number range ( $-2^{63} \leq Q63 \leq 2^{63}-1$ ), it is saturated to the range and the OV bit is set to 1. The result after saturation is written to Rd. The 32-bit contents are treated as signed integers.

**Operations:**

```
res = Rd - (Rs1.W[1] * Rs2.W[1]) - (Rs1.W[0] * Rs2.W[0]); // KMSDA32
res = Rd - (Rs1.W[1] * Rs2.W[0]) - (Rs1.W[0] * Rs2.W[1]); // KMSXDA32
if (res > (2^63)-1) {
    res = (2^63)-1;
    OV = 1;
} else if (res < -2^63) {
    res = -2^63;
    OV = 1;
}
Rd = res;
```

**Return** value stored in long type

**Parameters**

- [in] t: long type of value stored in t
- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMDS32(unsigned long a, unsigned long b)**  
SMDS32 (Signed Multiply Two Words and Subtract)

**Type:** DSP (RV64 Only)

**Syntax:**

```

SMDS32 Rd, Rs1, Rs2
SMDRS32 Rd, Rs1, Rs2
SMXDS32 Rd, Rs1, Rs2

```

**Purpose:** Do two signed 32-bit multiplications from the 1 32-bit element of two registers; and then perform a subtraction operation between the two 64-bit results.

- SMDS32: top\*top - bottom\*bottom
- SMDRS32: bottom\*bottom - top\*top
- SMXDS32: top\*bottom - bottom\*top

**Description:** For the SMDS32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2 and then subtracts the result from the result of multiplying the top 32-bit element of Rs1 with the top 32-bit element of Rs2. For the SMDRS32 instruction, it multiplies the top 32-bit element of Rs1 with the top 32-bit element of Rs2 and then subtracts the result from the result of multiplying the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2. For the SMXDS32 instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2 and then subtracts the result from the result of multiplying the top 32-bit element of Rs1 with the bottom 32-bit element of Rs2. The subtraction result is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

**Operations:**

```

Rt = (Rs1.W[1] * Rs2.W[1]) - (Rs1.W[0] * Rs2.W[0]); // SMDS32
Rt = (Rs1.W[0] * Rs2.W[0]) - (Rs1.W[1] * Rs2.W[1]); // SMDRS32
Rt = (Rs1.W[1] * Rs2.W[0]) - (Rs1.W[0] * Rs2.W[1]); // SMXDS32

```

**Return** value stored in long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE long \_\_RV\_SMDRS32(unsigned long a, unsigned long b)**  
SMDRS32 (Signed Multiply Two Words and Reverse Subtract)

**Type:** DSP (RV64 Only)

**Syntax:**

```

SMDS32 Rd, Rs1, Rs2
SMDRS32 Rd, Rs1, Rs2
SMXDS32 Rd, Rs1, Rs2

```

**Purpose:** Do two signed 32-bit multiplications from the 1 32-bit element of two registers; and then perform a subtraction operation between the two 64-bit results.

- SMDS32: top\*top - bottom\*bottom
- SMDRS32: bottom\*bottom - top\*top
- SMXDS32: top\*bottom - bottom\*top

**Description:** For the SMDS32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2 and then subtracts the result from the result of multiplying the top 32-bit element of Rs1 with the top 32-bit element of Rs2. For the SMDRS32 instruction, it multiplies the top 32-bit element of Rs1 with the top 32-bit element of Rs2 and then subtracts the result from the result of multiplying the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2. For the SMXDS32 instruction, it

multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2 and then subtracts the result from the result of multiplying the top 32-bit element of Rs1 with the bottom 32-bit element of Rs2. The subtraction result is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

#### Operations:

```
Rt = (Rs1.W[1] * Rs2.W[1]) - (Rs1.W[0] * Rs2.W[0]); // SMDS32
Rt = (Rs1.W[0] * Rs2.W[0]) - (Rs1.W[1] * Rs2.W[1]); // SMDRS32
Rt = (Rs1.W[1] * Rs2.W[0]) - (Rs1.W[0] * Rs2.W[1]); // SMXDS32
```

**Return** value stored in long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

```
__STATIC_FORCEINLINE long __RV_SMXDS32(unsigned long a, unsigned long b)  
SMXDS32 (Signed Crossed Multiply Two Words and Subtract)
```

**Type:** DSP (RV64 Only)

#### Syntax:

```
SMDS32 Rd, Rs1, Rs2  
SMDRS32 Rd, Rs1, Rs2  
SMXDS32 Rd, Rs1, Rs2
```

**Purpose:** Do two signed 32-bit multiplications from the 1 32-bit element of two registers; and then perform a subtraction operation between the two 64-bit results.

- SMDS32: top\*top - bottom\*bottom
- SMDRS32: bottom\*bottom - top\*top
- SMXDS32: top\*bottom - bottom\*top

**Description:** For the SMDS32 instruction, it multiplies the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2 and then subtracts the result from the result of multiplying the top 32-bit element of Rs1 with the top 32-bit element of Rs2. For the SMDRS32 instruction, it multiplies the top 32-bit element of Rs1 with the top 32-bit element of Rs2 and then subtracts the result from the result of multiplying the bottom 32-bit element of Rs1 with the bottom 32-bit element of Rs2. For the SMXDS32 instruction, it multiplies the bottom 32-bit element of Rs1 with the top 32-bit element of Rs2 and then subtracts the result from the result of multiplying the top 32-bit element of Rs1 with the bottom 32-bit element of Rs2. The subtraction result is written to Rd. The 32-bit contents of Rs1 and Rs2 are treated as signed integers.

#### Operations:

```
Rt = (Rs1.W[1] * Rs2.W[1]) - (Rs1.W[0] * Rs2.W[0]); // SMDS32
Rt = (Rs1.W[0] * Rs2.W[0]) - (Rs1.W[1] * Rs2.W[1]); // SMDRS32
Rt = (Rs1.W[1] * Rs2.W[0]) - (Rs1.W[0] * Rs2.W[1]); // SMXDS32
```

**Return** value stored in long type

#### Parameters

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

## (RV64 Only) Non-SIMD 32-bit Shift Instructions

```
__STATIC_FORCEINLINE long __RV_SRAIW_U(int a, unsigned int b)
```

group **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_NON\_SIMD\_32B\_SHIFT**

(RV64 Only) Non-SIMD 32-bit Shift Instructions

there are 1 (RV64 Only) Non-SIMD 32-bit Shift Instructions

### Functions

```
__STATIC_FORCEINLINE long __RV_SRAIW_U(int a, unsigned int b)
```

SRAIW.u (Rounding Shift Right Arithmetic Immediate Word)

**Type:** DSP (RV64 only)

**Syntax:**

```
SRAIW.u Rd, Rs1, imm5u
```

**Purpose:** Perform a 32-bit arithmetic right shift operation with rounding. The shift amount is an immediate value.

**Description:** This instruction right-shifts the lower 32-bit content of Rs1 arithmetically. The shifted out bits are filled with the sign-bit Rs1(31) and the shift amount is specified by the imm5u constant. For the rounding operation, a value of 1 is added to the most significant discarded bit of the data to calculate the final result. And the result is sign-extended and written to Rd.

**Operations:**

```
sa = imm5u;
if (sa != 0) {
    res[31:-1] = SE32(Rs1[31:(sa-1)]) + 1;
    Rd = SE32(res[31:0]);
} else {
    Rd = SE32(Rs1.W[0]);
}
```

**Return** value stored in long type

**Parameters**

- [in] a: int type of value stored in a
- [in] b: unsigned int type of value stored in b

## 32-bit Packing Instructions

```
__STATIC_FORCEINLINE unsigned long __RV_PKBB32(unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_PKBT32(unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_PKTT32(unsigned long a, unsigned long b)
```

```
__STATIC_FORCEINLINE unsigned long __RV_PKTB32(unsigned long a, unsigned long b)
```

group **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_32B\_PACK**

32-bit Packing Instructions

There are four 32-bit packing instructions here

## Functions

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PKBB32(unsigned long a, unsigned long b)**  
 PKBB32 (Pack Two 32-bit Data from Both Bottom Half)

**Type:** DSP (RV64 Only)

**Syntax:**

```
PKBB32 Rd, Rs1, Rs2
PKBT32 Rd, Rs1, Rs2
PKTT32 Rd, Rs1, Rs2
PKTB32 Rd, Rs1, Rs2
```

**Purpose:** Pack 32-bit data from 64-bit chunks in two registers.

- PKBB32: bottom.bottom
- PKBT32: bottom.top
- PKTT32: top.top
- PKTB32: top.bottom

**Description:** (PKBB32) moves Rs1.W[0] to Rd.W[1] and moves Rs2.W[0] to Rd.W[0]. (PKBT32) moves Rs1.W[0] to Rd.W[1] and moves Rs2.W[1] to Rd.W[0]. (PKTT32) moves Rs1.W[1] to Rd.W[1] and moves Rs2.W[1] to Rd.W[0]. (PKTB32) moves Rs1.W[1] to Rd.W[1] and moves Rs2.W[0] to Rd.W[0].

**Operations:**

```
Rd = CONCAT(Rs1.W[_*0*_*], Rs2.W[_*0*_*]); // PKBB32
Rd = CONCAT(Rs1.W[_*0*_*], Rs2.W[_*1*_*]); // PKBT32
Rd = CONCAT(Rs1.W[_*1*_*], Rs2.W[_*1*_*]); // PKTT32
Rd = CONCAT(Rs1.W[_*1*_*], Rs2.W[_*0*_*]); // PKTB32
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PKBT32(unsigned long a, unsigned long b)**  
 PKBT32 (Pack Two 32-bit Data from Bottom and Top Half)

**Type:** DSP (RV64 Only)

**Syntax:**

```
PKBB32 Rd, Rs1, Rs2
PKBT32 Rd, Rs1, Rs2
PKTT32 Rd, Rs1, Rs2
PKTB32 Rd, Rs1, Rs2
```

**Purpose:** Pack 32-bit data from 64-bit chunks in two registers.

- PKBB32: bottom.bottom
- PKBT32: bottom.top
- PKTT32: top.top
- PKTB32: top.bottom

**Description:** (PKBB32) moves Rs1.W[0] to Rd.W[1] and moves Rs2.W[0] to Rd.W[0]. (PKBT32) moves Rs1.W[0] to Rd.W[1] and moves Rs2.W[1] to Rd.W[0]. (PKTT32) moves Rs1.W[1] to Rd.W[1] and moves Rs2.W[1] to Rd.W[0]. (PKTB32) moves Rs1.W[1] to Rd.W[1] and moves Rs2.W[0] to Rd.W[0].

**Operations:**

```
Rd = CONCAT(Rs1.W[_*0*_*], Rs2.W[_*0*_*]); // PKBB32
Rd = CONCAT(Rs1.W[_*0*_*], Rs2.W[_*1*_*]); // PKBT32
Rd = CONCAT(Rs1.W[_*1*_*], Rs2.W[_*1*_*]); // PKTT32
Rd = CONCAT(Rs1.W[_*1*_*], Rs2.W[_*0*_*]); // PKTB32
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PKTT32(unsigned long a, unsigned long b)**  
PKTT32 (Pack Two 32-bit Data from Both Top Half)

**Type:** DSP (RV64 Only)

**Syntax:**

```
PKBB32 Rd, Rs1, Rs2
PKBT32 Rd, Rs1, Rs2
PKTT32 Rd, Rs1, Rs2
PKTB32 Rd, Rs1, Rs2
```

**Purpose:** Pack 32-bit data from 64-bit chunks in two registers.

- PKBB32: bottom.bottom
- PKBT32: bottom.top
- PKTT32: top.top
- PKTB32: top.bottom

**Description:** (PKBB32) moves Rs1.W[0] to Rd.W[1] and moves Rs2.W[0] to Rd.W[0]. (PKBT32) moves Rs1.W[0] to Rd.W[1] and moves Rs2.W[1] to Rd.W[0]. (PKTT32) moves Rs1.W[1] to Rd.W[1] and moves Rs2.W[1] to Rd.W[0]. (PKTB32) moves Rs1.W[1] to Rd.W[1] and moves Rs2.W[0] to Rd.W[0].

**Operations:**

```
Rd = CONCAT(Rs1.W[_*0*_*], Rs2.W[_*0*_*]); // PKBB32
Rd = CONCAT(Rs1.W[_*0*_*], Rs2.W[_*1*_*]); // PKBT32
Rd = CONCAT(Rs1.W[_*1*_*], Rs2.W[_*1*_*]); // PKTT32
Rd = CONCAT(Rs1.W[_*1*_*], Rs2.W[_*0*_*]); // PKTB32
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_PKTB32(unsigned long a, unsigned long b)**  
PKTB32 (Pack Two 32-bit Data from Top and Bottom Half)

**Type:** DSP (RV64 Only)

**Syntax:**

```
PKBB32 Rd, Rs1, Rs2
PKBT32 Rd, Rs1, Rs2
PKTT32 Rd, Rs1, Rs2
PKTB32 Rd, Rs1, Rs2
```

**Purpose:** Pack 32-bit data from 64-bit chunks in two registers.

- PKBB32: bottom.bottom
- PKBT32: bottom.top
- PKTT32: top.top
- PKTB32: top.bottom

**Description:** (PKBB32) moves Rs1.W[0] to Rd.W[1] and moves Rs2.W[0] to Rd.W[0]. (PKBT32) moves Rs1.W[0] to Rd.W[1] and moves Rs2.W[1] to Rd.W[0]. (PKTT32) moves Rs1.W[1] to Rd.W[1] and moves Rs2.W[1] to Rd.W[0]. (PKTB32) moves Rs1.W[1] to Rd.W[1] and moves Rs2.W[0] to Rd.W[0].

**Operations:**

```
Rd = CONCAT(Rs1.W[_*0*_*], Rs2.W[_*0*_*]); // PKBB32
Rd = CONCAT(Rs1.W[_*0*_*], Rs2.W[_*1*_*]); // PKBT32
Rd = CONCAT(Rs1.W[_*1*_*], Rs2.W[_*1*_*]); // PKTT32
Rd = CONCAT(Rs1.W[_*1*_*], Rs2.W[_*0*_*]); // PKTB32
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a
- [in] b: unsigned long type of value stored in b

*group* **NMSIS\_Core\_DSP\_Intrinsic\_RV64\_ONLY**  
RV64 Only Instructions.

**Nuclei Customized DSP Instructions**

```
__STATIC_FORCEINLINE unsigned long long __RV_DKHM8(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_DKHM16(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_DKABS8(unsigned long long a)
__STATIC_FORCEINLINE unsigned long long __RV_DKABS16(unsigned long long a)
__STATIC_FORCEINLINE unsigned long long __RV_DKSLRA8(unsigned long long a, int b)
__STATIC_FORCEINLINE unsigned long long __RV_DKSLRA16(unsigned long long a, int b)
__STATIC_FORCEINLINE unsigned long long __RV_DKADD8(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_DKADD16(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_DKSUB8(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long long __RV_DKSUB16(unsigned long long a, unsigned long long b)
__STATIC_FORCEINLINE unsigned long __RV_EXPD80(unsigned long a)
__STATIC_FORCEINLINE unsigned long __RV_EXPD81(unsigned long a)
```



```
__STATIC_FORCEINLINE unsigned long __RV_EXP82(unsigned long a)
```

```
__STATIC_FORCEINLINE unsigned long __RV_EXP83(unsigned long a)
```

group **NMSIS\_Core\_DSP\_Intrinsic\_NUCLEI\_CUSTOM**

(RV32 only)Nuclei Customized DSP Instructions

This is Nuclei customized DSP instructions only for RV32

## Functions

```
__STATIC_FORCEINLINE unsigned long long __RV_DKHM8(unsigned long long a, unsigned long long b)
DKHM8 (64-bit SIMD Signed Saturating Q7 Multiply)
```

**Type:** SIMD

**Syntax:**

```
DKHM8 Rd, Rs1, Rs2
# Rd, Rs1, Rs2 are all even/odd pair of registers
```

**Purpose:** Do Q7xQ7 element multiplications simultaneously. The Q14 results are then reduced to Q7 numbers again.

**Description:** For the DKHM8 instruction, multiply the top 8-bit Q7 content of 16-bit chunks in Rs1 with the top 8-bit Q7 content of 16-bit chunks in Rs2. At the same time, multiply the bottom 8-bit Q7 content of 16-bit chunks in Rs1 with the bottom 8-bit Q7 content of 16-bit chunks in Rs2.

The Q14 results are then right-shifted 7-bits and saturated into Q7 values. The Q7 results are then written into Rd. When both the two Q7 inputs of a multiplication are 0x80, saturation will happen. The result will be saturated to 0x7F and the overflow flag OV will be set.

**Operations:**

```
op1t = Rs1.B[x+1]; op2t = Rs2.B[x+1]; // top
op1b = Rs1.B[x]; op2b = Rs2.B[x]; // bottom
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    if (0x80 != aop | 0x80 != bop) {
        res = (aop s* bop) >> 7;
    } else {
        res = 0x7F;
        OV = 1;
    }
}
Rd.H[x/2] = concat(rest, resb);
for RV32, x=0,2,4,6
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

```
__STATIC_FORCEINLINE unsigned long long __RV_DKHM16(unsigned long long a, unsigned long long b)
DKHM16 (64-bit SIMD Signed Saturating Q15 Multiply)
```

**Type:** SIMD

**Syntax:**

```
DKHM16 Rd, Rs1, Rs2
# Rd, Rs1, Rs2 are all even/odd pair of registers
```

**Purpose:** Do Q15xQ15 element multiplications simultaneously. The Q30 results are then reduced to Q15 numbers again.

**Description:** For the DKHM16 instruction, multiply the top 16-bit Q15 content of 32-bit chunks in Rs1 with the top 16-bit Q15 content of 32-bit chunks in Rs2. At the same time, multiply the bottom 16-bit Q15 content of 32-bit chunks in Rs1 with the bottom 16-bit Q15 content of 32-bit chunks in Rs2.

The Q30 results are then right-shifted 15-bits and saturated into Q15 values. The Q15 results are then written into Rd. When both the two Q15 inputs of a multiplication are 0x8000, saturation will happen. The result will be saturated to 0x7FFF and the overflow flag OV will be set.

#### Operations:

```
op1t = Rs1.H[x+1]; op2t = Rs2.H[x+1]; // top
op1b = Rs1.H[x]; op2b = Rs2.H[x]; // bottom
for ((aop,bop,res) in [(op1t,op2t,rest), (op1b,op2b,resb)]) {
    if (0x8000 != aop & 0x8000 != bop) {
        res = (aop s* bop) >> 15;
    } else {
        res = 0x7FFF;
        OV = 1;
    }
}
Rd.W[x/2] = concat(rest, resb);
for RV32: x=0, 2
```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_DKABS8(unsigned long long a)**  
DKABS8 (64-bit SIMD 8-bit Saturating Absolute)

**Type:** SIMD

#### Syntax:

```
DKABS8 Rd, Rs1
# Rd, Rs1 are all even/odd pair of registers
```

**Purpose:** Get the absolute value of 8-bit signed integer elements simultaneously.

**Description:** This instruction calculates the absolute value of 8-bit signed integer elements stored in Rs1 and writes the element results to Rd. If the input number is 0x80, this instruction generates 0x7f as the output and sets the OV bit to 1.

#### Operations:

```
src = Rs1.B[x];
if (src == 0x80) {
    src = 0x7f;
    OV = 1;
} else if (src[7] == 1)
```

(continues on next page)

(continued from previous page)

```

    src = -src;
}
Rd.B[x] = src;
for RV32: x=7...0,

```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_DKABS16(unsigned long long a)**  
 DKABS16 (64-bit SIMD 16-bit Saturating Absolute)

**Type:** SIMD

**Syntax:**

```

DKABS16 Rd, Rs1
# Rd, Rs1 are all even/odd pair of registers

```

**Purpose:** Get the absolute value of 16-bit signed integer elements simultaneously.

**Description:** This instruction calculates the absolute value of 16-bit signed integer elements stored in Rs1 and writes the element results to Rd. If the input number is 0x8000, this instruction generates 0x7fff as the output and sets the OV bit to 1.

**Operations:**

```

src = Rs1.H[x];
if (src == 0x8000) {
    src = 0x7fff;
    OV = 1;
} else if (src[15] == 1)
    src = -src;
}
Rd.H[x] = src;
for RV32: x=3...0,

```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_DKSLRA8(unsigned long long a, int b)**  
 DKSLRA8 (64-bit SIMD 8-bit Shift Left Logical with Saturation or Shift Right Arithmetic)

**Type:** SIMD

**Syntax:**

```

DKSLRA8 Rd, Rs1, Rs2
# Rd, Rs1 are all even/odd pair of registers

```

**Purpose:** Do 8-bit elements logical left (positive) or arithmetic right (negative) shift operation with Q7 saturation for the left shift.

**Description:** The 8-bit data elements of Rs1 are left-shifted logically or right-shifted arithmetically based on the value of Rs2[3:0]. Rs2[3:0] is in the signed range of  $[-2^3, 2^3-1]$ . A positive Rs2[3:0] means

logical left shift and a negative Rs2[3:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[3:0]. However, the behavior of  $\text{Rs2}[3:0] == -2^3$  (0x8) is defined to be equivalent to the behavior of  $\text{Rs2}[3:0] == -(2^3-1)$  (0x9). The left-shifted results are saturated to the 8-bit signed integer range of  $[-2^7, 2^7-1]$ . If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:4] will not affect this instruction.

#### Operations:

```
if (Rs2[3:0] < 0) {
    sa = -Rs2[3:0];
    sa = (sa == 8)? 7 : sa;
    Rd.B[x] = SE8(Rs1.B[x][7:sa]);
} else {
    sa = Rs2[2:0];
    res[(7+sa):0] = Rs1.B[x] <<(logic) sa;
    if (res > (2^7)-1) {
        res[7:0] = 0x7f; OV = 1;
    } else if (res < -2^7) {
        res[7:0] = 0x80; OV = 1;
    }
    Rd.B[x] = res[7:0];
}
for RV32: x=7...0,
```

**Return** value stored in unsigned long long type

#### Parameters

- [in] a: unsigned long long type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_DKSLRA16(unsigned long long a, int b)**  
DKSLRA16 (64-bit SIMD 16-bit Shift Left Logical with Saturation or Shift Right Arithmetic)

**Type:** SIMD

#### Syntax:

```
DKSLRA16 Rd, Rs1, Rs2
# Rd, Rs1 are all even/odd pair of registers
```

**Purpose:** Do 16-bit elements logical left (positive) or arithmetic right (negative) shift operation with Q15 saturation for the left shift.

**Description:** The 16-bit data elements of Rs1 are left-shifted logically or right-shifted arithmetically based on the value of Rs2[4:0]. Rs2[4:0] is in the signed range of  $[-2^4, 2^4-1]$ . A positive Rs2[4:0] means logical left shift and a negative Rs2[4:0] means arithmetic right shift. The shift amount is the absolute value of Rs2[4:0]. However, the behavior of  $\text{Rs2}[4:0] == -2^4$  (0x10) is defined to be equivalent to the behavior of  $\text{Rs2}[4:0] == -(2^4-1)$  (0x11). The left-shifted results are saturated to the 16-bit signed integer range of  $[-2^{15}, 2^{15}-1]$ . After the shift, saturation, or rounding, the final results are written to Rd. If any saturation happens, this instruction sets the OV flag. The value of Rs2[31:5] will not affect this instruction.

#### Operations:

```
if (Rs2[4:0] < 0) {
    sa = -Rs2[4:0];
    sa = (sa == 16)? 15 : sa;
```

(continues on next page)

(continued from previous page)

```

    Rd.H[x] = SE16(Rs1.H[x][15:sa]);
} else {
    sa = Rs2[3:0];
    res[(15+sa):0] = Rs1.H[x] <<(logic) sa;
    if (res > (2^15)-1) {
        res[15:0] = 0x7fff; OV = 1;
    } else if (res < -2^15) {
        res[15:0] = 0x8000; OV = 1;
    }
    d.H[x] = res[15:0];
}
for RV32: x=3...0,

```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: int type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_DKADD8(unsigned long long a, unsigned long long b)**  
 DKADD8 (64-bit SIMD 8-bit Signed Saturating Addition)

**Type:** SIMD

**Syntax:**

```

DKADD8 Rd, Rs1, Rs2
# Rd, Rs1, Rs2 are all even/odd pair of registers

```

**Purpose:** Do 8-bit signed integer element saturating additions simultaneously.

**Description:** This instruction adds the 8-bit signed integer elements in Rs1 with the 8-bit signed integer elements in Rs2. If any of the results are beyond the Q7 number range ( $-2^7 \leq Q7 \leq 2^7-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```

res[x] = Rs1.B[x] + Rs2.B[x];
if (res[x] > 127) {
    res[x] = 127;
    OV = 1;
} else if (res[x] < -128) {
    res[x] = -128;
    OV = 1;
}
Rd.B[x] = res[x];
for RV32: x=7...0,

```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_DKADD16(unsigned long long a, unsigned long long b)**  
 DKADD16 (64-bit SIMD 16-bit Signed Saturating Addition)

**Type:** SIMD**Syntax:**

```
DKADD16 Rd, Rs1, Rs2
# Rd, Rs1, Rs2 are all even/odd pair of registers
```

**Purpose:** Do 16-bit signed integer element saturating additions simultaneously.**Description:** This instruction adds the 16-bit signed integer elements in Rs1 with the 16-bit signed integer elements in Rs2. If any of the results are beyond the Q15 number range ( $-2^{15} \leq Q15 \leq 2^{15}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.**Operations:**

```
res[x] = Rs1.H[x] + Rs2.H[x];
if (res[x] > 32767) {
    res[x] = 32767;
    OV = 1;
} else if (res[x] < -32768) {
    res[x] = -32768;
    OV = 1;
}
Rd.H[x] = res[x];
for RV32: x=3...0,
```

**Return** value stored in unsigned long long type**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_DKSUB8(unsigned long long a, unsigned long long b)**  
 DKSUB8 (64-bit SIMD 8-bit Signed Saturating Subtraction)

**Type:** SIMD**Syntax:**

```
DKSUB8 Rd, Rs1, Rs2
# Rd, Rs1, Rs2 are all even/odd pair of registers
```

**Purpose:** Do 8-bit signed elements saturating subtractions simultaneously.**Description:** This instruction subtracts the 8-bit signed integer elements in Rs2 from the 8-bit signed integer elements in Rs1. If any of the results are beyond the Q7 number range ( $-2^7 \leq Q7 \leq 2^7-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.**Operations:**

```
res[x] = Rs1.B[x] - Rs2.B[x];
if (res[x] > (2^7)-1) {
    res[x] = (2^7)-1;
    OV = 1;
} else if (res[x] < -2^7) {
    res[x] = -2^7;
    OV = 1;
}
```

(continues on next page)

(continued from previous page)

```
Rd.B[x] = res[x];
for RV32: x=7...0,
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long long \_\_RV\_DKSUB16(unsigned long long a, unsigned long long b)**  
DKSUB16 (64-bit SIMD 16-bit Signed Saturating Subtraction)

**Type:** SIMD

**Syntax:**

```
DKSUB16 Rd, Rs1, Rs2
# Rd, Rs1, Rs2 are all even/odd pair of registers
```

**Purpose:** Do 16-bit signed integer elements saturating subtractions simultaneously.

**Description:** This instruction subtracts the 16-bit signed integer elements in Rs2 from the 16-bit signed integer elements in Rs1. If any of the results are beyond the Q15 number range ( $-2^{15} \leq Q15 \leq 2^{15}-1$ ), they are saturated to the range and the OV bit is set to 1. The saturated results are written to Rd.

**Operations:**

```
res[x] = Rs1.H[x] - Rs2.H[x];
if (res[x] > (2^15)-1) {
    res[x] = (2^15)-1;
    OV = 1;
} else if (res[x] < -2^15) {
    res[x] = -2^15;
    OV = 1;
}
Rd.H[x] = res[x];
for RV32: x=3...0,
```

**Return** value stored in unsigned long long type

**Parameters**

- [in] a: unsigned long long type of value stored in a
- [in] b: unsigned long long type of value stored in b

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_EXPD80(unsigned long a)**  
EXPD80 (Expand and Copy Byte 0 to 32bit)

**Type:** DSP

**Syntax:**

```
EXPD80 Rd, Rs1
```

**Purpose:** Copy 8-bit data from 32-bit chunks into 4 bytes in a register.

**Description:** Moves Rs1.B[0][7:0] to Rd.[0][7:0], Rd.[1][7:0], Rd.[2][7:0], Rd.[3][7:0]

**Operations:**

```
Rd.W[x][31:0] = CONCAT(Rs1.B[0][7:0], Rs1.B[0][7:0], Rs1.B[0][7:0], Rs1.
↪B[0][7:0]);
for RV32: x=0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_EXP81(unsigned long a)**  
EXP81 (Expand and Copy Byte 1 to 32bit)

**Type:** DSP

**Syntax:**

```
EXP81 Rd, Rs1
```

**Purpose:** Copy 8-bit data from 32-bit chunks into 4 bytes in a register.

**Description:** Moves Rs1.B[1][7:0] to Rd.[0][7:0], Rd.[1][7:0], Rd.[2][7:0], Rd.[3][7:0]

**Operations:**

```
Rd.W[x][31:0] = CONCAT(Rs1.B[1][7:0], Rs1.B[1][7:0], Rs1.B[1][7:0], Rs1.
↪B[1][7:0]);
for RV32: x=0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

**\_\_STATIC\_FORCEINLINE unsigned long \_\_RV\_EXP82(unsigned long a)**  
EXP82 (Expand and Copy Byte 2 to 32bit)

**Type:** DSP

**Syntax:**

```
EXP82 Rd, Rs1
```

**Purpose:** Copy 8-bit data from 32-bit chunks into 4 bytes in a register.

**Description:** Moves Rs1.B[2][7:0] to Rd.[0][7:0], Rd.[1][7:0], Rd.[2][7:0], Rd.[3][7:0]

**Operations:**

```
Rd.W[x][31:0] = CONCAT(Rs1.B[2][7:0], Rs1.B[2][7:0], Rs1.B[2][7:0], Rs1.
↪B[2][7:0]);
for RV32: x=0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a



```
__STATIC_FORCEINLINE unsigned long __RV_EXPDP83(unsigned long a)
    EXPDP83 (Expand and Copy Byte 3 to 32bit)
```

**Type:** DSP

**Syntax:**

```
EXPDP83 Rd, Rs1
```

**Purpose:** Copy 8-bit data from 32-bit chunks into 4 bytes in a register.

**Description:** Moves Rs1.B[3][7:0] to Rd.[0][7:0], Rd.[1][7:0], Rd.[2][7:0], Rd.[3][7:0]

**Operations:**

```
Rd.W[x][31:0] = CONCAT(Rs1.B[3][7:0], Rs1.B[3][7:0], Rs1.B[3][7:0], Rs1.
↪B[3][7:0]);
for RV32: x=0
```

**Return** value stored in unsigned long type

**Parameters**

- [in] a: unsigned long type of value stored in a

#### group NMSIS\_Core\_DSP\_Intrinsic

Functions that generate RISC-V DSP SIMD instructions.

The following functions generate specified RISC-V SIMD instructions that cannot be directly accessed by compiler.

#### • DSP ISA Extension Instruction Summary

##### – Shorthand Definitions

- \* r.H == r.H1: r[31:16], r.L == r.H0: r[15:0]
- \* r.B3: r[31:24], r.B2: r[23:16], r.B1: r[15:8], r.B0: r[7:0]
- \* r.B[x]: r[(x\*8+7):(x\*8+0)]
- \* r.H[x]: r[(x\*16+7):(x\*16+0)]
- \* r.W[x]: r[(x\*32+31):(x\*32+0)]
- \* r[xU]: the upper 32-bit of a 64-bit number; xU represents the GPR number that contains this upper part 32-bit value.
- \* r[xL]: the lower 32-bit of a 64-bit number; xL represents the GPR number that contains this lower part 32-bit value.
- \* r[xU].r[xL]: a 64-bit number that is formed from a pair of GPRs.
- \* s>>: signed arithmetic right shift:
- \* u>>: unsigned logical right shift
- \* SAT.Qn(): Saturate to the range of  $[-2^n, 2^n-1]$ , if saturation happens, set PSW.OV.
- \* SAT.Um(): Saturate to the range of  $[0, 2^m-1]$ , if saturation happens, set PSW.OV.
- \* RUND(): Indicate rounding, i.e., add 1 to the most significant discarded bit for right shift or MSW-type multiplication instructions.
- \* Sign or Zero Extending functions:
  - SEm(data): Sign-Extend data to m-bit.:

- ZEm(data): Zero-Extend data to m-bit.
- \* ABS(x): Calculate the absolute value of x.
- \* CONCAT(x,y): Concatenate x and y to form a value.
- \* u<: Unsinged less than comparison.
- \* u<=: Unsinged less than & equal comparison.
- \* u>: Unsinged greater than comparison.
- \* s\*: Signed multiplication.
- \* u\*: Unsigned multiplication.

## 2.5.8 Peripheral Access

\_\_\_I volatile const

\_\_\_O volatile

\_\_\_IO volatile

\_\_\_IM volatile const

\_\_\_OM volatile

\_\_\_IOM volatile

\_\_VAL2FLD (field, value) (((uint32\_t)(value) << field ## \_Pos) & field ## \_Msk)

\_\_FLD2VAL (field, value) (((uint32\_t)(value) & field ## \_Msk) >> field ## \_Pos)

*group* **NMSIS\_Core\_PeriphAccess**

Naming conventions and optional features for accessing peripherals.

The section below describes the naming conventions, requirements, and optional features for accessing device specific peripherals. Most of the rules also apply to the core peripherals.

The **Device Header File <device.h>** contains typically these definition and also includes the core specific header files.

### Defines

\_\_\_I volatile const

Defines 'read only' permissions.

\_\_\_O volatile

Defines 'write only' permissions.

\_\_\_IO volatile

Defines 'read / write' permissions.

\_\_\_IM volatile const

Defines 'read only' structure member permissions.

\_\_\_OM volatile

Defines 'write only' structure member permissions.

\_\_\_IOM volatile

Defines 'read/write' structure member permissions.

**\_VAL2FLD** (field, value) (((uint32\_t)(value) << field ## \_Pos) & field ## \_Msk)

Mask and shift a bit field value for use in a register bit range.

The macro **\_VAL2FLD** uses the #define's **\_Pos** and **\_Msk** of the related bit field to shift bit-field values for assigning to a register.

**Example:**

```
ECLIC->CFG = _VAL2FLD(CLIC_CLICCFG_NLBIT, 3);
```

**Return** Masked and shifted value.

**Parameters**

- [in] **field**: Name of the register bit field.
- [in] **value**: Value of the bit field. This parameter is interpreted as an uint32\_t type.

**\_FLD2VAL** (field, value) (((uint32\_t)(value) & field ## \_Msk) >> field ## \_Pos)

Mask and shift a register value to extract a bit field value.

The macro **\_FLD2VAL** uses the #define's **\_Pos** and **\_Msk** of the related bit field to extract the value of a bit field from a register.

**Example:**

```
nlbits = _FLD2VAL(CLIC_CLICCFG_NLBIT, ECLIC->CFG);
```

**Return** Masked and shifted bit field value.

**Parameters**

- [in] **field**: Name of the register bit field.
- [in] **value**: Value of register. This parameter is interpreted as an uint32\_t type.

## 2.5.9 SysTick Timer(SysTimer)

Click [Nuclei Timer Unit](#)<sup>17</sup> to learn about Core Timer Unit in Nuclei ISA Spec.

### SysTimer API

```
__STATIC_FORCEINLINE void SysTimer_SetLoadValue(uint64_t value)
__STATIC_FORCEINLINE uint64_t SysTimer_GetLoadValue(void)
__STATIC_FORCEINLINE void SysTimer_SetCompareValue(uint64_t value)
__STATIC_FORCEINLINE uint64_t SysTimer_GetCompareValue(void)
__STATIC_FORCEINLINE void SysTimer_Start(void)
__STATIC_FORCEINLINE void SysTimer_Stop(void)
__STATIC_FORCEINLINE void SysTimer_SetControlValue(uint32_t mctl)
__STATIC_FORCEINLINE uint32_t SysTimer_GetControlValue(void)
__STATIC_FORCEINLINE void SysTimer_SetSWIRQ(void)
```

<sup>17</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/timer.html](https://doc.nucleisys.com/nuclei_spec/isa/timer.html)

```
__STATIC_FORCEINLINE void SysTimer_ClearSWIRQ(void)
__STATIC_FORCEINLINE uint32_t SysTimer_GetMsipValue(void)
__STATIC_FORCEINLINE void SysTimer_SetMsipValue(uint32_t msip)
__STATIC_FORCEINLINE void SysTimer_SoftwareReset(void)
__STATIC_INLINE uint32_t SysTick_Config(uint64_t ticks)
__STATIC_FORCEINLINE uint32_t SysTick_Reload(uint64_t ticks)
```

*group* **NMSIS\_Core\_SysTimer**

Functions that configure the Core System Timer.

## Functions

```
__STATIC_FORCEINLINE void SysTimer_SetLoadValue(uint64_t value)
```

Set system timer load value.

This function set the system timer load value in MTIMER register.

### Remark

- Load value is 64bits wide.
- SysTimer\_GetLoadValue

### Parameters

- [in] value: value to set system timer MTIMER register.

```
__STATIC_FORCEINLINE uint64_t SysTimer_GetLoadValue(void)
```

Get system timer load value.

This function get the system timer current value in MTIMER register.

**Return** current value(64bit) of system timer MTIMER register.

### Remark

- Load value is 64bits wide.
- SysTimer\_SetLoadValue

```
__STATIC_FORCEINLINE void SysTimer_SetCompareValue(uint64_t value)
```

Set system timer compare value.

This function set the system Timer compare value in MTIMERCMP register.

### Remark

- Compare value is 64bits wide.
- If compare value is larger than current value timer interrupt generate.
- Modify the load value or compare value less to clear the interrupt.
- SysTimer\_GetCompareValue

### Parameters

- [in] value: compare value to set system timer MTIMERCMP register.

```
__STATIC_FORCEINLINE uint64_t SysTimer_GetCompareValue(void)
```

Get system timer compare value.

This function get the system timer compare value in MTIMERCMP register.

**Return** compare value of system timer MTIMERCMP register.

**Remark**

- Compare value is 64bits wide.
- SysTimer\_SetCompareValue

**\_\_STATIC\_FORCEINLINE void SysTimer\_Start(void)**

Enable system timer counter running.

Enable system timer counter running by clear TIMESTOP bit in MTIMECTL register.

**\_\_STATIC\_FORCEINLINE void SysTimer\_Stop(void)**

Stop system timer counter running.

Stop system timer counter running by set TIMESTOP bit in MTIMECTL register.

**\_\_STATIC\_FORCEINLINE void SysTimer\_SetControlValue(uint32\_t mctl)**

Set system timer control value.

This function set the system timer MTIMECTL register value.

**Remark**

- Bit TIMESTOP is used to start and stop timer. Clear TIMESTOP bit to 0 to start timer, otherwise to stop timer.
- Bit CMPCLREN is used to enable auto MTIMER clear to zero when MTIMER >= MTIMER-CMP. Clear CMPCLREN bit to 0 to stop auto clear MTIMER feature, otherwise to enable it.
- Bit CLKSRC is used to select timer clock source. Clear CLKSRC bit to 0 to use *mtime\_toggle\_a*, otherwise use *core\_clk\_aon*
- SysTimer\_GetControlValue

**Parameters**

- [in] mctl: value to set MTIMECTL register

**\_\_STATIC\_FORCEINLINE uint32\_t SysTimer\_GetControlValue(void)**

Get system timer control value.

This function get the system timer MTIMECTL register value.

**Return** MTIMECTL register value

**Remark**

- SysTimer\_SetControlValue

**\_\_STATIC\_FORCEINLINE void SysTimer\_SetSWIRQ(void)**

Trigger or set software interrupt via system timer.

This function set the system timer MSIP bit in MSIP register.

**Remark**

- Set system timer MSIP bit and generate a SW interrupt.
- SysTimer\_ClearSWIRQ
- SysTimer\_GetMsipValue

**\_\_STATIC\_FORCEINLINE void SysTimer\_ClearSWIRQ(void)**

Clear system timer software interrupt pending request.

This function clear the system timer MSIP bit in MSIP register.

**Remark**

- Clear system timer MSIP bit in MSIP register to clear the software interrupt pending.
- SysTimer\_SetSWIRQ
- SysTimer\_GetMsipValue

**\_\_STATIC\_FORCEINLINE uint32\_t SysTimer\_GetMsipValue(void)**

Get system timer MSIP register value.

This function get the system timer MSIP register value.

**Return** Value of Timer MSIP register.

**Remark**

- Bit0 is SW interrupt flag. Bit0 is 1 then SW interrupt set. Bit0 is 0 then SW interrupt clear.
- SysTimer\_SetSWIRQ
- SysTimer\_ClearSWIRQ

**\_\_STATIC\_FORCEINLINE void SysTimer\_SetMsipValue(uint32\_t msip)**

Set system timer MSIP register value.

This function set the system timer MSIP register value.

**Parameters**

- [in] msip: value to set MSIP register

**\_\_STATIC\_FORCEINLINE void SysTimer\_SoftwareReset(void)**

Do software reset request.

This function will do software reset request through MTIMER

- Software need to write *SysTimer\_MSFRST\_KEY* (page 92) to generate software reset request
- The software request flag can be cleared by reset operation to clear

**Remark**

- The software reset is sent to SoC, SoC need to generate reset signal and send back to Core
- This function will not return, it will do while(1) to wait the Core reset happened

**\_\_STATIC\_INLINE uint32\_t SysTick\_Config(uint64\_t ticks)**

System Tick Configuration.

Initializes the System Timer and its non-vector interrupt, and starts the System Tick Timer.

In our default implementation, the timer counter will be set to zero, and it will start a timer compare non-vector interrupt when it matches the ticks user set, during the timer interrupt user should reload the system tick using SysTick\_Reload function or similar function written by user, so it can produce period timer interrupt.

**Return** 0 Function succeeded.

**Return** 1 Function failed.

**Remark**

- For *\_\_NUCLEI\_N\_REV* (page 58) >= 0x0104, the CMPCLREN bit in MTIMECTL is introduced, but we assume that the CMPCLREN bit is set to 0, so MTIMER register will not be auto cleared to 0 when MTIMER >= MTIMERCMP.

- When the variable `__Vendor_SysTickConfig` is set to 1, then the function `SysTick_Config` is not included.
- In this case, the file **<Device>.h** must contain a vendor-specific implementation of this function.
- If user need this function to start a period timer interrupt, then in timer interrupt handler routine code, user should call `SysTick_Reload` with ticks to reload the timer.
- This function only available when `__SYSTIMER_PRESENT == 1` and `__ECLIC_PRESENT == 1` and `__Vendor_SysTickConfig == 0`

**See**

- `SysTimer_SetCompareValue`; `SysTimer_SetLoadValue`

**Parameters**

- [in] `ticks`: Number of ticks between two interrupts.

**\_\_STATIC\_FORCEINLINE uint32\_t SysTick\_Reload(uint64\_t ticks)**

System Tick Reload.

Reload the System Timer Tick when the `MTIMECMP` reached `TIME` value

**Return** 0 Function succeeded.

**Return** 1 Function failed.

**Remark**

- For [\\_\\_NUCLEI\\_N\\_REV](#) (page 58)  $\geq 0x0104$ , the `CMPCLREN` bit in `MTIMECTL` is introduced, but for this `SysTick_Config` function, we assume this `CMPCLREN` bit is set to 0, so in interrupt handler function, user still need to set the `MTIMERCMP` or `MTIMER` to reload the system tick, if vendor want to use this timer's auto clear feature, they can define `__Vendor_SysTickConfig` to 1, and implement `SysTick_Config` and `SysTick_Reload` functions.
- When the variable `__Vendor_SysTickConfig` is set to 1, then the function `SysTick_Reload` is not included.
- In this case, the file **<Device>.h** must contain a vendor-specific implementation of this function.
- This function only available when `__SYSTIMER_PRESENT == 1` and `__ECLIC_PRESENT == 1` and `__Vendor_SysTickConfig == 0`
- Since the `MTIMERCMP` value might overflow, if overflowed, `MTIMER` will be set to 0, and `MTIMERCMP` set to ticks

**See**

- `SysTimer_SetCompareValue`
- `SysTimer_SetLoadValue`

**Parameters**

- [in] `ticks`: Number of ticks between two interrupts.

**SysTick Code Example**

The code below shows the usage of the function `SysTick_Config()` and `SysTick_Reload()` with an GD32VF103 SoC.

Listing 3: gd32vf103\_systick\_example.c

```

1  #include "gd32vf103.h"
2
3  volatile uint32_t msTicks = 0;           /* Variable to store_
   ↳millisecond ticks */
4
5  #define CONFIG_TICKS      (SOC_TIMER_FREQ / 1000)
6  #define SysTick_Handler  eclic_mtip_handler
7
8  void SysTick_Handler(void) {             /* SysTick interrupt_
   ↳Handler. */
9      SysTimer_Reload(CONFIG_TICKS);       /* Call SysTick_Reload to_
   ↳reload timer. */
10     msTicks++;                           /* See startup file_
   ↳startup_gd32vf103.S for SysTick vector */
11 }
12
13 int main (void) {
14     uint32_t returnCode;
15
16     returnCode = SysTick_Config(CONFIG_TICKS); /* Configure SysTick to_
   ↳generate an interrupt every millisecond */
17
18     if (returnCode != 0) {                 /* Check return code for_
   ↳errors */
19         // Error Handling
20     }
21
22     while(1);
23 }

```

## SysTimer Interrupt Code Example

The code below shows the usage of various NMSIS Timer Interrupt functions with an GD32VF103 device.

Listing 4: gd32vf103\_timer\_example1.c

```

1  #include "gd32vf103.h"
2
3  void eclic_mtip_handler(void)
4  {
5      uint64_t now = SysTimer_GetLoadValue();
6      SysTimer_SetCompareValue(now + SOC_TIMER_FREQ/100);
7  }
8
9  static uint32_t int_cnt = 0;
10 void eclic_msip_handler(void)
11 {
12     SysTimer_ClearSWIRQ();
13     int_cnt++;
14 }
15
16 void eclic_global_initialize(void)
17 {

```

(continues on next page)



(continued from previous page)

```

18     ECLIC_SetMth(0);
19     ECLIC_SetCfgNlbits(3);
20 }
21
22 int eclic_register_interrupt(IRQn_Type IRQn, uint8_t shv, uint32_t trig_mode, uint32_t
↳lvl, uint32_t priority, void * handler)
23 {
24     ECLIC_SetShvIRQ(IRQn, shv);
25     ECLIC_SetTrigIRQ(IRQn, trig_mode);
26     ECLIC_SetLevelIRQ(IRQn, lvl);
27     ECLIC_SetPriorityIRQ(IRQn, priority);
28     ECLIC_SetVector(IRQn, (rv_csr_t)(handler));
29     ECLIC_EnableIRQ(IRQn);
30     return 0;
31 }
32
33 void setup_timer(void)
34 {
35     SysTimer_SetLoadValue(0);
36     SysTimer_SetCompareValue(SOC_TIMER_FREQ/100);
37 }
38
39 int main (void)
40 {
41     uint32_t returnCode;
42
43     eclic_global_initialize();           /* initialize ECLIC */
44
45     setup_timer();                       /* initialize timer */
46
47     returnCode = eclic_register_interrupt(SysTimer_IRQn,1,2,8,0,eclic_mtip_handler);
↳/* register system timer interrupt */
48
49     returnCode = eclic_register_interrupt(SysTimerSW_IRQn,1,2,8,0,eclic_msip_handler);
↳ /* register system timer SW interrupt */
50
51     __enable_irq();                     /* enable global
↳interrupt */
52
53     SysTimer_SetSWIRQ();                 /* trigger timer SW
↳interrupt */
54
55     if (returnCode != 0) {               /* Check return code
↳for errors */
56         // Error Handling
57     }
58
59     while(1);
60 }

```

## 2.5.10 Interrupts and Exceptions

## Description

This section explains how to use interrupts and exceptions and access functions for the [Enhanced Core Local Interrupt Controller \(ECLIC\)](#)<sup>18</sup>.

Nuclei provides a template file `startup_device` for each supported compiler. The file must be adapted by the silicon vendor to include interrupt vectors for all device-specific interrupt handlers. Each interrupt handler is defined as a weak function to a dummy handler. These interrupt handlers can be used directly in application software without being adapted by the programmer.

Click [Interrupt](#)<sup>19</sup> to learn more about interrupt handling in Nuclei processor core.

## NMI Interrupt

[NMI interrupt](#)<sup>20</sup> entry is stored by `CSR_MNVEC`. If `CSR_MMSIC[9]` is 1 then NMI entry is the same as **Exception** which get from `CSR_MTVEC`. If `CSR_MMSIC[9]` is 1 NMI entry is reset vector.

## Exception

[Exception](#)<sup>21</sup> has only 1 entry address which stored by `CSR_MTVEC`. All the exceptions will jump to the same entry `exc_entry` defined in `intexc_<Device>.S`.

The table below lists the core exception code of the Nuclei N/NX processors.

Table 8: Core exception code of the Nuclei N/NX processors

Exception Code	Value	Description
<code>InsUnalign_EXCn</code>	0	Instruction address misaligned
<code>InsAccFault_EXCn</code>	1	Instruction access fault
<code>IlleIns_EXCn</code>	2	Illegal instruction
<code>Break_EXCn</code>	3	Breakpoint
<code>LdAddrUnalign_EXCn</code>	4	Load address misaligned
<code>LdFault_EXCn</code>	5	Load access fault
<code>StAddrUnalign_EXCn</code>	6	Store or AMO address misaligned
<code>StAccessFault_EXCn</code>	7	Store or AMO access fault
<code>UmodeEcall_EXCn</code>	8	Environment call from User mode
<code>MmodeEcall_EXCn</code>	11	Environment call from Machine mode
<code>NMI_EXCn</code>	0xff	NMI interrupt

## Vector Table

The Vector Table defines the entry addresses of the ECLIC managed interrupts.

It is typically located at the beginning of the program memory, and you can modify `CSR_MTVT` to reallocate the base address of this vector table, but you need to take care of the base address alignment according to the number of interrupts.

<sup>18</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/eclic.html](https://doc.nucleisys.com/nuclei_spec/isa/eclic.html)

<sup>19</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/interrupt.html](https://doc.nucleisys.com/nuclei_spec/isa/interrupt.html)

<sup>20</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/nmi.html](https://doc.nucleisys.com/nuclei_spec/isa/nmi.html)

<sup>21</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/exception.html](https://doc.nucleisys.com/nuclei_spec/isa/exception.html)

Table 9: base address alignment according to the number of interrupts

Number of Interrupt	Alignment Requirements of CSR MTVT
0 to 16	64-byte
17 to 32	128-byte
33 to 64	256-byte
65 to 128	512-byte
129 to 256	1KB
257 to 512	2KB
513 to 1024	4KB

Interrupt number 0~18 is reserved by Nuclei Core. 19~1023 could be used by Silicon Vendor Device.

Below is an example interrupt allocated table:

```

1 typedef enum IRQn {
2     /***** Nuclei N/NX Processor Core Internal Interrupt Numbers *****/
3     Reserved0_IRQn      = 0,      /*!< Internal reserved
4     Reserved1_IRQn      = 1,      /*!< Internal reserved
5     Reserved2_IRQn      = 2,      /*!< Internal reserved
6     SysTimerSW_IRQn     = 3,      /*!< System Timer SW interrupt
7     Reserved3_IRQn      = 4,      /*!< Internal reserved
8     Reserved4_IRQn      = 5,      /*!< Internal reserved
9     Reserved5_IRQn      = 6,      /*!< Internal reserved
10    SysTimer_IRQn        = 7,      /*!< System Timer Interrupt
11    Reserved6_IRQn       = 8,      /*!< Internal reserved
12    Reserved7_IRQn       = 9,      /*!< Internal reserved
13    Reserved8_IRQn       = 10,     /*!< Internal reserved
14    Reserved9_IRQn       = 11,     /*!< Internal reserved
15    Reserved10_IRQn      = 12,     /*!< Internal reserved
16    Reserved11_IRQn      = 13,     /*!< Internal reserved
17    Reserved12_IRQn      = 14,     /*!< Internal reserved
18    Reserved13_IRQn      = 15,     /*!< Internal reserved
19    Reserved14_IRQn      = 16,     /*!< Internal reserved
20    HardFault_IRQn       = 17,     /*!< Hard Fault, storage access error
21    Reserved15_IRQn      = 18,     /*!< Internal reserved
22

```

(continues on next page)

(continued from previous page)

```

23  /***** GD32VF103 Specific External Interrupt Numbers *****/
24  ↪ *****/
    WWDGT_IRQn      = 19,      /*!< window watchDog timer interrupt
25  ↪ */
    LVD_IRQn        = 20,      /*!< LVD through EXTI line detect
26  ↪ interrupt */
    TAMPER_IRQn     = 21,      /*!< tamper through EXTI line detect
27  ↪ */
    :               :
28  :               :
29  CAN1_EWMC_IRQn   = 85,      /*!< CAN1 EWMC interrupt
30  ↪ */
    USBFS_IRQn      = 86,      /*!< USBFS global interrupt
31  ↪ */
    SOC_INT_MAX,    /*!< Number of total Interrupts
32  ↪ */
    } IRQn_Type;

```

## ECLIC API Definitions

When macro NMSIS\_ECLIC\_VIRTUAL is defined, the ECLIC access functions in the table below must be implemented for virtualizing ECLIC access.

These functions should be implemented in a separate source module. The original NMSIS-Core \_\_ECLIC\_XXX functions are always available independent of NMSIS\_ECLIC\_VIRTUAL macro.

Table 10: ECLIC Access Functions

ECLIC ACCESS FUNCTIONS	NMSIS-CORE FUNCTIONS FOR ECLIC
<i>ECLIC_SetCfgNbBits</i> (page 370)	__ECLIC_SetCfgNbBits()
<i>ECLIC_GetCfgNbBits</i> (page 370)	__ECLIC_GetCfgNbBits()
<i>ECLIC_GetInfoVer</i> (page 370)	__ECLIC_GetInfoVer()
<i>ECLIC_GetInfoCtlBits</i> (page 370)	__ECLIC_GetInfoCtlBits()
<i>ECLIC_GetInfoNum</i> (page 370)	__ECLIC_GetInfoNum()
<i>ECLIC_SetMth</i> (page 370)	__ECLIC_SetMth()
<i>ECLIC_GetMth</i> (page 370)	__ECLIC_GetMth()
<i>ECLIC_EnableIRQ</i> (page 370)	__ECLIC_EnableIRQ()
<i>ECLIC_GetEnableIRQ</i> (page 370)	__ECLIC_GetEnableIRQ()
<i>ECLIC_DisableIRQ</i> (page 370)	__ECLIC_DisableIRQ()
<i>ECLIC_SetPendingIRQ</i> (page 370)	__ECLIC_SetPendingIRQ()
<i>ECLIC_GetPendingIRQ</i> (page 370)	__ECLIC_GetPendingIRQ()
<i>ECLIC_ClearPendingIRQ</i> (page 370)	__ECLIC_ClearPendingIRQ()
<i>ECLIC_SetTrigIRQ</i> (page 370)	__ECLIC_SetTrigIRQ()
<i>ECLIC_GetTrigIRQ</i> (page 370)	__ECLIC_GetTrigIRQ()
<i>ECLIC_SetShvIRQ</i> (page 370)	__ECLIC_SetShvIRQ()
<i>ECLIC_GetShvIRQ</i> (page 370)	__ECLIC_GetShvIRQ()
<i>ECLIC_SetCtrlIRQ</i> (page 370)	__ECLIC_SetCtrlIRQ()
<i>ECLIC_GetCtrlIRQ</i> (page 370)	__ECLIC_GetCtrlIRQ()
<i>ECLIC_SetLevelIRQ</i> (page 370)	__ECLIC_SetLevelIRQ()
<i>ECLIC_GetLevelIRQ</i> (page 370)	__ECLIC_GetLevelIRQ()
<i>ECLIC_SetPriorityIRQ</i> (page 370)	__ECLIC_SetPriorityIRQ()
<i>ECLIC_GetPriorityIRQ</i> (page 370)	__ECLIC_GetPriorityIRQ()

When NMSIS\_VECTAB\_VIRTUAL macro is defined, the functions in the table below must be replaced to virtualize the API access functions to the interrupt vector table.

The ECLIC vector table API should be implemented in a separate source module.

This allows, for example, alternate implementations to relocate the vector table from flash to RAM on the first vector table update.

The original NMSIS-Core functions are always available, but prefixed with `__ECLIC`.

Table 11: ECLIC Vector Access Functions

ECLIC Vector Table Access	NMSIS-CORE FUNCTIONS
<code>ECLIC_SetVector</code> (page 370)	<code>__ECLIC_SetVector()</code>
<code>ECLIC_GetVector</code> (page 370)	<code>__ECLIC_GetVector()</code>

## ECLIC Function Usage

The code below shows the usage of various NMSIS ECLIC flow with an GD32VF103 device.

Listing 5: gd32vf103\_interrupt\_example1.c

```

1  #include "gd32vf103.h"
2
3  // Vector interrupt which could be nested
4  __INTERCEPT void eclic_button1_handler(void)
5  {
6      SAVE_IRQ_CSR_CONTEXT(); /* save mepc,
7      ↳ mcause, msubm enable interrupts */
8      GPIO_REG(GPIO_OUTPUT_VAL) |= (1 << GREEN_LED_GPIO_OFFSET); /* Green LED On_
9      ↳ */
10     GPIO_REG(GPIO_RISE_IP) = (0x1 << BUTTON_1_GPIO_OFFSET); /* Clear the_
11     ↳ GPIO Pending interrupt by writing 1. */
12     RESTORE_IRQ_CSR_CONTEXT(); /* disable_
13     ↳ interrupts, restore mepc, mcause, msubm */
14 }
15 // Non-vector interrupt
16 void eclic_button2_handler(void)
17 {
18     GPIO_REG(GPIO_OUTPUT_VAL) |= (1 << GREEN_LED_GPIO_OFFSET); /* Green LED On_
19     ↳ */
20     GPIO_REG(GPIO_RISE_IP) = (0x1 << BUTTON_2_GPIO_OFFSET); /* Clear the_
21     ↳ GPIO Pending interrupt by writing 1. */
22 }
23 void eclic_global_initialize(void)
24 {
25     ECLIC_SetMth(0);
26     ECLIC_SetCfgNbBits(3);
27 }
28 int eclic_register_interrupt(IRQn_Type IRQn, uint8_t shv, uint32_t trig_mode, uint32_t
29 ↳ lvl, uint32_t priority, void * handler)
30 {

```

(continues on next page)

(continued from previous page)

```

29     ECLIC_SetShvIRQ(IRQn, shv);
30     ECLIC_SetTrigIRQ(IRQn, trig_mode);
31     ECLIC_SetLevelIRQ(IRQn, lvl);
32     ECLIC_SetPriorityIRQ(IRQn, priority);
33     ECLIC_SetVector(IRQn, (rv_csr_t)(handler));
34     ECLIC_EnableIRQ(IRQn);
35     return 0;
36 }
37
38 int main (void)
39 {
40     uint32_t returnCode;
41
42     eclic_global_initialize();           /* initialize ECLIC */
43
44     GPIO_init();                       /* initialize GPIO */
45
46     returnCode = eclic_register_interrupt(BTN1_IRQn, 1, 2, 1, 0, Button1_IRQHandler); /*
↪register system button1 interrupt */
47     returnCode = eclic_register_interrupt(BTN2_IRQn, 0, 2, 2, 0, Button2_IRQHandler); /*
↪register system button2 interrupt */
48
49     __enable_irq();                   /* enable global
↪interrupt */
50
51     if (returnCode != 0) {             /* Check return code
↪for errors */
52         // Error Handling
53     }
54
55     while(1);
56 }

```

## Interrupt and Exception API

**enum NMSIS\_Core\_IntExc::IRQn\_Type**

Values:

```

Reserved0_IRQn = 0
Reserved1_IRQn = 1
Reserved2_IRQn = 2
SysTimerSW_IRQn = 3
Reserved3_IRQn = 4
Reserved4_IRQn = 5
Reserved5_IRQn = 6
SysTimer_IRQn = 7
Reserved6_IRQn = 8
Reserved7_IRQn = 9
Reserved8_IRQn = 10

```

```

Reserved9_IRQn = 11
Reserved10_IRQn = 12
Reserved11_IRQn = 13
Reserved12_IRQn = 14
Reserved13_IRQn = 15
Reserved14_IRQn = 16
Reserved15_IRQn = 17
Reserved16_IRQn = 18
FirstDeviceSpecificInterrupt_IRQn = 19
SOC_INT_MAX

__STATIC_FORCEINLINE void __ECLIC_SetCfgNlbits(uint32_t nlbits)
__STATIC_FORCEINLINE uint32_t __ECLIC_GetCfgNlbits(void)
__STATIC_FORCEINLINE uint32_t __ECLIC_GetInfoVer(void)
__STATIC_FORCEINLINE uint32_t __ECLIC_GetInfoCtlbits(void)
__STATIC_FORCEINLINE uint32_t __ECLIC_GetInfoNum(void)
__STATIC_FORCEINLINE void __ECLIC_SetMth(uint8_t mth)
__STATIC_FORCEINLINE uint8_t __ECLIC_GetMth(void)
__STATIC_FORCEINLINE void __ECLIC_EnableIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE uint32_t __ECLIC_GetEnableIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_DisableIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE int32_t __ECLIC_GetPendingIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_SetPendingIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_ClearPendingIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_SetTrigIRQ(IRQn_Type IRQn, uint32_t trig)
__STATIC_FORCEINLINE uint32_t __ECLIC_GetTrigIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_SetShvIRQ(IRQn_Type IRQn, uint32_t shv)
__STATIC_FORCEINLINE uint32_t __ECLIC_GetShvIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_SetCtrlIRQ(IRQn_Type IRQn, uint8_t intctrl)
__STATIC_FORCEINLINE uint8_t __ECLIC_GetCtrlIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_SetLevelIRQ(IRQn_Type IRQn, uint8_t lvl_abs)
__STATIC_FORCEINLINE uint8_t __ECLIC_GetLevelIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_SetPriorityIRQ(IRQn_Type IRQn, uint8_t pri)
__STATIC_FORCEINLINE uint8_t __ECLIC_GetPriorityIRQ(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __ECLIC_SetVector(IRQn_Type IRQn, rv_csr_t vector)
__STATIC_FORCEINLINE rv_csr_t __ECLIC_GetVector(IRQn_Type IRQn)
__STATIC_FORCEINLINE void __set_exc_entry(rv_csr_t addr)

```

```
__STATIC_FORCEINLINE rv_csr_t __get_exc_entry(void)
__STATIC_FORCEINLINE void __set_nonvec_entry(rv_csr_t addr)
__STATIC_FORCEINLINE rv_csr_t __get_nonvec_entry(void)
__STATIC_FORCEINLINE rv_csr_t __get_nmi_entry(void)
ECLIC_SetCfgNlbits __ECLIC_SetCfgNlbits
ECLIC_GetCfgNlbits __ECLIC_GetCfgNlbits
ECLIC_GetInfoVer __ECLIC_GetInfoVer
ECLIC_GetInfoCtlbits __ECLIC_GetInfoCtlbits
ECLIC_GetInfoNum __ECLIC_GetInfoNum
ECLIC_SetMth __ECLIC_SetMth
ECLIC_GetMth __ECLIC_GetMth
ECLIC_EnableIRQ __ECLIC_EnableIRQ
ECLIC_GetEnableIRQ __ECLIC_GetEnableIRQ
ECLIC_DisableIRQ __ECLIC_DisableIRQ
ECLIC_SetPendingIRQ __ECLIC_SetPendingIRQ
ECLIC_GetPendingIRQ __ECLIC_GetPendingIRQ
ECLIC_ClearPendingIRQ __ECLIC_ClearPendingIRQ
ECLIC_SetTrigIRQ __ECLIC_SetTrigIRQ
ECLIC_GetTrigIRQ __ECLIC_GetTrigIRQ
ECLIC_SetShvIRQ __ECLIC_SetShvIRQ
ECLIC_GetShvIRQ __ECLIC_GetShvIRQ
ECLIC_SetCtrlIRQ __ECLIC_SetCtrlIRQ
ECLIC_GetCtrlIRQ __ECLIC_GetCtrlIRQ
ECLIC_SetLevelIRQ __ECLIC_SetLevelIRQ
ECLIC_GetLevelIRQ __ECLIC_GetLevelIRQ
ECLIC_SetPriorityIRQ __ECLIC_SetPriorityIRQ
ECLIC_GetPriorityIRQ __ECLIC_GetPriorityIRQ
ECLIC_SetVector __ECLIC_SetVector
ECLIC_GetVector __ECLIC_GetVector
SAVE_IRQ_CSR_CONTEXT ()
RESTORE_IRQ_CSR_CONTEXT ()
```

*group* **NMSIS\_Core\_IntExc**

Functions that manage interrupts and exceptions via the ECLIC.



## Defines

```

ECLIC_SetCfgNlbits __ECLIC_SetCfgNlbits
ECLIC_GetCfgNlbits __ECLIC_GetCfgNlbits
ECLIC_GetInfoVer __ECLIC_GetInfoVer
ECLIC_GetInfoCtlbits __ECLIC_GetInfoCtlbits
ECLIC_GetInfoNum __ECLIC_GetInfoNum
ECLIC_SetMth __ECLIC_SetMth
ECLIC_GetMth __ECLIC_GetMth
ECLIC_EnableIRQ __ECLIC_EnableIRQ
ECLIC_GetEnableIRQ __ECLIC_GetEnableIRQ
ECLIC_DisableIRQ __ECLIC_DisableIRQ
ECLIC_SetPendingIRQ __ECLIC_SetPendingIRQ
ECLIC_GetPendingIRQ __ECLIC_GetPendingIRQ
ECLIC_ClearPendingIRQ __ECLIC_ClearPendingIRQ
ECLIC_SetTrigIRQ __ECLIC_SetTrigIRQ
ECLIC_GetTrigIRQ __ECLIC_GetTrigIRQ
ECLIC_SetShvIRQ __ECLIC_SetShvIRQ
ECLIC_GetShvIRQ __ECLIC_GetShvIRQ
ECLIC_SetCtrlIRQ __ECLIC_SetCtrlIRQ
ECLIC_GetCtrlIRQ __ECLIC_GetCtrlIRQ
ECLIC_SetLevelIRQ __ECLIC_SetLevelIRQ
ECLIC_GetLevelIRQ __ECLIC_GetLevelIRQ
ECLIC_SetPriorityIRQ __ECLIC_SetPriorityIRQ
ECLIC_GetPriorityIRQ __ECLIC_GetPriorityIRQ
ECLIC_SetVector __ECLIC_SetVector
ECLIC_GetVector __ECLIC_GetVector

```

### **SAVE\_IRQ\_CSR\_CONTEXT** ( )

Save necessary CSRs into variables for vector interrupt nesting.

This macro is used to declare variables which are used for saving CSRs(MCAUSE, MEPC, MSUB), and it will read these CSR content into these variables, it need to be used in a vector-interrupt if nesting is required.

### **Remark**

- Interrupt will be enabled after this macro is called
- It need to be used together with **RESTORE\_IRQ\_CSR\_CONTEXT**
- Don't use variable names `__mcause`, `__mpec`, `__msubm` in your ISR code
- If you want to enable interrupt nesting feature for vector interrupt, you can do it like this:

```
// __INTERRUPT attribute will generates function entry and exit_
↪sequences suitable
// for use in an interrupt handler when this attribute is present
__INTERRUPT void eclic_mtip_handler(void)
{
    // Must call this to save CSRs
    SAVE_IRQ_CSR_CONTEXT();
    // !!!Interrupt is enabled here!!!
    // !!!Higher priority interrupt could nest it!!!

    // put you own interrupt handling code here

    // Must call this to restore CSRs
    RESTORE_IRQ_CSR_CONTEXT();
}
```

**RESTORE\_IRQ\_CSR\_CONTEXT()**

Restore necessary CSRs from variables for vector interrupt nesting.

This macro is used restore CSRs(MCAUSE, MEPC, MSUB) from pre-defined variables in SAVE\_IRQ\_CSR\_CONTEXT macro.

**Remark**

- Interrupt will be disabled after this macro is called
- It need to be used together with SAVE\_IRQ\_CSR\_CONTEXT

**Enums****enum IRQn\_Type**

Definition of IRQn numbers.

The core interrupt enumeration names for IRQn values are defined in the file **<Device>.h**.

- Interrupt ID(IRQn) from 0 to 18 are reserved for core internal interrupts.
- Interrupt ID(IRQn) start from 19 represent device-specific external interrupts.
- The first device-specific interrupt has the IRQn value 19.

The table below describes the core interrupt names and their availability in various Nuclei Cores.

*Values:*

**Reserved0\_IRQn** = 0

Internal reserved.

**Reserved1\_IRQn** = 1

Internal reserved.

**Reserved2\_IRQn** = 2

Internal reserved.

**SysTimerSW\_IRQn** = 3

System Timer SW interrupt.

**Reserved3\_IRQn** = 4

Internal reserved.

**Reserved4\_IRQn** = 5

Internal reserved.

**Reserved5\_IRQn** = 6  
Internal reserved.

**SysTimer\_IRQn** = 7  
System Timer Interrupt.

**Reserved6\_IRQn** = 8  
Internal reserved.

**Reserved7\_IRQn** = 9  
Internal reserved.

**Reserved8\_IRQn** = 10  
Internal reserved.

**Reserved9\_IRQn** = 11  
Internal reserved.

**Reserved10\_IRQn** = 12  
Internal reserved.

**Reserved11\_IRQn** = 13  
Internal reserved.

**Reserved12\_IRQn** = 14  
Internal reserved.

**Reserved13\_IRQn** = 15  
Internal reserved.

**Reserved14\_IRQn** = 16  
Internal reserved.

**Reserved15\_IRQn** = 17  
Internal reserved.

**Reserved16\_IRQn** = 18  
Internal reserved.

**FirstDeviceSpecificInterrupt\_IRQn** = 19  
First Device Specific Interrupt.

**SOC\_INT\_MAX**  
Number of total interrupts.

## Functions

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_SetCfgNlbits(uint32\_t nlbits)**  
Set nlbits value.

This function set the nlbits value of CLICCFG register.

### Remark

- nlbits is used to set the width of level in the CLICINTCTL[i].

### See

- ECLIC\_GetCfgNlbits

### Parameters

- [in] nlbits: nlbits value

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_ECLIC\_GetCfgNlbits(void)**

Get nlbits value.

This function get the nlbits value of CLICCFG register.

**Return** nlbits value of CLICCFG register

**Remark**

- nlbits is used to set the width of level in the CLICINTCTL[i].

**See**

- ECLIC\_SetCfgNlbits

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_ECLIC\_GetInfoVer(void)**

Get the ECLIC version number.

This function gets the hardware version information from CLICINFO register.

**Return** hardware version number in CLICINFO register.

**Remark**

- This function gets hardware version information from CLICINFO register.
- Bit 20:17 for architecture version, bit 16:13 for implementation version.

**See**

- ECLIC\_GetInfoNum

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_ECLIC\_GetInfoCtlbits(void)**

Get CLICINTCTLBITS.

This function gets CLICINTCTLBITS from CLICINFO register.

**Return** CLICINTCTLBITS from CLICINFO register.

**Remark**

- In the CLICINTCTL[i] registers, with  $2 \leq \text{CLICINTCTLBITS} \leq 8$ .
- The implemented bits are kept left-justified in the most-significant bits of each 8-bit CLICINTCTL[I] register, with the lower unimplemented bits treated as hardwired to 1.

**See**

- ECLIC\_GetInfoNum

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_ECLIC\_GetInfoNum(void)**

Get number of maximum interrupt inputs supported.

This function gets number of maximum interrupt inputs supported from CLICINFO register.

**Return** number of maximum interrupt inputs supported from CLICINFO register.

**Remark**

- This function gets number of maximum interrupt inputs supported from CLICINFO register.
- The num\_interrupt field specifies the actual number of maximum interrupt inputs supported in this implementation.

**See**

- ECLIC\_GetInfoCtlbits

---

```
__STATIC_FORCEINLINE void __ECLIC_SetMth(uint8_t mth)
```

Set Machine Mode Interrupt Level Threshold.

This function sets machine mode interrupt level threshold.

See

- ECLIC\_GetMth

#### Parameters

- [in] mth: Interrupt Level Threshold.

```
__STATIC_FORCEINLINE uint8_t __ECLIC_GetMth(void)
```

Get Machine Mode Interrupt Level Threshold.

This function gets machine mode interrupt level threshold.

**Return** Interrupt Level Threshold.

See

- ECLIC\_SetMth

```
__STATIC_FORCEINLINE void __ECLIC_EnableIRQ(IRQn_Type IRQn)
```

Enable a specific interrupt.

This function enables the specific interrupt *IRQn*.

#### Remark

- IRQn must not be negative.

See

- ECLIC\_DisableIRQ

#### Parameters

- [in] IRQn: Interrupt number

```
__STATIC_FORCEINLINE uint32_t __ECLIC_GetEnableIRQ(IRQn_Type IRQn)
```

Get a specific interrupt enable status.

This function returns the interrupt enable status for the specific interrupt *IRQn*.

#### Return

- 0 Interrupt is not enabled
- 1 Interrupt is pending

#### Remark

- IRQn must not be negative.

See

- ECLIC\_EnableIRQ
- ECLIC\_DisableIRQ

#### Parameters

- [in] IRQn: Interrupt number

```
__STATIC_FORCEINLINE void __ECLIC_DisableIRQ(IRQn_Type IRQn)
```

Disable a specific interrupt.

This function disables the specific interrupt *IRQn*.

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_EnableIRQ

**Parameters**

- [in] IRQn: Number of the external interrupt to disable

**\_\_STATIC\_FORCEINLINE int32\_t \_\_ECLIC\_GetPendingIRQ(IRQn\_Type IRQn)**

Get the pending specific interrupt.

This function returns the pending status of the specific interrupt *IRQn*.

**Return**

- 0 Interrupt is not pending
- 1 Interrupt is pending

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_SetPendingIRQ
- ECLIC\_ClearPendingIRQ

**Parameters**

- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_SetPendingIRQ(IRQn\_Type IRQn)**

Set a specific interrupt to pending.

This function sets the pending bit for the specific interrupt *IRQn*.

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_GetPendingIRQ
- ECLIC\_ClearPendingIRQ

**Parameters**

- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_ClearPendingIRQ(IRQn\_Type IRQn)**

Clear a specific interrupt from pending.

This function removes the pending state of the specific interrupt *IRQn*. *IRQn* cannot be a negative number.

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_SetPendingIRQ
- ECLIC\_GetPendingIRQ

**Parameters**

- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_SetTrigIRQ(IRQn\_Type IRQn, uint32\_t trig)**

Set trigger mode and polarity for a specific interrupt.

This function set trigger mode and polarity of the specific interrupt *IRQn*.

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_GetTrigIRQ

**Parameters**

- [in] IRQn: Interrupt number
- [in] trig:
  - 00 level trigger, [ECLIC\\_LEVEL\\_TRIGGER](#) (page 91)
  - 01 positive edge trigger, [ECLIC\\_POSTIVE\\_EDGE\\_TRIGGER](#) (page 91)
  - 02 level trigger, [ECLIC\\_LEVEL\\_TRIGGER](#) (page 91)
  - 03 negative edge trigger, [ECLIC\\_NEGTIVE\\_EDGE\\_TRIGGER](#) (page 91)

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_ECLIC\_GetTrigIRQ(IRQn\_Type IRQn)**

Get trigger mode and polarity for a specific interrupt.

This function get trigger mode and polarity of the specific interrupt *IRQn*.

**Return**

- 00 level trigger, [ECLIC\\_LEVEL\\_TRIGGER](#) (page 91)
- 01 positive edge trigger, [ECLIC\\_POSTIVE\\_EDGE\\_TRIGGER](#) (page 91)
- 02 level trigger, [ECLIC\\_LEVEL\\_TRIGGER](#) (page 91)
- 03 negative edge trigger, [ECLIC\\_NEGTIVE\\_EDGE\\_TRIGGER](#) (page 91)

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_SetTrigIRQ

**Parameters**

- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_SetShvIRQ(IRQn\_Type IRQn, uint32\_t shv)**

Set interrupt working mode for a specific interrupt.

This function set selective hardware vector or non-vector working mode of the specific interrupt *IRQn*.

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_GetShvIRQ

**Parameters**

- [in] IRQn: Interrupt number
- [in] shv:
  - 0 non-vector mode, *ECLIC\_NON\_VECTOR\_INTERRUPT* (page 90)
  - 1 vector mode, *ECLIC\_VECTOR\_INTERRUPT* (page 90)

**\_\_STATIC\_FORCEINLINE uint32\_t \_\_ECLIC\_GetShvIRQ(IRQn\_Type IRQn)**

Get interrupt working mode for a specific interrupt.

This function get selective hardware vector or non-vector working mode of the specific interrupt *IRQn*.

**Return** shv

- 0 non-vector mode, *ECLIC\_NON\_VECTOR\_INTERRUPT* (page 90)
- 1 vector mode, *ECLIC\_VECTOR\_INTERRUPT* (page 90)

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_SetShvIRQ

**Parameters**

- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_SetCtrlIRQ(IRQn\_Type IRQn, uint8\_t intctrl)**

Modify ECLIC Interrupt Input Control Register for a specific interrupt.

This function modify ECLIC Interrupt Input Control(CLICINTCTL[i]) register of the specific interrupt *IRQn*.

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_GetCtrlIRQ

**Parameters**

- [in] IRQn: Interrupt number
- [in] intctrl: Set value for CLICINTCTL[i] register

**\_\_STATIC\_FORCEINLINE uint8\_t \_\_ECLIC\_GetCtrlIRQ(IRQn\_Type IRQn)**

Get ECLIC Interrupt Input Control Register value for a specific interrupt.

This function modify ECLIC Interrupt Input Control register of the specific interrupt *IRQn*.

**Return** value of ECLIC Interrupt Input Control register

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_SetCtrlIRQ

**Parameters**



- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_SetLevelIRQ(IRQn\_Type IRQn, uint8\_t lvl\_abs)**  
Set ECLIC Interrupt level of a specific interrupt.

This function set interrupt level of the specific interrupt *IRQn*.

#### Remark

- IRQn must not be negative.
- If lvl\_abs to be set is larger than the max level allowed, it will be force to be max level.
- When you set level value you need use clciinfo.nbits to get the width of level. Then we could know the maximum of level. CLICINTCTLBITS is how many total bits are present in the CLICINTCTL register.

#### See

- ECLIC\_GetLevelIRQ

#### Parameters

- [in] IRQn: Interrupt number
- [in] lvl\_abs: Interrupt level

**\_\_STATIC\_FORCEINLINE uint8\_t \_\_ECLIC\_GetLevelIRQ(IRQn\_Type IRQn)**  
Get ECLIC Interrupt level of a specific interrupt.

This function get interrupt level of the specific interrupt *IRQn*.

**Return** Interrupt level

#### Remark

- IRQn must not be negative.

#### See

- ECLIC\_SetLevelIRQ

#### Parameters

- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_SetPriorityIRQ(IRQn\_Type IRQn, uint8\_t pri)**  
Get ECLIC Interrupt priority of a specific interrupt.

This function get interrupt priority of the specific interrupt *IRQn*.

#### Remark

- IRQn must not be negative.
- If pri to be set is larger than the max priority allowed, it will be force to be max priority.
- Priority width is CLICINTCTLBITS minus clciinfo.nbits if clciinfo.nbits is less than CLICINTCTLBITS. Otherwise priority width is 0.

#### See

- ECLIC\_GetPriorityIRQ

#### Parameters

- [in] IRQn: Interrupt number
- [in] pri: Interrupt priority

**\_\_STATIC\_FORCEINLINE uint8\_t \_\_ECLIC\_GetPriorityIRQ(IRQn\_Type IRQn)**

Get ECLIC Interrupt priority of a specific interrupt.

This function get interrupt priority of the specific interrupt *IRQn*.

**Return** Interrupt priority

**Remark**

- IRQn must not be negative.

**See**

- ECLIC\_SetPriorityIRQ

**Parameters**

- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_ECLIC\_SetVector(IRQn\_Type IRQn, rv\_csr\_t vector)**

Set Interrupt Vector of a specific interrupt.

This function set interrupt handler address of the specific interrupt *IRQn*.

**Remark**

- IRQn must not be negative.
- You can set the CSR\_CSR\_MTVT to set interrupt vector table entry address.
- If your vector table is placed in readonly section, the vector for IRQn will not be modified. For this case, you need to use the correct irq handler name defined in your vector table as your irq handler function name.
- This function will only work correctly when the vector table is placed in an read-write enabled section.

**See**

- ECLIC\_GetVector

**Parameters**

- [in] IRQn: Interrupt number
- [in] vector: Interrupt handler address

**\_\_STATIC\_FORCEINLINE rv\_csr\_t \_\_ECLIC\_GetVector(IRQn\_Type IRQn)**

Get Interrupt Vector of a specific interrupt.

This function get interrupt handler address of the specific interrupt *IRQn*.

**Return** Interrupt handler address

**Remark**

- IRQn must not be negative.
- You can read CSR\_CSR\_MTVT to get interrupt vector table entry address.

**See**

- ECLIC\_SetVector

**Parameters**

- [in] IRQn: Interrupt number

**\_\_STATIC\_FORCEINLINE void \_\_set\_exc\_entry(rv\_csr\_t addr)**

Set Exception entry address.

This function set exception handler address to 'CSR\_MTVEC'.

**Remark**

- This function use to set exception handler address to 'CSR\_MTVEC'. Address is 4 bytes align.

**See**

- \_\_get\_exc\_entry

**Parameters**

- [in] addr: Exception handler address

**\_\_STATIC\_FORCEINLINE rv\_csr\_t \_\_get\_exc\_entry(void)**

Get Exception entry address.

This function get exception handler address from 'CSR\_MTVEC'.

**Return** Exception handler address

**Remark**

- This function use to get exception handler address from 'CSR\_MTVEC'. Address is 4 bytes align

**See**

- \_\_set\_exc\_entry

**\_\_STATIC\_FORCEINLINE void \_\_set\_nonvec\_entry(rv\_csr\_t addr)**

Set Non-vector interrupt entry address.

This function set Non-vector interrupt address.

**Remark**

- This function use to set non-vector interrupt entry address to 'CSR\_MTVT2' if
- CSR\_MTVT2 bit0 is 1. If 'CSR\_MTVT2' bit0 is 0 then set address to 'CSR\_MTVEC'

**See**

- \_\_get\_nonvec\_entry

**Parameters**

- [in] addr: Non-vector interrupt entry address

**\_\_STATIC\_FORCEINLINE rv\_csr\_t \_\_get\_nonvec\_entry(void)**

Get Non-vector interrupt entry address.

This function get Non-vector interrupt address.

**Return** Non-vector interrupt handler address

**Remark**

- This function use to get non-vector interrupt entry address from 'CSR\_MTVT2' if
- CSR\_MTVT2 bit0 is 1. If 'CSR\_MTVT2' bit0 is 0 then get address from 'CSR\_MTVEC'.

**See**

- \_\_set\_nonvec\_entry

**\_\_STATIC\_FORCEINLINE rv\_csr\_t \_\_get\_nmi\_entry(void)**

Get NMI interrupt entry from 'CSR\_MNVEC'.

This function get NMI interrupt address from 'CSR\_MNVEC'.

**Return** NMI interrupt handler address

**Remark**

- This function use to get NMI interrupt handler address from 'CSR\_MNVEC'. If CSR\_MMISC\_CTL[9] = 1 'CSR\_MNVEC'
- will be equal as mtvec. If CSR\_MMISC\_CTL[9] = 0 'CSR\_MNVEC' will be equal as reset vector.
- NMI entry is defined via [CSR\\_MMISC\\_CTL](#) (page 70), writing to [CSR\\_MNVEC](#) (page 70) will be ignored.

## 2.5.11 FPU Functions

### group NMSIS\_Core\_FPU\_Functions

Functions that related to the RISC-V FPU (F and D extension).

Nuclei provided floating point unit by RISC-V F and D extension.

- F extension adds single-precision floating-point computational instructions compliant with the IEEE 754-2008 arithmetic standard, \_\_RISCV\_FLEN = 32. The F extension adds 32 floating-point registers, f0-f31, each 32 bits wide, and a floating-point control and status register fcsr, which contains the operating mode and exception status of the floating-point unit.
- D extension adds double-precision floating-point computational instructions compliant with the IEEE 754-2008 arithmetic standard. The D extension widens the 32 floating-point registers, f0-f31, to 64 bits, \_\_RISCV\_FLEN = 64

### Defines

**\_\_RISCV\_FLEN** 64

**\_\_get\_FCSR()** [\\_\\_RV\\_CSR\\_READ](#) (page 61)([CSR\\_FCSR](#) (page 63))

Get FCSR CSR Register.

**\_\_set\_FCSR(val)** [\\_\\_RV\\_CSR\\_WRITE](#) (page 61)([CSR\\_FCSR](#) (page 63), (val))

Set FCSR CSR Register with val.

**\_\_get\_FRM()** [\\_\\_RV\\_CSR\\_READ](#) (page 61)([CSR\\_FRM](#) (page 63))

Get FRM CSR Register.

**\_\_set\_FRM(val)** [\\_\\_RV\\_CSR\\_WRITE](#) (page 61)([CSR\\_FRM](#) (page 63), (val))

Set FRM CSR Register with val.

**\_\_get\_FFLAGS()** [\\_\\_RV\\_CSR\\_READ](#) (page 61)([CSR\\_FFLAGS](#) (page 63))

Get FFLAGS CSR Register.

**\_\_set\_FFLAGS(val)** [\\_\\_RV\\_CSR\\_WRITE](#) (page 61)([CSR\\_FFLAGS](#) (page 63), (val))

Set FFLAGS CSR Register with val.

**\_\_enable\_FPU()** [\\_\\_RV\\_CSR\\_SET](#) (page 61)([CSR\\_MSTATUS](#) (page 64), [MSTATUS\\_FS](#) (page 71))

Enable FPU Unit.

`__disable_FPU()` [\\_\\_RV\\_CSR\\_CLEAR](#) (page 62) [\(CSR\\_MSTATUS](#) (page 64), [MSTATUS\\_FS](#) (page 71))  
Disable FPU Unit.

- We can save power by disable FPU Unit.
- When FPU Unit is disabled, any access to FPU related CSR registers and FPU instructions will cause illegal Instruction Exception.

`__RV_FLW` (freg, addr, ofs)

Load a single-precision value from memory into float point register freg using flw instruction.

The FLW instruction loads a single-precision floating point value from memory address (addr + ofs) into floating point register freg(f0-f31)

#### Remark

- FLW and FSW operations need to make sure the address is 4 bytes aligned, otherwise it will cause exception code 4(Load address misaligned) or 6 (Store/AMO address misaligned)
- FLW and FSW do not modify the bits being transferred; in particular, the payloads of non-canonical NaNs are preserved

#### Parameters

- [in] freg: The floating point register, eg. [FREG\(0\)](#) (page 77), f0
- [in] addr: The memory base address, 4 byte aligned required
- [in] ofs: a 12-bit immediate signed byte offset value, should be an const value

`__RV_FSW` (freg, addr, ofs)

Store a single-precision value from float point freg into memory using fsw instruction.

The FSW instruction stores a single-precision value from floating point register to memory

#### Remark

- FLW and FSW operations need to make sure the address is 4 bytes aligned, otherwise it will cause exception code 4(Load address misaligned) or 6 (Store/AMO address misaligned)
- FLW and FSW do not modify the bits being transferred; in particular, the payloads of non-canonical NaNs are preserved

#### Parameters

- [in] freg: The floating point register(f0-f31), eg. [FREG\(0\)](#) (page 77), f0
- [in] addr: The memory base address, 4 byte aligned required
- [in] ofs: a 12-bit immediate signed byte offset value, should be an const value

`__RV_FLD` (freg, addr, ofs)

Load a double-precision value from memory into float point register freg using fld instruction.

The FLD instruction loads a double-precision floating point value from memory address (addr + ofs) into floating point register freg(f0-f31)

#### Attention

- Function only available for double precision floating point unit, FLEN = 64

#### Remark

- FLD and FSD operations need to make sure the address is 8 bytes aligned, otherwise it will cause exception code 4(Load address misaligned) or 6 (Store/AMO address misaligned)

- FLD and FSD do not modify the bits being transferred; in particular, the payloads of non-canonical NaNs are preserved.

**Parameters**

- [in] *freg*: The floating point register, eg. *FREG(0)* (page 77), *f0*
- [in] *addr*: The memory base address, 8 byte aligned required
- [in] *ofs*: a 12-bit immediate signed byte offset value, should be an const value

**\_\_RV\_FSD** (*freg*, *addr*, *ofs*)

Store a double-precision value from float point *freg* into memory using *fsd* instruction.

The FSD instruction stores double-precision value from floating point register to memory

**Attention**

- Function only available for double precision floating point unit, *FLEN* = 64

**Remark**

- FLD and FSD operations need to make sure the address is 8 bytes aligned, otherwise it will cause exception code 4(Load address misaligned) or 6 (Store/AMO address misaligned)
- FLD and FSD do not modify the bits being transferred; in particular, the payloads of non-canonical NaNs are preserved.

**Parameters**

- [in] *freg*: The floating point register(*f0-f31*), eg. *FREG(0)* (page 77), *f0*
- [in] *addr*: The memory base address, 8 byte aligned required
- [in] *ofs*: a 12-bit immediate signed byte offset value, should be an const value

**\_\_RV\_FLOAD** *\_\_RV\_FLD* (page 383)

Load a float point value from memory into float point register *freg* using *flw/fld* instruction.

- For Single-Precision Floating-Point Mode(*\_\_FPU\_PRESENT* == 1, *\_\_RISCV\_FLEN* == 32): It will call *\_\_RV\_FLW* (page 383) to load a single-precision floating point value from memory to floating point register
- For Double-Precision Floating-Point Mode(*\_\_FPU\_PRESENT* == 2, *\_\_RISCV\_FLEN* == 64): It will call *\_\_RV\_FLD* (page 383) to load a double-precision floating point value from memory to floating point register

**Attention** Function behaviour is different for *\_\_FPU\_PRESENT* = 1 or 2, please see the real function this macro represent

**\_\_RV\_FSTORE** *\_\_RV\_FSD* (page 384)

Store a float value from float point *freg* into memory using *fsw/fsd* instruction.

- For Single-Precision Floating-Point Mode(*\_\_FPU\_PRESENT* == 1, *\_\_RISCV\_FLEN* == 32): It will call *\_\_RV\_FSW* (page 383) to store floating point register into memory
- For Double-Precision Floating-Point Mode(*\_\_FPU\_PRESENT* == 2, *\_\_RISCV\_FLEN* == 64): It will call *\_\_RV\_FSD* (page 384) to store floating point register into memory

**Attention** Function behaviour is different for *\_\_FPU\_PRESENT* = 1 or 2, please see the real function this macro represent

**SAVE\_FPU\_CONTEXT ( )**

Save FPU context into variables for interrupt nesting.

This macro is used to declare variables which are used for saving FPU context, and it will store the necessary fpu registers into these variables, it needs to be used in an interrupt when in this interrupt fpu registers are used.

**Remark**

- It needs to be used together with [RESTORE\\_FPU\\_CONTEXT](#) (page 385)
- Don't use variable names `__fpu_context` in your ISR code
- If your ISR code will use fpu registers, and this interrupt is nested. Then you can do it like this:

```
void eclic_mtip_handler(void)
{
    // !!!Interrupt is enabled here!!!
    // !!!Higher priority interrupt could nest it!!!

    // Necessary only when you need to use fpu registers
    // in this isr handler functions
    SAVE_FPU_CONTEXT();

    // put your own interrupt handling code here

    // pair of SAVE_FPU_CONTEXT()
    RESTORE_FPU_CONTEXT();
}
```

**RESTORE\_FPU\_CONTEXT ( )**

Restore necessary fpu registers from variables for interrupt nesting.

This macro is used to restore necessary fpu registers from pre-defined variables in [SAVE\\_FPU\\_CONTEXT](#) (page 385) macro.

**Remark**

- It needs to be used together with [SAVE\\_FPU\\_CONTEXT](#) (page 385)

**Typedefs**

**typedef** uint64\_t **rv\_fpu\_t**

Type of FPU register, depends on the FLEN defined in RISC-V.

**2.5.12 PMP Functions**

Click [Nuclei PMP Unit](#)<sup>22</sup> to learn about Core PMP Unit in Nuclei ISA Spec.

```
__STATIC_INLINE uint8_t __get_PMPxCFG(uint32_t idx)
__STATIC_INLINE void __set_PMPxCFG(uint32_t idx, uint8_t pmpcfg)
__STATIC_INLINE rv_csr_t __get_PMPCFGx(uint32_t idx)
__STATIC_INLINE void __set_PMPCFGx(uint32_t idx, rv_csr_t pmpcfg)
__STATIC_INLINE rv_csr_t __get_PMPADDRx(uint32_t idx)
```

<sup>22</sup> [https://doc.nucleisys.com/nuclei\\_spec/isa/pmp.html](https://doc.nucleisys.com/nuclei_spec/isa/pmp.html)

```
__STATIC_INLINE void __set_PMPADDRx(uint32_t idx, rv_csr_t pmpaddr)
```

*group* **NMSIS\_Core\_PMP\_Functions**

Functions that related to the RISC-V Physical Memory Protection.

Optional physical memory protection (PMP) unit provides per-hart machine-mode control registers to allow physical memory access privileges (read, write, execute) to be specified for each physical memory region.

The PMP can supports region access control settings as small as four bytes.

## Functions

```
__STATIC_INLINE uint8_t __get_PMPxCFG(uint32_t idx)
```

Get 8bit PMPxCFG Register by PMP entry index.

Return the content of the PMPxCFG Register.

**Return** PMPxCFG Register value

**Parameters**

- [in] idx: PMP region index(0-15)

```
__STATIC_INLINE void __set_PMPxCFG(uint32_t idx, uint8_t pmpxcfg)
```

Set 8bit PMPxCFG by pmp entry index.

Set the given pmpxcfg value to the PMPxCFG Register.

**Parameters**

- [in] idx: PMPx region index(0-15)
- [in] pmpxcfg: PMPxCFG register value to set

```
__STATIC_INLINE rv_csr_t __get_PMPCFGx(uint32_t idx)
```

Get PMPCFGx Register by index.

Return the content of the PMPCFGx Register.

**Return** PMPCFGx Register value

**Remark**

- For RV64, only idx = 0 and idx = 2 is allowed. pmpcfg0 and pmpcfg2 hold the configurations for the 16 PMP entries, pmpcfg1 and pmpcfg3 are illegal
- For RV32, pmpcfg0–pmpcfg3, hold the configurations pmp0cfg–pmp15cfg for the 16 PMP entries

**Parameters**

- [in] idx: PMPCFG CSR index(0-3)

```
__STATIC_INLINE void __set_PMPCFGx(uint32_t idx, rv_csr_t pmpcfg)
```

Set PMPCFGx by index.

Write the given value to the PMPCFGx Register.

**Remark**

- For RV64, only idx = 0 and idx = 2 is allowed. pmpcfg0 and pmpcfg2 hold the configurations for the 16 PMP entries, pmpcfg1 and pmpcfg3 are illegal
- For RV32, pmpcfg0–pmpcfg3, hold the configurations pmp0cfg–pmp15cfg for the 16 PMP entries



**Parameters**

- [in] `idx`: PMPCFG CSR index(0-3)
- [in] `pmpcfg`: PMPCFGx Register value to set

**\_\_STATIC\_INLINE rv\_csr\_t \_\_get\_PMPADDRx(uint32\_t idx)**

Get PMPADDRx Register by index.

Return the content of the PMPADDRx Register.

**Return** PMPADDRx Register value

**Parameters**

- [in] `idx`: PMP region index(0-15)

**\_\_STATIC\_INLINE void \_\_set\_PMPADDRx(uint32\_t idx, rv\_csr\_t pmpaddr)**

Set PMPADDRx by index.

Write the given value to the PMPADDRx Register.

**Parameters**

- [in] `idx`: PMP region index(0-15)
- [in] `pmpaddr`: PMPADDRx Register value to set

## 2.5.13 Cache Functions

**General**

**enum NMSIS\_Core\_Cache::CCM\_OP\_FINFO\_Type**

*Values:*

**CCM\_OP\_SUCCESS** = 0x0

**CCM\_OP\_EXCEED\_ERR** = 0x1

**CCM\_OP\_PERM\_CHECK\_ERR** = 0x2

**CCM\_OP\_REFILL\_BUS\_ERR** = 0x3

**CCM\_OP\_ECC\_ERR** = 0x4

**enum NMSIS\_Core\_Cache::CCM\_CMD\_Type**

*Values:*

**CCM\_DC\_INVALID** = 0x0

**CCM\_DC\_WB** = 0x1

**CCM\_DC\_WBINVAL** = 0x2

**CCM\_DC\_LOCK** = 0x3

**CCM\_DC\_UNLOCK** = 0x4

**CCM\_DC\_WBINVAL\_ALL** = 0x6

**CCM\_DC\_WB\_ALL** = 0x7

**CCM\_DC\_INVALID\_ALL** = 0x17

**CCM\_IC\_INVALID** = 0x8

**CCM\_IC\_LOCK** = 0xb

```
CCM_IC_UNLOCK = 0xc
CCM_IC_INVALID_ALL = 0xd
__STATIC_FORCEINLINE void EnableSUCCM(void)
__STATIC_FORCEINLINE void DisableSUCCM(void)
__STATIC_FORCEINLINE void FlushPipeCCM(void)
CCM_SUEN_SUEN_Pos 0U
CCM_SUEN_SUEN_Msk (1UL << CCM_SUEN_SUEN_Pos)
```

#### *group* **NMSIS\_Core\_Cache**

Functions that configure Instruction and Data Cache.

Nuclei provide Cache Control and Maintenance (CCM) for software to control and maintain the internal L1 I/D Cache of the RISC-V Core, software can manage the cache flexibly to meet the actual application scenarios.

The CCM operations have 3 types: by single address, by all and flush pipeline. The CCM operations are done via CSR registers, M/S/U mode has its own CSR registers to do CCM operations. By default, CCM operations are not allowed in S/U mode, you can execute EnableSUCCM in M-Mode to enable it.

- API names started with M<operations>, such as MInvalICacheLine must be called in M-Mode only.
- API names started with S<operations>, such as SInvalICacheLine should be called in S-Mode.
- API names started with U<operations>, such as UInvalICacheLine should be called in U-Mode.

### Defines

```
CCM_SUEN_SUEN_Pos 0U
    CSR CCM_SUEN: SUEN bit Position.
CCM_SUEN_SUEN_Msk (1UL << CCM_SUEN_SUEN_Pos)
    CSR CCM_SUEN: SUEN Mask.
```

### Enums

#### **enum CCM\_OP\_FINFO\_Type**

Cache CCM Operation Fail Info.

*Values:*

```
CCM_OP_SUCCESS = 0x0
    Lock Succeed.
```

```
CCM_OP_EXCEED_ERR = 0x1
    Exceed the the number of lockable ways(N-Way I/D-Cache, lockable is N-1)
```

```
CCM_OP_PERM_CHECK_ERR = 0x2
    PMP/sPMP/Page-Table X(I-Cache)/R(D-Cache) permission check failed, or belong to Device/Non-
    Cacheable address range.
```

```
CCM_OP_REFILL_BUS_ERR = 0x3
    Refill has Bus Error.
```

```
CCM_OP_ECC_ERR = 0x4
    ECC Error.
```

**enum CCM\_CMD\_Type**

Cache CCM Command Types.

*Values:*

**CCM\_DC\_INVALID** = 0x0

Unlock and invalidate D-Cache line specified by CSR CCM\_XBEGINADDR.

**CCM\_DC\_WB** = 0x1

Flush the specific D-Cache line specified by CSR CCM\_XBEGINADDR.

**CCM\_DC\_WBINVAL** = 0x2

Unlock, flush and invalidate the specific D-Cache line specified by CSR CCM\_XBEGINADDR.

**CCM\_DC\_LOCK** = 0x3

Lock the specific D-Cache line specified by CSR CCM\_XBEGINADDR.

**CCM\_DC\_UNLOCK** = 0x4

Unlock the specific D-Cache line specified by CSR CCM\_XBEGINADDR.

**CCM\_DC\_WBINVAL\_ALL** = 0x6

Unlock and flush and invalidate all the valid and dirty D-Cache lines.

**CCM\_DC\_WB\_ALL** = 0x7

Flush all the valid and dirty D-Cache lines.

**CCM\_DC\_INVALID\_ALL** = 0x17

Unlock and invalidate all the D-Cache lines.

**CCM\_IC\_INVALID** = 0x8

Unlock and invalidate I-Cache line specified by CSR CCM\_XBEGINADDR.

**CCM\_IC\_LOCK** = 0xb

Lock the specific I-Cache line specified by CSR CCM\_XBEGINADDR.

**CCM\_IC\_UNLOCK** = 0xc

Unlock the specific I-Cache line specified by CSR CCM\_XBEGINADDR.

**CCM\_IC\_INVALID\_ALL** = 0xd

Unlock and invalidate all the I-Cache lines.

**Functions**

**\_\_STATIC\_FORCEINLINE void EnableSUCCM(void)**

Enable CCM operation in Supervisor/User Mode.

This function enable CCM operation in Supervisor/User Mode. If enabled, CCM operations in supervisor/user mode will be allowed.

**Remark**

- This function can be called in M-Mode only.

**See**

- DisableSUCCM

**\_\_STATIC\_FORCEINLINE void DisableSUCCM(void)**

Disable CCM operation in Supervisor/User Mode.

This function disable CCM operation in Supervisor/User Mode. If not enabled, CCM operations in supervisor/user mode will trigger a *illegal intruction* exception.

**Remark**

- This function can be called in M-Mode only.

See

- EnableSUCCM

**\_\_STATIC\_FORCEINLINE void FlushPipeCCM(void)**

Flush pipeline after CCM operation.

This function is used to flush pipeline after CCM operations on Cache, it will ensure latest instructions or data can be seen by pipeline.

**Remark**

- This function can be called in M/S/U-Mode only.

**struct CacheInfo\_Type**

*#include <core\_feature\_cache.h>* Cache Information Type.

## I-Cache Functions

**\_\_STATIC\_FORCEINLINE void EnableICache(void)**

**\_\_STATIC\_FORCEINLINE void DisableICache(void)**

**\_\_STATIC\_FORCEINLINE int32\_t GetICacheInfo(CacheInfo\_Type \* info)**

**\_\_STATIC\_FORCEINLINE int32\_t GetDCacheInfo(CacheInfo\_Type \* info)**

**\_\_STATIC\_FORCEINLINE void MInvalICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE void MInvalICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE void SInvalICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE void SInvalICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE void UInvalICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE void UInvalICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE unsigned long MLockICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE unsigned long MLockICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE unsigned long SLockICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE unsigned long SLockICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE unsigned long ULockICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE unsigned long ULockICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE void MUnlockICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE void MUnlockICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE void SUnlockICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE void SUnlockICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE void UUnlockICacheLine(unsigned long addr)**

**\_\_STATIC\_FORCEINLINE void UUnlockICacheLines(unsigned long addr, unsigned long cnt)**

**\_\_STATIC\_FORCEINLINE void MInvalICache(void)**

**\_\_STATIC\_FORCEINLINE void SInvalICache(void)**

---

```
__STATIC_FORCEINLINE void UInvalICache(void)
```

*group* **NMSIS\_Core\_ICache**

Functions that configure Instruction Cache.

## Functions

```
__STATIC_FORCEINLINE void EnableICache(void)
```

Enable ICache.

This function enable I-Cache

### Remark

- This function can be called in M-Mode only.
- This [CSR\\_MCACHE\\_CTL](#) (page 70) register control I Cache enable.

### See

- DisableICache

```
__STATIC_FORCEINLINE void DisableICache(void)
```

Disable ICache.

This function Disable I-Cache

### Remark

- This function can be called in M-Mode only.
- This [CSR\\_MCACHE\\_CTL](#) (page 70) register control I Cache enable.

### See

- EnableICache

```
__STATIC_FORCEINLINE int32_t GetICacheInfo(CacheInfo_Type * info)
```

Get I-Cache Information.

This function get I-Cache Information

### Remark

- This function can be called in M-Mode only.
- You can use this function in combination with cache lines operations

### See

- GetDCacheInfo

```
__STATIC_FORCEINLINE int32_t GetDCacheInfo(CacheInfo_Type * info)
```

Get D-Cache Information.

This function get D-Cache Information

### Remark

- This function can be called in M-Mode only.
- You can use this function in combination with cache lines operations

### See

- GetICacheInfo

**\_\_STATIC\_FORCEINLINE void MInvalidCacheLine(unsigned long addr)**

Invalidate one I-Cache line specified by address in M-Mode.

This function unlock and invalidate one I-Cache line specified by the address. Command CCM\_IC\_INVALID is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void MInvalidCacheLines(unsigned long addr, unsigned long cnt)**

Invalidate several I-Cache lines specified by address in M-Mode.

This function unlock and invalidate several I-Cache lines specified by the address and line count. Command CCM\_IC\_INVALID is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be invalidated
- [in] cnt: count of cache lines to be invalidated

**\_\_STATIC\_FORCEINLINE void SInvalidCacheLine(unsigned long addr)**

Invalidate one I-Cache line specified by address in S-Mode.

This function unlock and invalidate one I-Cache line specified by the address. Command CCM\_IC\_INVALID is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void SInvalidCacheLines(unsigned long addr, unsigned long cnt)**

Invalidate several I-Cache lines specified by address in S-Mode.

This function unlock and invalidate several I-Cache lines specified by the address and line count. Command CCM\_IC\_INVALID is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be invalidated
- [in] cnt: count of cache lines to be invalidated

**\_\_STATIC\_FORCEINLINE void UInvalidCacheLine(unsigned long addr)**

Invalidate one I-Cache line specified by address in U-Mode.

This function unlock and invalidate one I-Cache line specified by the address. Command CCM\_IC\_INVALID is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void UInvalidCacheLines(unsigned long addr, unsigned long cnt)**

Invalidate several I-Cache lines specified by address in U-Mode.

This function unlock and invalidate several I-Cache lines specified by the address and line count. Command CCM\_IC\_INVALID is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

#### Parameters

- [in] addr: start address to be invalidated
- [in] cnt: count of cache lines to be invalidated

**\_\_STATIC\_FORCEINLINE unsigned long MLockICacheLine(unsigned long addr)**

Lock one I-Cache line specified by address in M-Mode.

This function lock one I-Cache line specified by the address. Command CCM\_IC\_LOCK is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

#### Parameters

- [in] addr: start address to be locked

**\_\_STATIC\_FORCEINLINE unsigned long MLockICacheLines(unsigned long addr, unsigned long cnt)**

Lock several I-Cache lines specified by address in M-Mode.

This function lock several I-Cache lines specified by the address and line count. Command CCM\_IC\_LOCK is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

#### Parameters

- [in] addr: start address to be locked
- [in] cnt: count of cache lines to be locked

**\_\_STATIC\_FORCEINLINE unsigned long SLockICacheLine(unsigned long addr)**

Lock one I-Cache line specified by address in S-Mode.

This function lock one I-Cache line specified by the address. Command CCM\_IC\_LOCK is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

#### Parameters

- [in] addr: start address to be locked

**\_\_STATIC\_FORCEINLINE unsigned long SLockICacheLines(unsigned long addr, unsigned long cnt)**

Lock several I-Cache lines specified by address in S-Mode.

This function lock several I-Cache lines specified by the address and line count. Command CCM\_IC\_LOCK is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

#### Parameters

- [in] addr: start address to be locked

- [in] cnt: count of cache lines to be locked

**\_\_STATIC\_FORCEINLINE unsigned long ULockICacheLine(unsigned long addr)**

Lock one I-Cache line specified by address in U-Mode.

This function lock one I-Cache line specified by the address. Command CCM\_IC\_LOCK is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

**Parameters**

- [in] addr: start address to be locked

**\_\_STATIC\_FORCEINLINE unsigned long ULockICacheLines(unsigned long addr, unsigned long cnt)**

Lock several I-Cache lines specified by address in U-Mode.

This function lock several I-Cache lines specified by the address and line count. Command CCM\_IC\_LOCK is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

**Parameters**

- [in] addr: start address to be locked
- [in] cnt: count of cache lines to be locked

**\_\_STATIC\_FORCEINLINE void MUnlockICacheLine(unsigned long addr)**

Unlock one I-Cache line specified by address in M-Mode.

This function unlock one I-Cache line specified by the address. Command CCM\_IC\_UNLOCK is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be unlocked

**\_\_STATIC\_FORCEINLINE void MUnlockICacheLines(unsigned long addr, unsigned long cnt)**

Unlock several I-Cache lines specified by address in M-Mode.

This function unlock several I-Cache lines specified by the address and line count. Command CCM\_IC\_UNLOCK is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be unlocked
- [in] cnt: count of cache lines to be unlocked

**\_\_STATIC\_FORCEINLINE void SUnlockICacheLine(unsigned long addr)**

Unlock one I-Cache line specified by address in S-Mode.

This function unlock one I-Cache line specified by the address. Command CCM\_IC\_UNLOCK is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**



- [in] addr: start address to be unlocked

**\_\_STATIC\_FORCEINLINE void SUnlockICacheLines(unsigned long addr, unsigned long cnt)**  
Unlock several I-Cache lines specified by address in S-Mode.

This function unlock several I-Cache lines specified by the address and line count. Command CCM\_IC\_UNLOCK is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be unlocked
- [in] cnt: count of cache lines to be unlocked

**\_\_STATIC\_FORCEINLINE void UUnlockICacheLine(unsigned long addr)**  
Unlock one I-Cache line specified by address in U-Mode.

This function unlock one I-Cache line specified by the address. Command CCM\_IC\_UNLOCK is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be unlocked

**\_\_STATIC\_FORCEINLINE void UUnlockICacheLines(unsigned long addr, unsigned long cnt)**  
Unlock several I-Cache lines specified by address in U-Mode.

This function unlock several I-Cache lines specified by the address and line count. Command CCM\_IC\_UNLOCK is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be unlocked
- [in] cnt: count of cache lines to be unlocked

**\_\_STATIC\_FORCEINLINE void MInvalidICache(void)**  
Invalidate all I-Cache lines in M-Mode.

This function invalidate all I-Cache lines. Command CCM\_IC\_INVALID\_ALL is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void SInvalidICache(void)**  
Invalidate all I-Cache lines in S-Mode.

This function invalidate all I-Cache lines. Command CCM\_IC\_INVALID\_ALL is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

```
__STATIC_FORCEINLINE void UInvalICache(void)
```

Invalidate all I-Cache lines in U-Mode.

This function invalidate all I-Cache lines. Command CCM\_IC\_INVALID\_ALL is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

## D-Cache Functions

```
__STATIC_FORCEINLINE void EnabledDCache(void)
```

```
__STATIC_FORCEINLINE void DisabledDCache(void)
```

```
__STATIC_FORCEINLINE void MInvalDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void MInvalDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void SInvalDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void SInvalDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void UInvalDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void UInvalDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void MFlushDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void MFlushDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void SFlushDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void SFlushDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void UFlushDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void UFlushDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void MFlushInvalDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void MFlushInvalDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void SFlushInvalDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void SFlushInvalDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void UFlushInvalDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void UFlushInvalDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE unsigned long MLockDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE unsigned long MLockDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE unsigned long SLockDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE unsigned long SLockDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE unsigned long ULockDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE unsigned long ULockDCacheLines(unsigned long addr, unsigned long cnt)
```

```
__STATIC_FORCEINLINE void MUnlockDCacheLine(unsigned long addr)
```

```
__STATIC_FORCEINLINE void MUnlockDCacheLines(unsigned long addr, unsigned long cnt)
```

```

__STATIC_FORCEINLINE void SUnlockDCacheLine(unsigned long addr)
__STATIC_FORCEINLINE void SUnlockDCacheLines(unsigned long addr, unsigned long cnt)
__STATIC_FORCEINLINE void UUnlockDCacheLine(unsigned long addr)
__STATIC_FORCEINLINE void UUnlockDCacheLines(unsigned long addr, unsigned long cnt)
__STATIC_FORCEINLINE void MInvalDCache(void)
__STATIC_FORCEINLINE void SInvalDCache(void)
__STATIC_FORCEINLINE void UInvalDCache(void)
__STATIC_FORCEINLINE void MFlushDCache(void)
__STATIC_FORCEINLINE void SFlushDCache(void)
__STATIC_FORCEINLINE void UFlushDCache(void)
__STATIC_FORCEINLINE void MFlushInvalDCache(void)
__STATIC_FORCEINLINE void SFlushInvalDCache(void)
__STATIC_FORCEINLINE void UFlushInvalDCache(void)

```

*group* **NMSIS\_Core\_DCache**

Functions that configure Data Cache.

## Functions

```
__STATIC_FORCEINLINE void EnabledDCache(void)
```

Enable DCache.

This function enable D-Cache

### Remark

- This function can be called in M-Mode only.
- This [CSR\\_MCACHE\\_CTL](#) (page 70) register control D Cache enable.

### See

- DisabledDCache

```
__STATIC_FORCEINLINE void DisabledDCache(void)
```

Disable DCache.

This function Disable D-Cache

### Remark

- This function can be called in M-Mode only.
- This [CSR\\_MCACHE\\_CTL](#) (page 70) register control D Cache enable.

### See

- EnabledDCache

```
__STATIC_FORCEINLINE void MInvalDCacheLine(unsigned long addr)
```

Invalidate one D-Cache line specified by address in M-Mode.

This function unlock and invalidate one D-Cache line specified by the address. Command CCM\_DC\_INVALID is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void MInvalDCacheLines(unsigned long addr, unsigned long cnt)**

Invalidate several D-Cache lines specified by address in M-Mode.

This function unlock and invalidate several D-Cache lines specified by the address and line count. Command CCM\_DC\_INVALID is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be invalidated
- [in] cnt: count of cache lines to be invalidated

**\_\_STATIC\_FORCEINLINE void SInvalDCacheLine(unsigned long addr)**

Invalidate one D-Cache line specified by address in S-Mode.

This function unlock and invalidate one D-Cache line specified by the address. Command CCM\_DC\_INVALID is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void SInvalDCacheLines(unsigned long addr, unsigned long cnt)**

Invalidate several D-Cache lines specified by address in S-Mode.

This function unlock and invalidate several D-Cache lines specified by the address and line count. Command CCM\_DC\_INVALID is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be invalidated
- [in] cnt: count of cache lines to be invalidated

**\_\_STATIC\_FORCEINLINE void UInvalDCacheLine(unsigned long addr)**

Invalidate one D-Cache line specified by address in U-Mode.

This function unlock and invalidate one D-Cache line specified by the address. Command CCM\_DC\_INVALID is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void UInvalDCacheLines(unsigned long addr, unsigned long cnt)**

Invalidate several D-Cache lines specified by address in U-Mode.

This function unlock and invalidate several D-Cache lines specified by the address and line count. Command CCM\_DC\_INVALID is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

- [in] cnt: count of cache lines to be invalidated

**\_\_STATIC\_FORCEINLINE void MFlushDCacheLine(unsigned long addr)**

Flush one D-Cache line specified by address in M-Mode.

This function flush one D-Cache line specified by the address. Command CCM\_DC\_WB is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be flushed

**\_\_STATIC\_FORCEINLINE void MFlushDCacheLines(unsigned long addr, unsigned long cnt)**

Flush several D-Cache lines specified by address in M-Mode.

This function flush several D-Cache lines specified by the address and line count. Command CCM\_DC\_WB is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be flushed
- [in] cnt: count of cache lines to be flushed

**\_\_STATIC\_FORCEINLINE void SFlushDCacheLine(unsigned long addr)**

Flush one D-Cache line specified by address in S-Mode.

This function flush one D-Cache line specified by the address. Command CCM\_DC\_WB is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be flushed

**\_\_STATIC\_FORCEINLINE void SFlushDCacheLines(unsigned long addr, unsigned long cnt)**

Flush several D-Cache lines specified by address in S-Mode.

This function flush several D-Cache lines specified by the address and line count. Command CCM\_DC\_WB is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be flushed
- [in] cnt: count of cache lines to be flushed

**\_\_STATIC\_FORCEINLINE void UFlushDCacheLine(unsigned long addr)**

Flush one D-Cache line specified by address in U-Mode.

This function flush one D-Cache line specified by the address. Command CCM\_DC\_WB is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be flushed

**\_\_STATIC\_FORCEINLINE void UFlushDCacheLines(unsigned long addr, unsigned long cnt)**  
Flush several D-Cache lines specified by address in U-Mode.

This function flush several D-Cache lines specified by the address and line count. Command CCM\_DC\_WB is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be flushed
- [in] cnt: count of cache lines to be flushed

**\_\_STATIC\_FORCEINLINE void MFlushInvalDCacheLine(unsigned long addr)**  
Flush and invalidate one D-Cache line specified by address in M-Mode.

This function flush and invalidate one D-Cache line specified by the address. Command CCM\_DC\_WBINVAL is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be flushed and invalidated

**\_\_STATIC\_FORCEINLINE void MFlushInvalDCacheLines(unsigned long addr, unsigned long cnt)**  
Flush and invalidate several D-Cache lines specified by address in M-Mode.

This function flush and invalidate several D-Cache lines specified by the address and line count. Command CCM\_DC\_WBINVAL is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be flushed and invalidated
- [in] cnt: count of cache lines to be flushed and invalidated

**\_\_STATIC\_FORCEINLINE void SFlushInvalDCacheLine(unsigned long addr)**  
Flush and invalidate one D-Cache line specified by address in S-Mode.

This function flush and invalidate one D-Cache line specified by the address. Command CCM\_DC\_WBINVAL is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be flushed and invalidated

**\_\_STATIC\_FORCEINLINE void SFlushInvalDCacheLines(unsigned long addr, unsigned long cnt)**  
Flush and invalidate several D-Cache lines specified by address in S-Mode.

This function flush and invalidate several D-Cache lines specified by the address and line count. Command CCM\_DC\_WBINVAL is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be flushed and invalidated
- [in] cnt: count of cache lines to be flushed and invalidated

**\_\_STATIC\_FORCEINLINE void UFlushInvalDCacheLine(unsigned long addr)**

Flush and invalidate one D-Cache line specified by address in U-Mode.

This function flush and invalidate one D-Cache line specified by the address. Command CCM\_DC\_WBINVAL is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be flushed and invalidated

**\_\_STATIC\_FORCEINLINE void UFlushInvalDCacheLines(unsigned long addr, unsigned long cnt)**

Flush and invalidate several D-Cache lines specified by address in U-Mode.

This function flush and invalidate several D-Cache lines specified by the address and line count. Command CCM\_DC\_WBINVAL is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be flushed and invalidated
- [in] cnt: count of cache lines to be flushed and invalidated

**\_\_STATIC\_FORCEINLINE unsigned long MLockDCacheLine(unsigned long addr)**

Lock one D-Cache line specified by address in M-Mode.

This function lock one D-Cache line specified by the address. Command CCM\_DC\_LOCK is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

**Parameters**

- [in] addr: start address to be locked

**\_\_STATIC\_FORCEINLINE unsigned long MLockDCacheLines(unsigned long addr, unsigned long cnt)**

Lock several D-Cache lines specified by address in M-Mode.

This function lock several D-Cache lines specified by the address and line count. Command CCM\_DC\_LOCK is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

**Parameters**

- [in] addr: start address to be locked
- [in] cnt: count of cache lines to be locked

**\_\_STATIC\_FORCEINLINE unsigned long SLockDCacheLine(unsigned long addr)**

Lock one D-Cache line specified by address in S-Mode.

This function lock one D-Cache line specified by the address. Command CCM\_DC\_LOCK is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

**Parameters**

- [in] addr: start address to be locked

**\_\_STATIC\_FORCEINLINE unsigned long SLockDCacheLines(unsigned long addr, unsigned long cnt)**  
Lock several D-Cache lines specified by address in S-Mode.

This function lock several D-Cache lines specified by the address and line count. Command CCM\_DC\_LOCK is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

**Parameters**

- [in] addr: start address to be locked
- [in] cnt: count of cache lines to be locked

**\_\_STATIC\_FORCEINLINE unsigned long ULockDCacheLine(unsigned long addr)**  
Lock one D-Cache line specified by address in U-Mode.

This function lock one D-Cache line specified by the address. Command CCM\_DC\_LOCK is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

**Parameters**

- [in] addr: start address to be locked

**\_\_STATIC\_FORCEINLINE unsigned long ULockDCacheLines(unsigned long addr, unsigned long cnt)**  
Lock several D-Cache lines specified by address in U-Mode.

This function lock several D-Cache lines specified by the address and line count. Command CCM\_DC\_LOCK is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Return** result of CCM lock operation, see enum CCM\_OP\_FINFO

**Parameters**

- [in] addr: start address to be locked
- [in] cnt: count of cache lines to be locked

**\_\_STATIC\_FORCEINLINE void MUnlockDCacheLine(unsigned long addr)**  
Unlock one D-Cache line specified by address in M-Mode.

This function unlock one D-Cache line specified by the address. Command CCM\_DC\_UNLOCK is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be unlocked

**\_\_STATIC\_FORCEINLINE void MUnlockDCacheLines(unsigned long addr, unsigned long cnt)**  
Unlock several D-Cache lines specified by address in M-Mode.

This function unlock several D-Cache lines specified by the address and line count. Command CCM\_DC\_UNLOCK is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.



**Parameters**

- [in] addr: start address to be unlocked
- [in] cnt: count of cache lines to be unlocked

**\_\_STATIC\_FORCEINLINE void SUnlockDCacheLine(unsigned long addr)**

Unlock one D-Cache line specified by address in S-Mode.

This function unlock one D-Cache line specified by the address. Command CCM\_DC\_UNLOCK is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be unlocked

**\_\_STATIC\_FORCEINLINE void SUnlockDCacheLines(unsigned long addr, unsigned long cnt)**

Unlock several D-Cache lines specified by address in S-Mode.

This function unlock several D-Cache lines specified by the address and line count. Command CCM\_DC\_UNLOCK is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be unlocked
- [in] cnt: count of cache lines to be unlocked

**\_\_STATIC\_FORCEINLINE void UUnlockDCacheLine(unsigned long addr)**

Unlock one D-Cache line specified by address in U-Mode.

This function unlock one D-Cache line specified by the address. Command CCM\_DC\_UNLOCK is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be unlocked

**\_\_STATIC\_FORCEINLINE void UUnlockDCacheLines(unsigned long addr, unsigned long cnt)**

Unlock several D-Cache lines specified by address in U-Mode.

This function unlock several D-Cache lines specified by the address and line count. Command CCM\_DC\_UNLOCK is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be unlocked
- [in] cnt: count of cache lines to be unlocked

**\_\_STATIC\_FORCEINLINE void MInvalidDCache(void)**

Invalidate all D-Cache lines in M-Mode.

This function invalidate all D-Cache lines. Command CCM\_DC\_INVALID\_ALL is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void SInvalDCache(void)**

Invalidate all D-Cache lines in S-Mode.

This function invalidate all D-Cache lines. Command CCM\_DC\_INVAL\_ALL is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void UInvalDCache(void)**

Invalidate all D-Cache lines in U-Mode.

This function invalidate all D-Cache lines. In U-Mode, this operation will be automatically translated to flush and invalidate operations by hardware. Command CCM\_DC\_INVAL\_ALL is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be invalidated

**\_\_STATIC\_FORCEINLINE void MFlushDCache(void)**

Flush all D-Cache lines in M-Mode.

This function flush all D-Cache lines. Command CCM\_DC\_WB\_ALL is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be flushed

**\_\_STATIC\_FORCEINLINE void SFlushDCache(void)**

Flush all D-Cache lines in S-Mode.

This function flush all D-Cache lines. Command CCM\_DC\_WB\_ALL is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be flushed

**\_\_STATIC\_FORCEINLINE void UFlushDCache(void)**

Flush all D-Cache lines in U-Mode.

This function flush all D-Cache lines. Command CCM\_DC\_WB\_ALL is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be flushed

**\_\_STATIC\_FORCEINLINE void MFlushInvalDCache(void)**

Flush and invalidate all D-Cache lines in M-Mode.

This function flush and invalidate all D-Cache lines. Command CCM\_DC\_WBINVAL\_ALL is written to CSR [CSR\\_CCM\\_MCOMMAND](#) (page 70).

**Remark** This function must be executed in M-Mode only.

**Parameters**

- [in] addr: start address to be flushed and locked

**\_\_STATIC\_FORCEINLINE void SFlushInvalDCache(void)**

Flush and invalidate all D-Cache lines in S-Mode.

This function flush and invalidate all D-Cache lines. Command CCM\_DC\_WBINVAL\_ALL is written to CSR [CSR\\_CCM\\_SCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S-Mode only.

**Parameters**

- [in] addr: start address to be flushed and locked

**\_\_STATIC\_FORCEINLINE void UFlushInvalDCache(void)**

Flush and invalidate all D-Cache lines in U-Mode.

This function flush and invalidate all D-Cache lines. Command CCM\_DC\_WBINVAL\_ALL is written to CSR [CSR\\_CCM\\_UCOMMAND](#) (page 71).

**Remark** This function must be executed in M/S/U-Mode only.

**Parameters**

- [in] addr: start address to be flushed and locked

## 2.5.14 System Device Configuration

### group NMSIS\_Core\_SystemConfig

Functions for system init, clock setup and interrupt/exception/nmi functions available in system\_<device>.c.

Nuclei provides a template file **system\_Device.c** that must be adapted by the silicon vendor to match their actual device. As a **minimum requirement**, this file must provide:

- A device-specific system configuration function, [SystemInit](#) (page 406).
- A global variable that contains the system frequency, [SystemCoreClock](#) (page 406).
- A global eclic configuration initialization, [ECLIC\\_Init](#) (page 408).
- Global [\\_premain\\_init](#) (page 406) and [\\_postmain\\_fini](#) (page 406) functions called right before calling main function.
- Vendor customized interrupt, exception and nmi handling code, see [Interrupt and Exception and NMI Handling](#) (page 407)

The file configures the device and, typically, initializes the oscillator (PLL) that is part of the microcontroller device. This file might export other functions or variables that provide a more flexible configuration of the microcontroller system.

And this file also provided common interrupt, exception and NMI exception handling framework template, Silicon vendor can customize these template code as they want.

**Note** Please pay special attention to the static variable `SystemCoreClock`. This variable might be used throughout the whole system initialization and runtime to calculate frequency/time related values. Thus one must assure that the variable always reflects the actual system clock speed.

**Note** `_init` (page 406) function and `_fini` (page 406) function are now deprecated, but still need to provide a empty implementation here, due to newlibc still need it. Please put your pre- and post- main steps in `_premain_init` (page 406) and `_postmain_fini` (page 406)

**Attention** Be aware that a value stored to `SystemCoreClock` during low level initialization (i.e. `SystemInit()` (page 406)) might get overwritten by C library startup code and/or .bss section initialization. Thus its highly recommended to call `SystemCoreClockUpdate` (page 406) at the beginning of the user `main()` routine.

## Functions

void **SystemCoreClockUpdate** (void)

Function to update the variable `SystemCoreClock` (page 406).

Updates the variable `SystemCoreClock` (page 406) and must be called whenever the core clock is changed during program execution. The function evaluates the clock register settings and calculates the current core clock.

void **SystemInit** (void)

Function to Initialize the system.

Initializes the microcontroller system. Typically, this function configures the oscillator (PLL) that is part of the microcontroller device. For systems with a variable clock speed, it updates the variable `SystemCoreClock` (page 406). `SystemInit` is called from the file **startup**.

void **\_premain\_init** (void)

early init function before main

This function is executed right before main function. For RISC-V gnu toolchain, `_init` function might not be called by `__libc_init_array` function, so we defined a new function to do initialization

void **\_postmain\_fini** (int *status*)

finish function after main

This function is executed right after main function. For RISC-V gnu toolchain, `_fini` function might not be called by `__libc_fini_array` function, so we defined a new function to do initialization

### Parameters

- [in] `status`: status code return from main

void **\_init** (void)

`_init` function called in `__libc_init_array()`

This `__libc_init_array()` function is called during startup code, user need to implement this function, otherwise when link it will error `init.c:(.text.__libc_init_array+0x26): undefined reference to '_init'`

**Note** Please use `_premain_init` (page 406) function now

void **\_fini** (void)

`_fini` function called in `__libc_fini_array()`

This `__libc_fini_array()` function is called when exit main. user need to implement this function, otherwise when link it will error `fini.c:(.text.__libc_fini_array+0x28): undefined reference to '_fini'`

**Note** Please use `_postmain_fini` (page 406) function now

## Variables

uint32\_t **SystemCoreClock** = `SYSTEM_CLOCK`

Variable to hold the system core clock value.

Holds the system core clock, which is the system clock frequency supplied to the SysTick timer and the processor core clock. This variable can be used by debuggers to query the frequency of the debug timer or to configure the trace clock speed.

**Attention** Compilers must be configured to avoid removing this variable in case the application program is not using it. Debugging systems require the variable to be physically present in memory so that it can be examined to configure the debugger.

## Interrupt Exception NMI Handling

### *group* **NMSIS\_Core\_IntExcNMI\_Handling**

Functions for interrupt, exception and nmi handle available in `system_<device>.c`.

Nuclei provide a template for interrupt, exception and NMI handling. Silicon Vendor could adapt according to their requirement. Silicon vendor could implement interface for different exception code and replace current implementation.

### Defines

#### **MAX\_SYSTEM\_EXCEPTION\_NUM** 12

Max exception handler number, don't include the NMI(0xFFFF) one.

### Typedefs

**typedef** void (\***EXC\_HANDLER**) (unsigned long mcause, unsigned long sp)

Exception Handler Function Typedef.

**Note** This typedef is only used internal in this `system_<Device>.c` file. It is used to do type conversion for registered exception handler before calling it.

### Functions

**static** void **system\_default\_exception\_handler** (unsigned long mcause, unsigned long sp)

System Default Exception Handler.

This function provided a default exception and NMI handling code for all exception ids. By default, It will just print some information for debug, Vendor can customize it according to its requirements.

**static** void **Exception\_Init** (void)

Initialize all the default core exception handlers.

The core exception handler for each exception id will be initialized to *system\_default\_exception\_handler* (page 407).

**Note** Called in *\_init* (page 406) function, used to initialize default exception handlers for all exception IDs

void **Exception\_DumpFrame** (unsigned long sp)

Dump Exception Frame.

This function provided feature to dump exception frame stored in stack.

void **Exception\_Register\_EXC** (uint32\_t EXCn, unsigned long exc\_handler)

Register an exception handler for exception code EXCn.

- For `EXCn < MAX_SYSTEM_EXCEPTION_NUM` (page 407), it will be registered into `SystemExceptionHandlers[EXCn-1]`.
- For `EXCn == NMI_EXCn`, it will be registered into `SystemExceptionHandlers[MAX_SYSTEM_EXCEPTION_NUM]`.

**Parameters**

- `EXCn`: See `EXCn_Type`
- `exc_handler`: The exception handler for this exception code `EXCn`

unsigned long **Exception\_Get\_EXC** (uint32\_t *EXCn*)

Get current exception handler for exception code `EXCn`.

- For `EXCn < MAX_SYSTEM_EXCEPTION_NUM` (page 407), it will return `SystemExceptionHandlers[EXCn-1]`.
- For `EXCn == NMI_EXCn`, it will return `SystemExceptionHandlers[MAX_SYSTEM_EXCEPTION_NUM]`.

**Return** Current exception handler for exception code `EXCn`, if not found, return 0.

**Parameters**

- `EXCn`: See `EXCn_Type`

uint32\_t **core\_exception\_handler** (unsigned long *mcause*, unsigned long *sp*)

Common NMI and Exception handler entry.

This function provided a command entry for NMI and exception. Silicon Vendor could modify this template implementation according to requirement.

**Remark**

- RISC-V provided common entry for all types of exception. This is proposed code template for exception entry function, Silicon Vendor could modify the implementation.
- For the `core_exception_handler` template, we provided exception register function `Exception_Register_EXCn` which can help developer to register your exception handler for specific exception number.

void **ECLIC\_Init** (void)

Initialize Global ECLIC Config.

ECLIC needs be initialized after boot up, Vendor could also change the initialization configuration.

int32\_t **ECLIC\_Register\_IRQ** (IRQn\_Type *IRQn*, uint8\_t *shv*, ECLIC\_TRIGGER\_Type *trig\_mode*,  
uint8\_t *lvl*, uint8\_t *priority*, void *\*handler*)

Initialize a specific IRQ and register the handler.

This function set vector mode, trigger mode and polarity, interrupt level and priority, assign handler for specific `IRQn`.

**Return** -1 means invalid input parameter. 0 means successful.

**Remark**

- This function use to configure specific eclic interrupt and register its interrupt handler and enable its interrupt.
- If the vector table is placed in read-only section (FLASHXIP mode), handler could not be installed

**Parameters**

- [in] IRQn: NMI interrupt handler address
- [in] shv: *ECLIC\_NON\_VECTOR\_INTERRUPT* (page 90) means non-vector mode, and *ECLIC\_VECTOR\_INTERRUPT* (page 90) is vector mode
- [in] trig\_mode: see *ECLIC\_TRIGGER\_Type* (page 90)
- [in] lvl: interrupt level
- [in] priority: interrupt priority
- [in] handler: interrupt handler, if NULL, handler will not be installed

## Variables

unsigned long **SystemExceptionHandlers**[MAX\_SYSTEM\_EXCEPTION\_NUM + 1]  
Store the exception handlers for each exception ID.

### Note

- This SystemExceptionHandlers are used to store all the handlers for all the exception codes Nuclei N/NX core provided.
- Exception code 0 - 11, totally 12 exceptions are mapped to SystemExceptionHandlers[0:11]
- Exception for NMI is also re-routed to exception handling(exception code 0xFFFF) in startup code configuration, the handler itself is mapped to SystemExceptionHandlers[MAX\_SYSTEM\_EXCEPTION\_NUM]

## 2.5.15 ARM Compatible Functions

### group NMSIS\_Core\_ARMCompatible\_Functions

A few functions that compatible with ARM CMSIS-Core.

Here we provided a few functions that compatible with ARM CMSIS-Core, mostly used in the DSP and NN library.

### Defines

**\_\_ISB** () **\_\_RWMB**()  
Instruction Synchronization Barrier, compatible with ARM.

**\_\_DSB** () **\_\_RWMB**()  
Data Synchronization Barrier, compatible with ARM.

**\_\_DMB** () **\_\_RWMB**()  
Data Memory Barrier, compatible with ARM.

**\_\_LDRBT** (ptr) **\_\_LB**((ptr))  
LDRT Unprivileged (8 bit), ARM Compatible.

**\_\_LDRHT** (ptr) **\_\_LH**((ptr))  
LDRT Unprivileged (16 bit), ARM Compatible.

**\_\_LDRT** (ptr) **\_\_LW**((ptr))  
LDRT Unprivileged (32 bit), ARM Compatible.

**\_\_STRBT** (val, ptr) **\_\_SB**((ptr), (val))  
STRT Unprivileged (8 bit), ARM Compatible.

**\_\_STRHT** (val, ptr) **\_\_SH**((ptr), (val))  
STRT Unprivileged (16 bit), ARM Compatible.

**\_\_STRT** (val, ptr) **\_\_SW**((ptr), (val))  
STRT Unprivileged (32 bit), ARM Compatible.

**\_\_SSAT** (val, sat) **\_\_RV\_SCLIP32**((val), (sat-1))  
Signed Saturate.

Saturates a signed value.

**Return** Saturated value

**Parameters**

- [in] value: Value to be saturated
- [in] sat: Bit position to saturate to (1..32)

**\_\_USAT** (val, sat) **\_\_RV\_UCLIP32**((val), (sat))  
Unsigned Saturate.

Saturates an unsigned value.

**Return** Saturated value

**Parameters**

- [in] value: Value to be saturated
- [in] sat: Bit position to saturate to (0..31)

**\_\_RBIT** (value) **\_\_RV\_BITREVI**((value), 31)  
Reverse bit order of value.

Reverses the bit order of the given value.

**Return** Reversed value

**Parameters**

- [in] value: Value to reverse

**\_\_CLZ** (data) **\_\_RV\_CLZ32**(data)  
Count leading zeros.

Counts the number of leading zeros of a data value.

**Return** number of leading zeros in value

**Parameters**

- [in] data: Value to count the leading zeros

## Functions

**\_\_STATIC\_FORCEINLINE** uint32\_t **\_\_REV**(uint32\_t value)  
Reverse byte order (32 bit)

Reverses the byte order in unsigned integer value. For example, 0x12345678 becomes 0x78563412.

**Return** Reversed value

**Parameters**

- [in] value: Value to reverse



---

```
__STATIC_FORCEINLINE uint32_t __REV16(uint32_t value)
```

Reverse byte order (16 bit)

Reverses the byte order within each halfword of a word. For example, 0x12345678 becomes 0x34127856.

**Return** Reversed value

**Parameters**

- [in] value: Value to reverse

```
__STATIC_FORCEINLINE int16_t __REVSH(int16_t value)
```

Reverse byte order (16 bit)

Reverses the byte order in a 16-bit value and returns the signed 16-bit result. For example, 0x0080 becomes 0x8000.

**Return** Reversed value

**Parameters**

- [in] value: Value to reverse

```
__STATIC_FORCEINLINE uint32_t __ROR(uint32_t op1, uint32_t op2)
```

Rotate Right in unsigned value (32 bit)

Rotate Right (immediate) provides the value of the contents of a register rotated by a variable number of bits.

**Return** Rotated value

**Parameters**

- [in] op1: Value to rotate
- [in] op2: Number of Bits to rotate(0-31)



## 3.1 Overview

### 3.1.1 Introduction

This user manual describes the NMSIS DSP software library, a suite of common signal processing functions for use on Nuclei N/NX Class Processors based devices.

The library is divided into a number of functions each covering a specific category:

- Basic math functions
- Fast math functions
- Complex math functions
- Filters
- Matrix functions
- Transform functions
- Motor control functions
- Statistical functions
- Support functions
- Interpolation functions

The library has separate functions for operating on 8-bit integers, 16-bit integers, 32-bit integer and 32-bit floating-point values.

### 3.1.2 Using the Library

The library functions are declared in the public file `riscv_math.h` which is placed in the *NMSIS/DSP/Include* folder.

Simply include this file and link the appropriate library in the application and begin calling the library functions.

The Library supports single public header file `riscv_math.h` for Nuclei N/NX Class Processors cores with little endian. Same header file will be used for floating point unit(FPU) variants.

### 3.1.3 Examples

The library ships with *a number of examples* (page 416) which demonstrate how to use the library functions.

### 3.1.4 Toolchain Support

The library has been developed and tested with RISC-V GCC Toolchain.

The library is being tested in GCC toolchain and updates on this activity will be made available shortly.

### 3.1.5 Building the Library

The library installer contains a **Makefile** to rebuild libraries on Nuclei RISC-V GCC toolchain in the **NMSIS/** folder.

The libraries can be built by running `make gen_dsp_lib`, it will build and install DSP library into **Library/DSP/GCC** folder.

### 3.1.6 Preprocessor Macros

Each library project has different preprocessor macros.

**RISCV\_MATH\_MATRIX\_CHECK:** Define macro `RISCV_MATH_MATRIX_CHECK` for checking on the input and output sizes of matrices

**RISCV\_MATH\_ROUNDING:** Define macro `RISCV_MATH_ROUNDING` for rounding on support functions

**RISCV\_MATH\_LOOPUNROLL:** Define macro `RISCV_MATH_LOOPUNROLL` to enable manual loop unrolling in DSP functions

## 3.2 Using NMSIS-DSP

Here we will describe how to run the nmsis dsp examples in Nuclei Spike.

### 3.2.1 Preparation

- Nuclei Modified Spike - `xl_spike`
- Nuclei SDK modified for `xl_spike` branch `dev_xlspike`
- Nuclei RISC-V GNU Toolchain, `riscv-gcc 10` with vector extension is required.
- CMake `>= 3.5`

### 3.2.2 Tool Setup

1. Export **PATH** correctly for `xl_spike` and `riscv-nuclei-elf-gcc`

```
export PATH=/path/to/xl_spike/bin:/path/to/riscv-nuclei-elf-gcc/bin/:$PATH
```

### 3.2.3 Build NMSIS DSP Library

1. Download or clone NMSIS source code into **NMSIS** directory.
2. `cd` to `NMSIS/NMSIS/` directory
3. Build NMSIS DSP library using `make gen_dsp_lib`

4. Strip debug informations using `make strip_dsp_lib` to make the generated library smaller
5. The dsp library will be generated into `./Library/DSP/GCC` folder
6. The dsp libraries will be look like this:

```
$ ls -lh Library/DSP/GCC/
total 52M
-rw-rw-r-- 1 hqfang nucleisys 2.5M Aug  6 16:07 libnmsis_dsp_rv32imac.a
-rw-rw-r-- 1 hqfang nucleisys 3.2M Aug  6 16:07 libnmsis_dsp_rv32imacp.a
-rw-rw-r-- 1 hqfang nucleisys 2.3M Aug  6 16:07 libnmsis_dsp_rv32imaafc.a
-rw-rw-r-- 1 hqfang nucleisys 3.0M Aug  6 16:07 libnmsis_dsp_rv32imaafcp.a
-rw-rw-r-- 1 hqfang nucleisys 2.3M Aug  6 16:07 libnmsis_dsp_rv32imaafdc.a
-rw-rw-r-- 1 hqfang nucleisys 3.0M Aug  6 16:07 libnmsis_dsp_rv32imaafdcv.a
-rw-rw-r-- 1 hqfang nucleisys 3.1M Aug  6 16:07 libnmsis_dsp_rv64imac.a
-rw-rw-r-- 1 hqfang nucleisys 4.3M Aug  6 16:07 libnmsis_dsp_rv64imacp.a
-rw-rw-r-- 1 hqfang nucleisys 3.0M Aug  6 16:07 libnmsis_dsp_rv64imaafc.a
-rw-rw-r-- 1 hqfang nucleisys 4.0M Aug  6 16:07 libnmsis_dsp_rv64imaafcp.a
-rw-rw-r-- 1 hqfang nucleisys 3.9M Aug  6 16:08 libnmsis_dsp_rv64imaafcpv.a
-rw-rw-r-- 1 hqfang nucleisys 3.6M Aug  6 16:08 libnmsis_dsp_rv64imaafcv.a
-rw-rw-r-- 1 hqfang nucleisys 3.0M Aug  6 16:08 libnmsis_dsp_rv64imaafdc.a
-rw-rw-r-- 1 hqfang nucleisys 4.0M Aug  6 16:08 libnmsis_dsp_rv64imaafdcv.a
-rw-rw-r-- 1 hqfang nucleisys 3.9M Aug  6 16:08 libnmsis_dsp_rv64imaafdcvp.a
-rw-rw-r-- 1 hqfang nucleisys 3.6M Aug  6 16:08 libnmsis_dsp_rv64imaafdcv.a
```

7. library name with extra `p` is build with RISC-V DSP enabled.
  - `libnmsis_dsp_rv32imac.a`: Build for **RISCV\_ARCH=rv32imac** without DSP.
  - `libnmsis_dsp_rv32imacp.a`: Build for **RISCV\_ARCH=rv32imac** with DSP enabled.
8. library name with extra `v` is build with RISC-V Vector enabled, only valid for RISC-V 64bit processor.
  - `libnmsis_dsp_rv64imac.a`: Build for **RISCV\_ARCH=rv64imac** without Vector.
  - `libnmsis_dsp_rv64imacv.a`: Build for **RISCV\_ARCH=rv64imac** with Vector enabled.

---

#### Note:

- You can also directly build both DSP and NN library using `make gen`
  - You can strip the generated DSP and NN library using `make strip`
  - DSP and Vector extension can be combined, such as `p`, `v` and `pv`
  - Vector extension currently is only available with RISC-V 64 bit processor
- 

### 3.2.4 How to run

1. Set environment variables `NUCLEI_SDK_ROOT` and `NUCLEI_SDK_NMSIS`, and set Nuclei SDK SoC to *xl-spike*

```
export NUCLEI_SDK_ROOT=/path/to/nuclei_sdk
export NUCLEI_SDK_NMSIS=/path/to/NMSIS/NMSIS
export SOC=xlspike
```

2. Let us take `./riscv_class_marks_example/` for example
3. `cd ./riscv_class_marks_example/`

## 4. Run with RISC-V DSP enabled and Vector enabled NMSIS-DSP library for CORE ux600

```
# Clean project
make DSP_ENABLE=ON VECTOR_ENABLE=ON CORE=ux600 clean
# Build project
make DSP_ENABLE=ON VECTOR_ENABLE=ON CORE=ux600 all
# Run application using xl_spike
make DSP_ENABLE=ON VECTOR_ENABLE=ON CORE=ux600 run
```

## 5. Run with RISC-V DSP disabled and Vector disabled NMSIS-DSP library for CORE ux600

```
make DSP_ENABLE=OFF VECTOR_ENABLE=OFF CORE=ux600 clean
make DSP_ENABLE=OFF VECTOR_ENABLE=OFF CORE=ux600 all
make DSP_ENABLE=OFF VECTOR_ENABLE=OFF CORE=ux600 run
```

---

**Note:**

- You can easily run this example in your hardware, if you have enough memory to run it, just modify the SOC to the one you are using in step 1.
- 

## 3.3 NMSIS DSP API

If you want to access doxygen generated NMSIS DSP API, please click [NMSIS DSP API Doxygen Documentation](#).

### 3.3.1 Examples

#### Bayes Example

group **BayesExample**

**Description:**

Demonstrates the use of Bayesian classifier functions. It is complementing the tutorial about classical ML with NMSIS-DSP and python scikit-learn: <https://developer.arm.com/solutions/machine-learning-on-arm/developer-material/how-to-guides/implement-classical-ml-with-arm-nmsis-dsp-libraries>

#### Class Marks Example

group **ClassMarks**

**Refer** riscv\_class\_marks\_example\_f32.c

**Description:**

Demonstrates the use of the Maximum, Minimum, Mean, Standard Deviation, Variance and Matrix functions to calculate statistical values of marks obtained in a class.

**Note** This example also demonstrates the usage of static initialization.

**Variables Description:**

- testMarks\_f32 points to the marks scored by 20 students in 4 subjects
- max\_marks Maximum of all marks

- `min_marks` Minimum of all marks
- `mean` Mean of all marks
- `var` Variance of the marks
- `std` Standard deviation of the marks
- `numStudents` Total number of students in the class

#### NMSIS DSP Software Library Functions Used:

- `riscv_mat_init_f32()`
- `riscv_mat_mult_f32()`
- `riscv_max_f32()`
- `riscv_min_f32()`
- `riscv_mean_f32()`
- `riscv_std_f32()`
- `riscv_var_f32()`

### Convolution Example

#### *group* ConvolutionExample

**Refer** `riscv_convolution_example_f32.c`

#### **Description:**

Demonstrates the convolution theorem with the use of the Complex FFT, Complex-by-Complex Multiplication, and Support Functions.

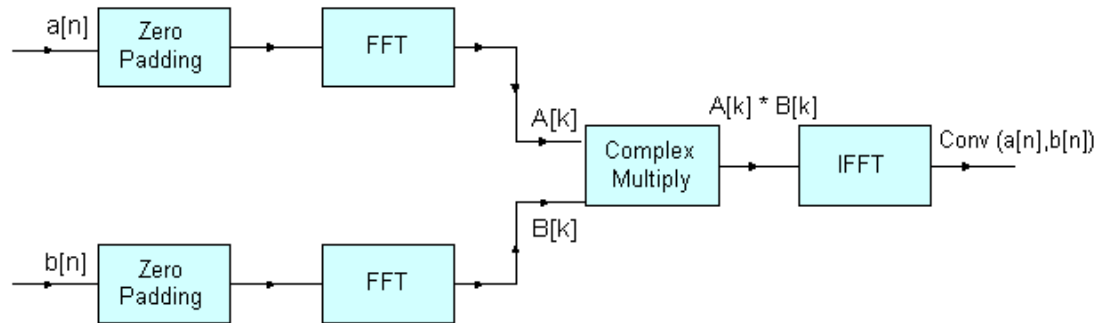
#### **Algorithm:**

The convolution theorem states that convolution in the time domain corresponds to multiplication in the frequency domain. Therefore, the Fourier transform of the convolution of two signals is equal to the product of their individual Fourier transforms. The Fourier transform of a signal can be evaluated efficiently using the Fast Fourier Transform (FFT).

Two input signals,  $a[n]$  and  $b[n]$ , with lengths  $n_1$  and  $n_2$  respectively, are zero padded so that their lengths become  $N$ , which is greater than or equal to  $(n_1+n_2-1)$  and is a power of 4 as FFT implementation is radix-4. The convolution of  $a[n]$  and  $b[n]$  is obtained by taking the FFT of the input signals, multiplying the Fourier transforms of the two signals, and taking the inverse FFT of the multiplied result.

This is denoted by the following equations: where  $A[k]$  and  $B[k]$  are the  $N$ -point FFTs of the signals  $a[n]$  and  $b[n]$  respectively. The length of the convolved signal is  $(n_1+n_2-1)$ .

#### **Block Diagram:**

**Variables Description:**

- testInputA\_f32 points to the first input sequence
- srcALen length of the first input sequence
- testInputB\_f32 points to the second input sequence
- srcBLen length of the second input sequence
- outLen length of convolution output sequence,  $(srcALen + srcBLen - 1)$
- AxB points to the output array where the product of individual FFTs of inputs is stored.

**NMSIS DSP Software Library Functions Used:**

- riscv\_fill\_f32()
- riscv\_copy\_f32()
- riscv\_cfft\_radix4\_init\_f32()
- riscv\_cfft\_radix4\_f32()
- riscv\_cmplx\_mult\_cmplx\_f32()

**Dot Product Example**

*group* **DotproductExample**

**Refer** riscv\_dotproduct\_example\_f32.c

**Description:**

Demonstrates the use of the Multiply and Add functions to perform the dot product. The dot product of two vectors is obtained by multiplying corresponding elements and summing the products.

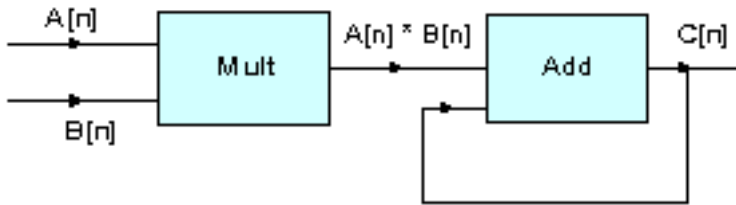
**Algorithm:**

The two input vectors A and B with length n, are multiplied element-by-element and then added to obtain dot product.

This is denoted by the following equation:

**Block Diagram:**



**Variables Description:**

- srcA\_buf\_f32 points to first input vector
- srcB\_buf\_f32 points to second input vector
- testOutput stores dot product of the two input vectors.

**NMSIS DSP Software Library Functions Used:**

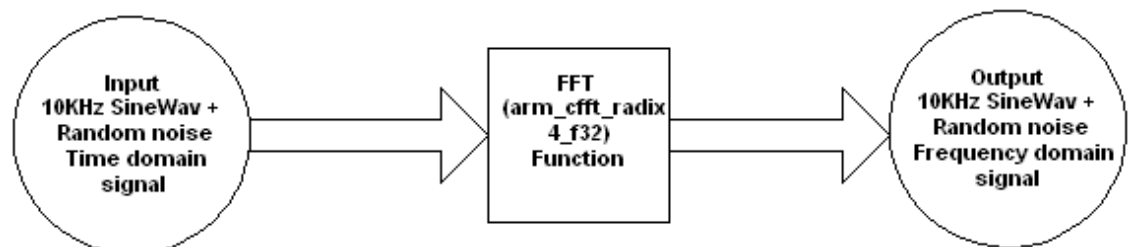
- riscv\_mult\_f32()
- riscv\_add\_f32()

**Frequency Bin Example***group* **FrequencyBin****Refer** riscv\_fft\_bin\_example\_f32.c**Description**

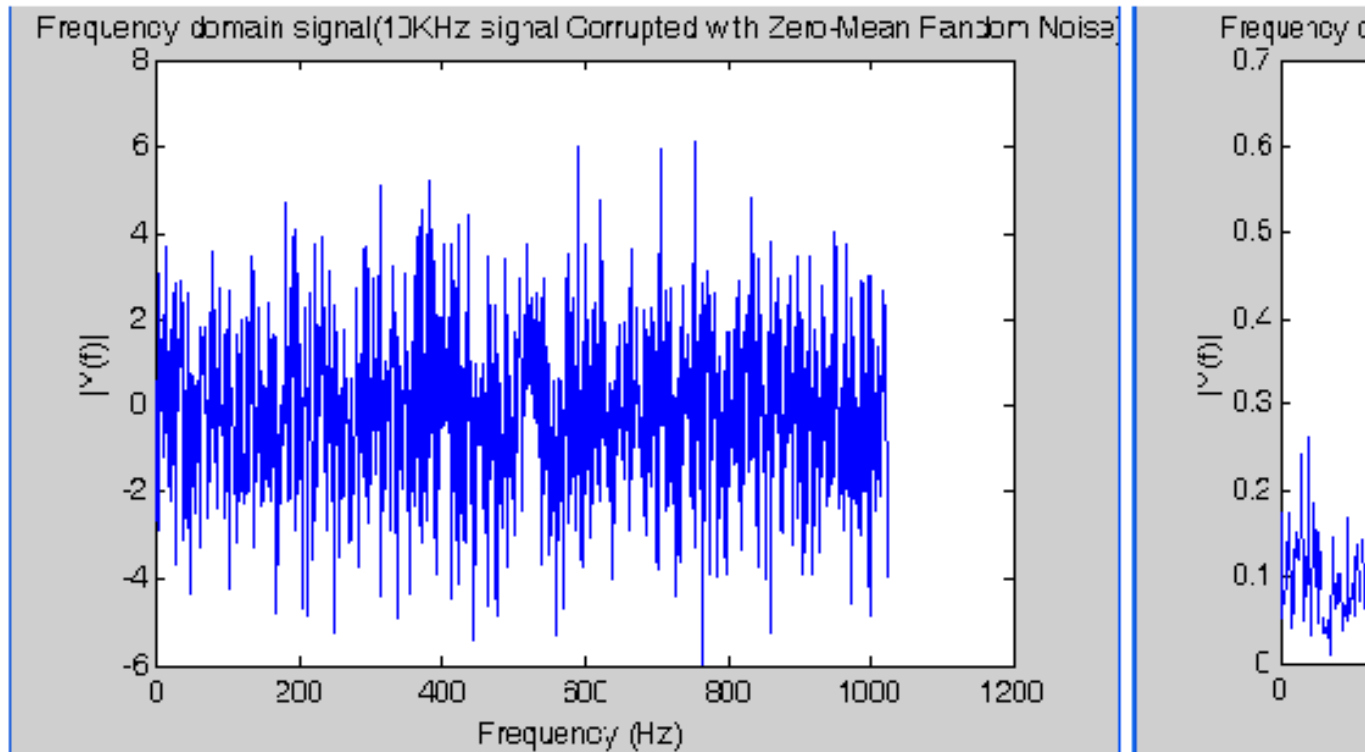
Demonstrates the calculation of the maximum energy bin in the frequency domain of the input signal with the use of Complex FFT, Complex Magnitude, and Maximum functions.

**Algorithm:**

The input test signal contains a 10 kHz signal with uniformly distributed white noise. Calculating the FFT of the input signal will give us the maximum energy of the bin corresponding to the input frequency of 10 kHz.

**Block Diagram:**

The figure below shows the time domain signal of 10 kHz signal with uniformly distributed white noise, and the next figure shows the input in the frequency domain. The bin with maximum energy corresponds to 10 kHz signal.



#### Variables Description:

- `testInput_f32_10khz` points to the input data
- `testOutput` points to the output data
- `fftSize` length of FFT
- `ifftFlag` flag for the selection of CFFT/CIFFT
- `doBitReverse` Flag for selection of normal order or bit reversed order
- `refIndex` reference index value at which maximum energy of bin occurs
- `testIndex` calculated index value at which maximum energy of bin occurs

#### NMSIS DSP Software Library Functions Used:

- `riscv_cfft_f32()`
- `riscv_cmplx_mag_f32()`
- `riscv_max_f32()`

#### FIR Lowpass Filter Example

*group* **FIRLPF**

**Refer** `riscv_fir_example_f32.c`

#### Description:

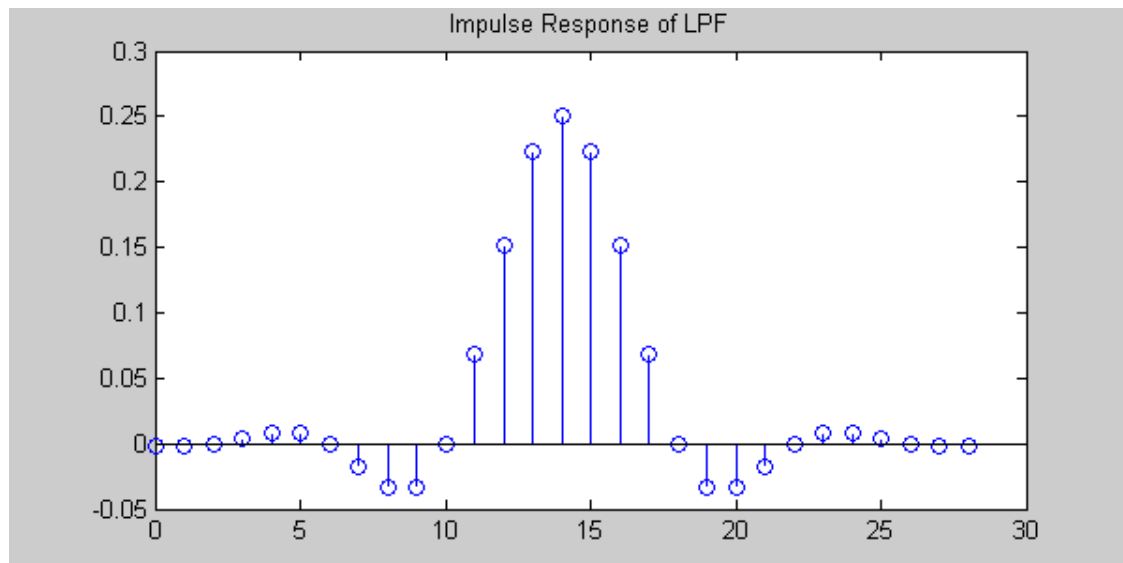
Removes high frequency signal components from the input using an FIR lowpass filter. The example demonstrates how to configure an FIR filter and then pass data through it in a block-by-block fashion.



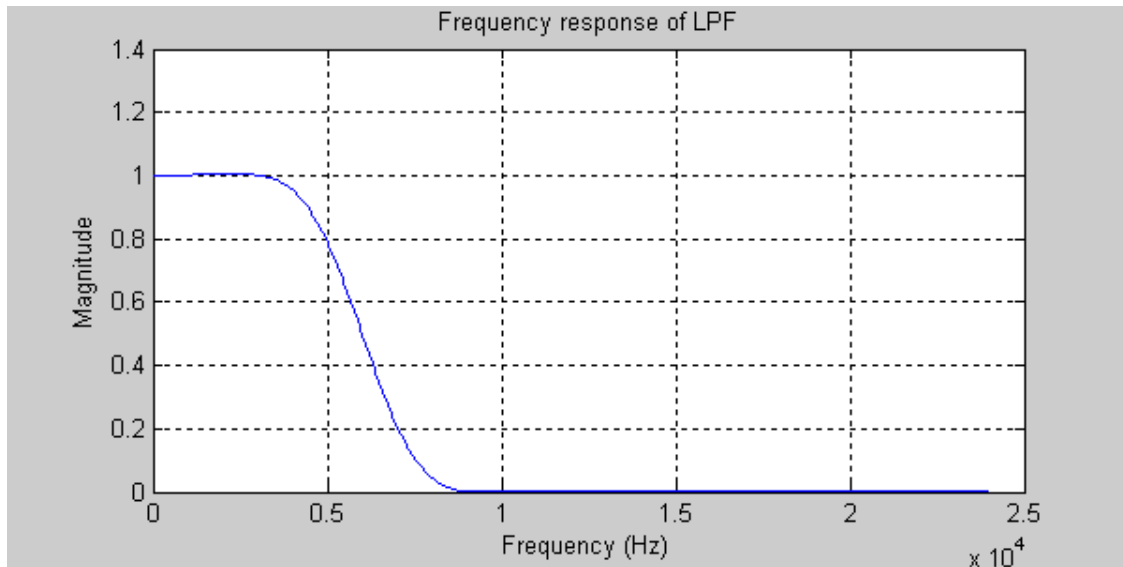
#### Algorithm:

The input signal is a sum of two sine waves: 1 kHz and 15 kHz. This is processed by an FIR lowpass filter with cutoff frequency 6 kHz. The lowpass filter eliminates the 15 kHz signal leaving only the 1 kHz sine wave at the output.

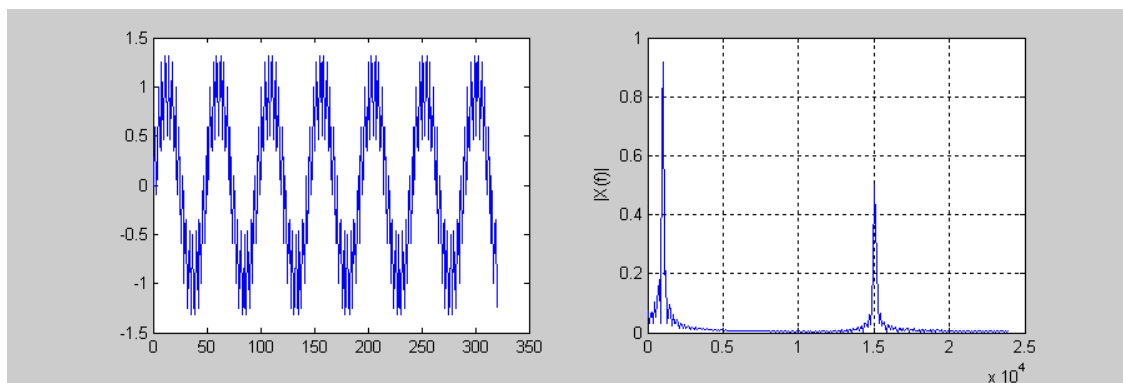
The lowpass filter was designed using MATLAB with a sample rate of 48 kHz and a length of 29 points. The MATLAB code to generate the filter coefficients is shown below: The first argument is the “order” of the filter and is always one less than the desired length. The second argument is the normalized cutoff frequency. This is in the range 0 (DC) to 1.0 (Nyquist). A 6 kHz cutoff with a Nyquist frequency of 24 kHz lies at a normalized frequency of  $6/24 = 0.25$ . The NMSIS FIR filter function requires the coefficients to be in time reversed order. The resulting filter coefficients are shown below. Note that the filter is symmetric (a property of linear phase FIR filters) and the point of symmetry is sample 14. Thus the filter will have a delay of 14 samples for all frequencies.



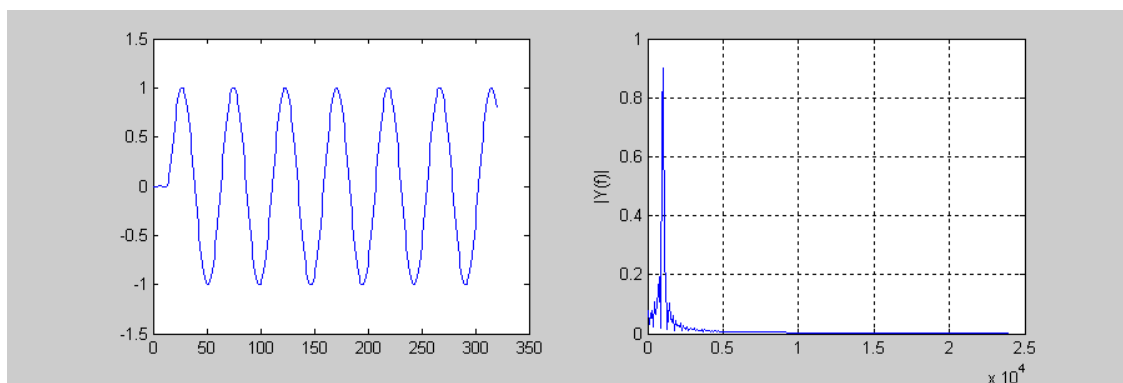
The frequency response of the filter is shown next. The passband gain of the filter is 1.0 and it reaches 0.5 at the cutoff frequency 6 kHz.



The input signal is shown below. The left hand side shows the signal in the time domain while the right hand side is a frequency domain representation. The two sine wave components can be clearly seen.



The output of the filter is shown below. The 15 kHz component has been eliminated.



#### Variables Description:

- `testInput_f32_1kHz_15kHz` points to the input data
- `refOutput` points to the reference output data
- `testOutput` points to the test output data

- `firStateF32` points to state buffer
- `firCoeffs32` points to coefficient buffer
- `blockSize` number of samples processed at a time
- `numBlocks` number of frames

#### NMSIS DSP Software Library Functions Used:

- `riscv_fir_init_f32()`
- `riscv_fir_f32()`

### Graphic Audio Equalizer Example

group **GEQ5Band**

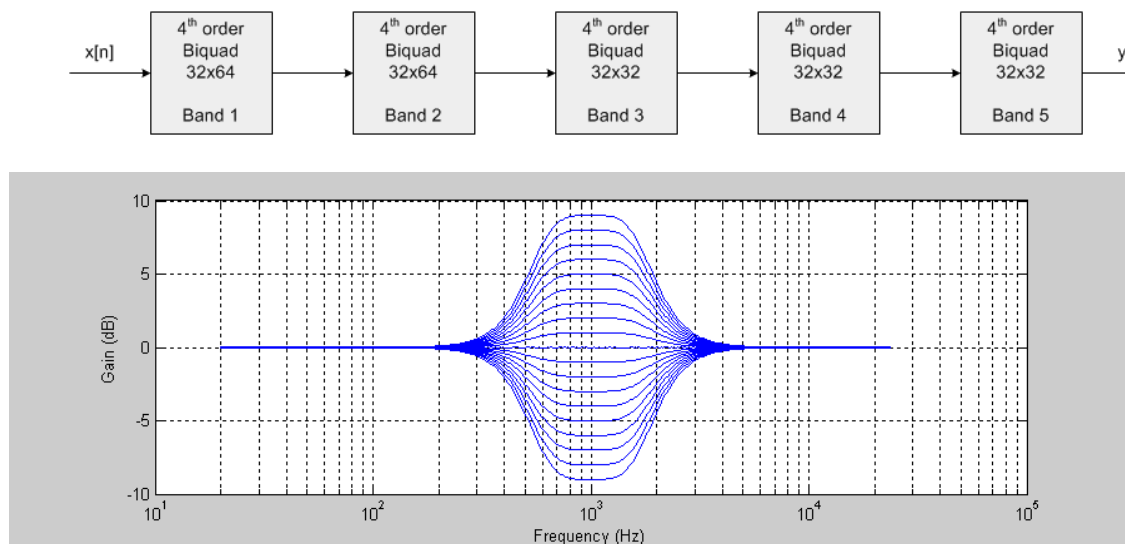
Refer `riscv_graphic_equalizer_example_q31.c`

#### Description:

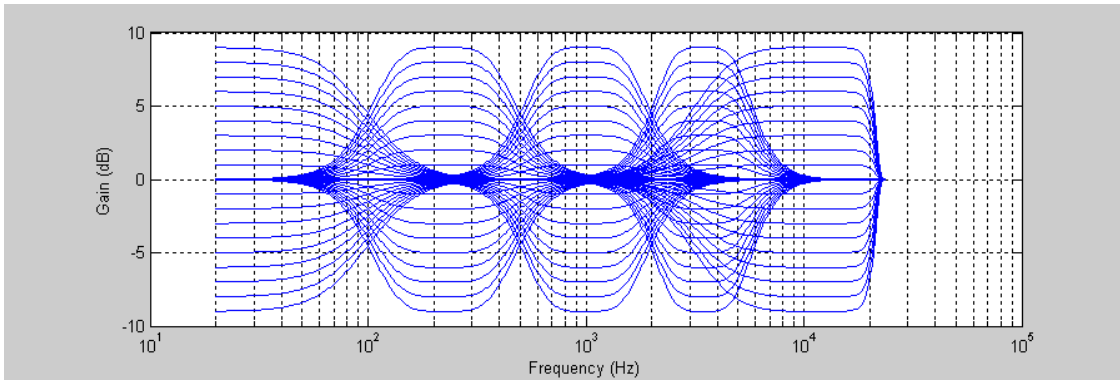
This example demonstrates how a 5-band graphic equalizer can be constructed using the Biquad cascade functions. A graphic equalizer is used in audio applications to vary the tonal quality of the audio.

#### Block Diagram:

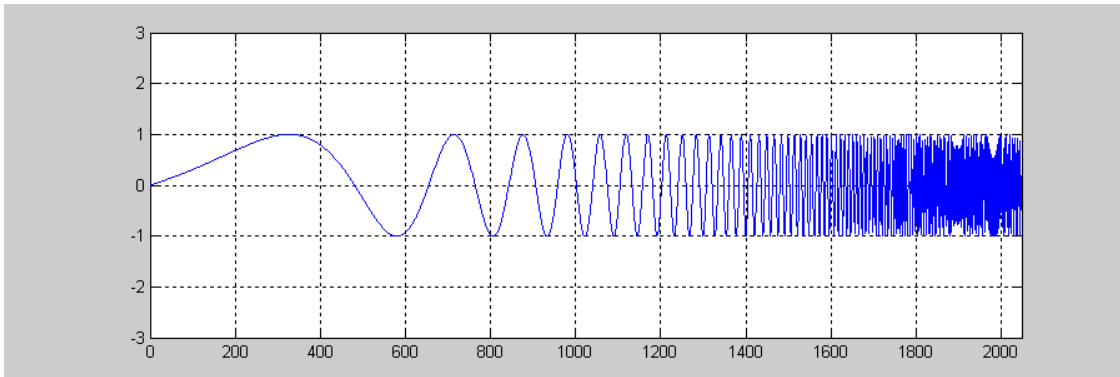
The design is based on a cascade of 5 filter sections. Each filter section is 4th order and consists of a cascade of two Biquads. Each filter has a nominal gain of 0 dB (1.0 in linear units) and boosts or cuts signals within a specific frequency range. The edge frequencies between the 5 bands are 100, 500, 2000, and 6000 Hz. Each band has an adjustable boost or cut in the range of +/- 9 dB. For example, the band that extends from 500 to 2000 Hz has the response shown below:



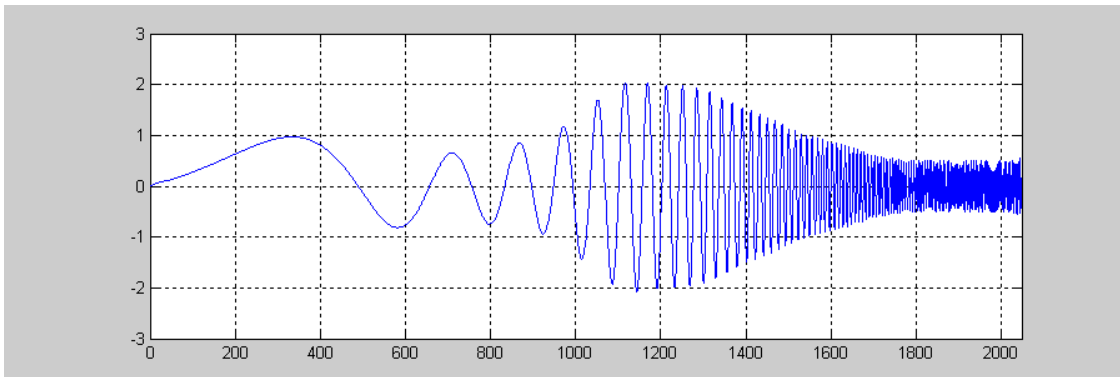
With 1 dB steps, each filter has a total of 19 different settings. The filter coefficients for all possible 19 settings were precomputed in MATLAB and stored in a table. With 5 different tables, there are a total of  $5 \times 19 = 95$  different 4th order filters. All 95 responses are shown below:



Each 4th order filter has 10 coefficients for a grand total of 950 different filter coefficients that must be tabulated. The input and output data is in Q31 format. For better noise performance, the two low frequency bands are implemented using the high precision 32x64-bit Biquad filters. The remaining 3 high frequency bands use standard 32x32-bit Biquad filters. The input signal used in the example is a logarithmic chirp.



The array `bandGains` specifies the gain in dB to apply in each band. For example, if `bandGains={0, -3, 6, 4, -6}`; then the output signal will be:



**Note** The output chirp signal follows the gain or boost of each band.

#### Variables Description:

- `testInput_f32` points to the input data

- testRefOutput\_f32 points to the reference output data
- testOutput points to the test output data
- inputQ31 temporary input buffer
- outputQ31 temporary output buffer
- biquadStateBand1Q31 points to state buffer for band1
- biquadStateBand2Q31 points to state buffer for band2
- biquadStateBand3Q31 points to state buffer for band3
- biquadStateBand4Q31 points to state buffer for band4
- biquadStateBand5Q31 points to state buffer for band5
- coeffTable points to coefficient buffer for all bands
- gainDB gain buffer which has gains applied for all the bands

#### NMSIS DSP Software Library Functions Used:

- riscv\_biquad\_cas\_df1\_32x64\_init\_q31()
- riscv\_biquad\_cas\_df1\_32x64\_q31()
- riscv\_biquad\_cascade\_df1\_init\_q31()
- riscv\_biquad\_cascade\_df1\_q31()
- riscv\_scale\_q31()
- riscv\_scale\_f32()
- riscv\_float\_to\_q31()
- riscv\_q31\_to\_float()

### Linear Interpolate Example

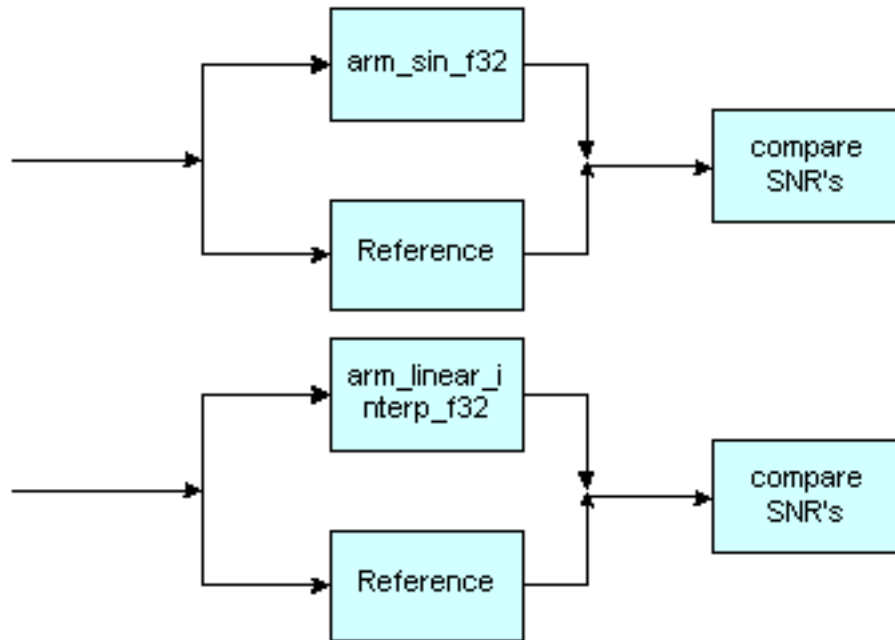
#### *group* **LinearInterpExample**

##### **NMSIS DSP Software Library Linear Interpolate Example**

**Description** This example demonstrates usage of linear interpolate modules and fast math modules. Method 1 uses fast math sine function to calculate sine values using cubic interpolation and method 2 uses linear interpolation function and results are compared to reference output. Example shows linear interpolation function can be used to get higher precision compared to fast math sin calculation.

**Refer** riscv\_linear\_interp\_example\_f32.c

**Block Diagram:**

**Variables Description:**

- testInputSin\_f32 points to the input values for sine calculation
- testRefSinOutput32\_f32 points to the reference values caculated from sin() matlab function
- testOutput points to output buffer calculation from cubic interpolation
- testLinIntOutput points to output buffer calculation from linear interpolation
- snr1 Signal to noise ratio for reference and cubic interpolation output
- snr2 Signal to noise ratio for reference and linear interpolation output

**NMSIS DSP Software Library Functions Used:**

- riscv\_sin\_f32()
- riscv\_linear\_interp\_f32()

**Matrix Example**

*group* **MatrixExample**

**Refer** riscv\_matrix\_example\_f32.c

**Description:**

Demonstrates the use of Matrix Transpose, Matrix Mulplication, and Matrix Inverse functions to apply least squares fitting to input data. Least squares fitting is the procedure for finding the best-fitting curve that minimizes the sum of the squares of the offsets (least square error) from a given set of data.

**Algorithm:**

The linear combination of parameters considered is as follows:

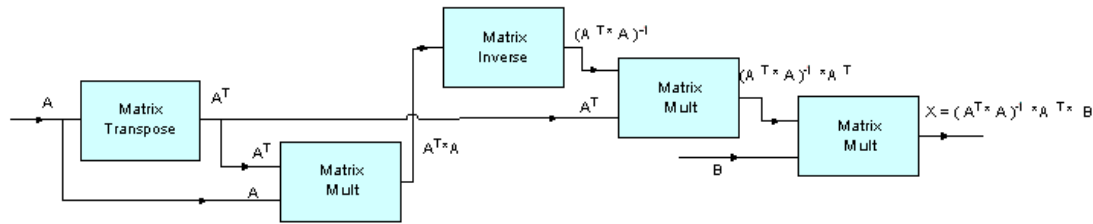
$A * X = B$ , where X is the unknown value and can be estimated from A & B.



The least squares estimate  $X$  is given by the following equation:

$$X = \text{Inverse}(A^T * A) * A^T * B$$

**Block Diagram:**



**Variables Description:**

- $A\_f32$  input matrix in the linear combination equation
- $B\_f32$  output matrix in the linear combination equation
- $X\_f32$  unknown matrix estimated using  $A\_f32$  &  $B\_f32$  matrices

**NMSIS DSP Software Library Functions Used:**

- `riscv_mat_init_f32()`
- `riscv_mat_trans_f32()`
- `riscv_mat_mult_f32()`
- `riscv_mat_inverse_f32()`

## Signal Convergence Example

*group* **SignalConvergence**

**Refer** `riscv_signal_converge_example_f32.c`

**Description:**

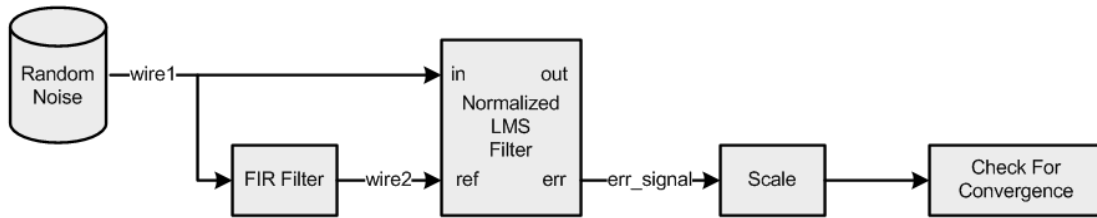
Demonstrates the ability of an adaptive filter to “learn” the transfer function of a FIR lowpass filter using the Normalized LMS Filter, Finite Impulse Response (FIR) Filter, and Basic Math Functions.

**Algorithm:**

The figure below illustrates the signal flow in this example. Uniformly distributed white noise is passed through an FIR lowpass filter. The output of the FIR filter serves as the reference input of the adaptive filter (normalized LMS filter). The white noise is input to the adaptive filter. The adaptive filter learns the transfer function of the FIR filter. The filter outputs two signals: (1) the output of the internal adaptive FIR filter, and (2) the error signal which is the difference between the adaptive filter and the reference output of the FIR filter. Over time as the adaptive filter learns the transfer function of the FIR filter, the first output approaches the reference output of the FIR filter, and the error signal approaches zero.

The adaptive filter converges properly even if the input signal has a large dynamic range (i.e., varies from small to large values). The coefficients of the adaptive filter are initially zero, and then converge over 1536 samples. The internal function `test_signal_converge()` implements the stopping condition. The function checks if all of the values of the error signal have a magnitude below a threshold DELTA.

**Block Diagram:**

**Variables Description:**

- testInput\_f32 points to the input data
- firStateF32 points to FIR state buffer
- lmsStateF32 points to Normalised Least mean square FIR filter state buffer
- FIRCoeff\_f32 points to coefficient buffer
- lmsNormCoeff\_f32 points to Normalised Least mean square FIR filter coefficient buffer
- wire1, wire2, wire3 temporary buffers
- errOutput, err\_signal temporary error buffers

**NMSIS DSP Software Library Functions Used:**

- riscv\_lms\_norm\_init\_f32()
- riscv\_fir\_init\_f32()
- riscv\_fir\_f32()
- riscv\_lms\_norm\_f32()
- riscv\_scale\_f32()
- riscv\_abs\_f32()
- riscv\_sub\_f32()
- riscv\_min\_f32()
- riscv\_copy\_f32()

**SineCosine Example**

*group* **SinCosExample**

**Refer** riscv\_sin\_cos\_example\_f32.c

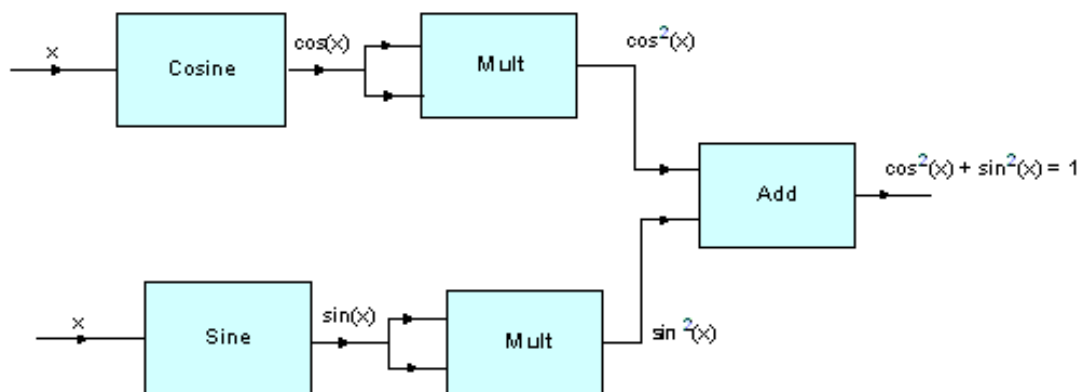
**Description:**

Demonstrates the Pythagorean trigonometric identity with the use of Cosine, Sine, Vector Multiplication, and Vector Addition functions.

**Algorithm:**

Mathematically, the Pythagorean trigonometric identity is defined by the following equation: where  $x$  is the angle in radians.

**Block Diagram:**

**Variables Description:**

- testInput\_f32 array of input angle in radians
- testOutput stores sum of the squares of sine and cosine values of input angle

**NMSIS DSP Software Library Functions Used:**

- riscv\_cos\_f32()
- riscv\_sin\_f32()
- riscv\_mult\_f32()
- riscv\_add\_f32()

**SVM Example**

*group* **SVMExample**

**Description:**

Demonstrates the use of SVM functions. It is complementing the tutorial about classical ML with NMSIS-DSP and python scikit-learn: <https://developer.arm.com/solutions/machine-learning-on-arm/developer-material/how-to-guides/implement-classical-ml-with-arm-nmsis-dsp-libraries>

**Variance Example**

*group* **VarianceExample**

**Refer** riscv\_variance\_example\_f32.c

**Description:**

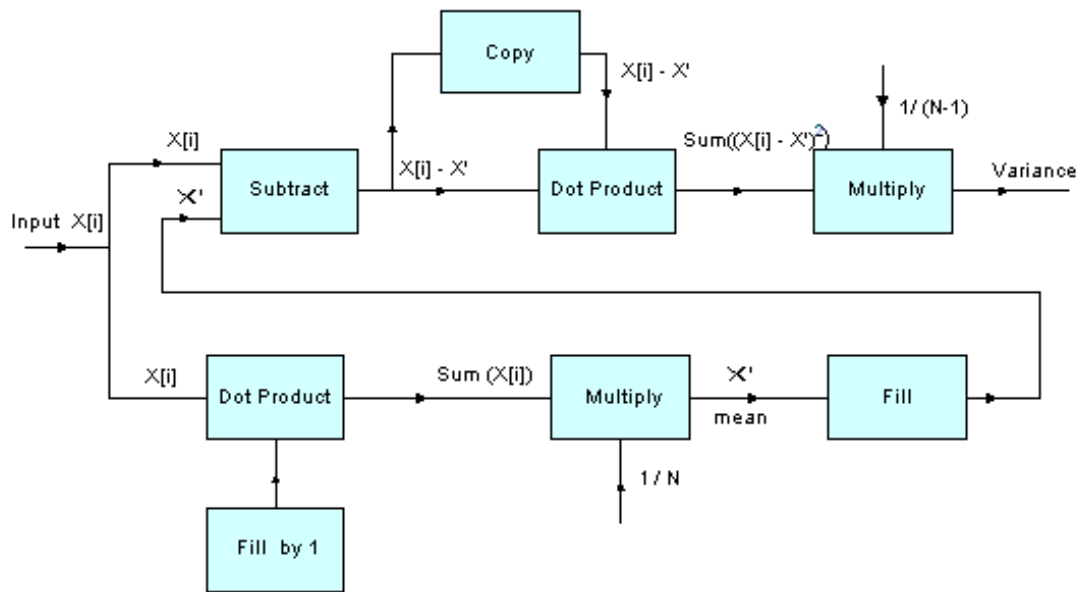
Demonstrates the use of Basic Math and Support Functions to calculate the variance of an input sequence with N samples. Uniformly distributed white noise is taken as input.

**Algorithm:**

The variance of a sequence is the mean of the squared deviation of the sequence from its mean.

This is denoted by the following equation: where,  $x[n]$  is the input sequence, N is the number of input samples, and  $\bar{x}$  is the mean value of the input sequence,  $x[n]$ .

The mean value  $\bar{x}$  is defined as:

**Block Diagram:****Variables Description:**

- testInput\_f32 points to the input data
- wire1, wir2, wire3 temporary buffers
- blockSize number of samples processed at a time
- refVarianceOut reference variance value

**NMSIS DSP Software Library Functions Used:**

- riscv\_dot\_prod\_f32()
- riscv\_mult\_f32()
- riscv\_sub\_f32()
- riscv\_fill\_f32()
- riscv\_copy\_f32()

*group* **groupExamples**

### 3.3.2 Basic Math Functions

#### Vector Absolute Value

void **riscv\_abs\_f16** (const float16\_t \*pSrc, float16\_t \*pDst, uint32\_t blockSize)

void **riscv\_abs\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_abs\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_abs\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_abs\_q7** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

*group* **BasicAbs**

Computes the absolute value of a vector on an element-by-element basis.

The functions support in-place computation allowing the source and destination pointers to reference the same memory buffer. There are separate functions for floating-point, Q7, Q15, and Q31 data types.

**Functions**

void **riscv\_abs\_f16** (**const** float16\_t \*pSrc, float16\_t \*pDst, uint32\_t blockSize)  
Floating-point vector absolute value.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_abs\_f32** (**const** float32\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)  
Floating-point vector absolute value.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_abs\_q15** (**const** q15\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)  
Q15 vector absolute value.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. The Q15 value -1 (0x8000) will be saturated to the maximum allowable positive value 0x7FFF.

**Parameters**

- [in] pSrc: points to the input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_abs\_q31** (**const** q31\_t \*pSrc, q31\_t \*pDst, uint32\_t blockSize)  
Q31 vector absolute value.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. The Q31 value -1 (0x80000000) will be saturated to the maximum allowable positive value 0x7FFFFFFF.

**Parameters**

- [in] pSrc: points to the input vector

- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_abs\_q7** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)  
Q7 vector absolute value.

**Return** none

**Conditions for optimum performance** Input and output buffers should be aligned by 32-bit

**Scaling and Overflow Behavior** The function uses saturating arithmetic. The Q7 value -1 (0x80) will be saturated to the maximum allowable positive value 0x7F.

**Parameters**

- [in] pSrc: points to the input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

## Vector Addition

void **riscv\_add\_f16** (const float16\_t \*pSrcA, const float16\_t \*pSrcB, float16\_t \*pDst, uint32\_t blockSize)

void **riscv\_add\_f32** (const float32\_t \*pSrcA, const float32\_t \*pSrcB, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_add\_q15** (const q15\_t \*pSrcA, const q15\_t \*pSrcB, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_add\_q31** (const q31\_t \*pSrcA, const q31\_t \*pSrcB, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_add\_q7** (const q7\_t \*pSrcA, const q7\_t \*pSrcB, q7\_t \*pDst, uint32\_t blockSize)

group **BasicAdd**

Element-by-element addition of two vectors.

There are separate functions for floating-point, Q7, Q15, and Q31 data types.

## Functions

void **riscv\_add\_f16** (const float16\_t \*pSrcA, const float16\_t \*pSrcB, float16\_t \*pDst, uint32\_t blockSize)  
Floating-point vector addition.

**Return** none

**Parameters**

- [in] pSrcA: points to first input vector
- [in] pSrcB: points to second input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_add\_f32** (const float32\_t \*pSrcA, const float32\_t \*pSrcB, float32\_t \*pDst, uint32\_t blockSize)  
Floating-point vector addition.

**Return** none

**Parameters**

- [in] `pSrcA`: points to first input vector
- [in] `pSrcB`: points to second input vector
- [out] `pDst`: points to output vector
- [in] `blockSize`: number of samples in each vector

void **riscv\_add\_q15** (**const** q15\_t \**pSrcA*, **const** q15\_t \**pSrcB*, q15\_t \**pDst*, uint32\_t *blockSize*)  
Q15 vector addition.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

**Parameters**

- [in] `pSrcA`: points to the first input vector
- [in] `pSrcB`: points to the second input vector
- [out] `pDst`: points to the output vector
- [in] `blockSize`: number of samples in each vector

void **riscv\_add\_q31** (**const** q31\_t \**pSrcA*, **const** q31\_t \**pSrcB*, q31\_t \**pDst*, uint32\_t *blockSize*)  
Q31 vector addition.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range [0x80000000 0x7FFFFFFF] are saturated.

**Parameters**

- [in] `pSrcA`: points to the first input vector
- [in] `pSrcB`: points to the second input vector
- [out] `pDst`: points to the output vector
- [in] `blockSize`: number of samples in each vector

void **riscv\_add\_q7** (**const** q7\_t \**pSrcA*, **const** q7\_t \**pSrcB*, q7\_t \**pDst*, uint32\_t *blockSize*)  
Q7 vector addition.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q7 range [0x80 0x7F] are saturated.

**Parameters**

- [in] `pSrcA`: points to the first input vector
- [in] `pSrcB`: points to the second input vector
- [out] `pDst`: points to the output vector
- [in] `blockSize`: number of samples in each vector

## Vector bitwise AND

void **riscv\_and\_u16**(**const** uint16\_t \*pSrcA, **const** uint16\_t \*pSrcB, uint16\_t \*pDst, uint32\_t *blockSize*)

void **riscv\_and\_u32**(**const** uint32\_t \*pSrcA, **const** uint32\_t \*pSrcB, uint32\_t \*pDst, uint32\_t *blockSize*)

void **riscv\_and\_u8**(**const** uint8\_t \*pSrcA, **const** uint8\_t \*pSrcB, uint8\_t \*pDst, uint32\_t *blockSize*)

### group And

Compute the logical bitwise AND.

There are separate functions for uint32\_t, uint16\_t, and uint8\_t data types.

## Functions

void **riscv\_and\_u16**(**const** uint16\_t \*pSrcA, **const** uint16\_t \*pSrcB, uint16\_t \*pDst, uint32\_t *blockSize*)

Compute the logical bitwise AND of two fixed-point vectors.

**Return** none

### Parameters

- [in] pSrcA: points to input vector A
- [in] pSrcB: points to input vector B
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_and\_u32**(**const** uint32\_t \*pSrcA, **const** uint32\_t \*pSrcB, uint32\_t \*pDst, uint32\_t *blockSize*)

Compute the logical bitwise AND of two fixed-point vectors.

**Return** none

### Parameters

- [in] pSrcA: points to input vector A
- [in] pSrcB: points to input vector B
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_and\_u8**(**const** uint8\_t \*pSrcA, **const** uint8\_t \*pSrcB, uint8\_t \*pDst, uint32\_t *blockSize*)

Compute the logical bitwise AND of two fixed-point vectors.

**Return** none

### Parameters

- [in] pSrcA: points to input vector A
- [in] pSrcB: points to input vector B
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector



## Elementwise clipping

void **riscv\_clip\_f16** (const float16\_t \*pSrc, float16\_t \*pDst, float16\_t low, float16\_t high, uint32\_t numSamples)

void **riscv\_clip\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, float32\_t low, float32\_t high, uint32\_t numSamples)

void **riscv\_clip\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, q15\_t low, q15\_t high, uint32\_t numSamples)

void **riscv\_clip\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, q31\_t low, q31\_t high, uint32\_t numSamples)

void **riscv\_clip\_q7** (const q7\_t \*pSrc, q7\_t \*pDst, q7\_t low, q7\_t high, uint32\_t numSamples)

### group BasicClip

Element-by-element clipping of a value.

The value is constrained between 2 bounds.

There are separate functions for floating-point, Q7, Q15, and Q31 data types.

## Functions

void **riscv\_clip\_f16** (const float16\_t \*pSrc, float16\_t \*pDst, float16\_t low, float16\_t high, uint32\_t numSamples)

Elementwise floating-point clipping.

**Return** none

### Parameters

- [in] pSrc: points to input values
- [out] pDst: points to output clipped values
- [in] low: lower bound
- [in] high: higher bound
- [in] numSamples: number of samples to clip

void **riscv\_clip\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, float32\_t low, float32\_t high, uint32\_t numSamples)

Elementwise floating-point clipping.

**Return** none

### Parameters

- [in] pSrc: points to input values
- [out] pDst: points to output clipped values
- [in] low: lower bound
- [in] high: higher bound
- [in] numSamples: number of samples to clip

void **riscv\_clip\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, q15\_t low, q15\_t high, uint32\_t numSamples)

Elementwise fixed-point clipping.

**Return** none

**Parameters**

- [in] pSrc: points to input values
- [out] pDst: points to output clipped values
- [in] low: lower bound
- [in] high: higher bound
- [in] numSamples: number of samples to clip

void **riscv\_clip\_q31** (**const** q31\_t \*pSrc, q31\_t \*pDst, q31\_t low, q31\_t high, uint32\_t numSamples)  
Elementwise fixed-point clipping.

**Return** none

**Parameters**

- [in] pSrc: points to input values
- [out] pDst: points to output clipped values
- [in] low: lower bound
- [in] high: higher bound
- [in] numSamples: number of samples to clip

void **riscv\_clip\_q7** (**const** q7\_t \*pSrc, q7\_t \*pDst, q7\_t low, q7\_t high, uint32\_t numSamples)  
Elementwise fixed-point clipping.

**Return** none

**Parameters**

- [in] pSrc: points to input values
- [out] pDst: points to output clipped values
- [in] low: lower bound
- [in] high: higher bound
- [in] numSamples: number of samples to clip

**Vector Dot Product**

void **riscv\_dot\_prod\_f16** (**const** float16\_t \*pSrcA, **const** float16\_t \*pSrcB, uint32\_t blockSize, float16\_t \*result)

void **riscv\_dot\_prod\_f32** (**const** float32\_t \*pSrcA, **const** float32\_t \*pSrcB, uint32\_t blockSize, float32\_t \*result)

void **riscv\_dot\_prod\_q15** (**const** q15\_t \*pSrcA, **const** q15\_t \*pSrcB, uint32\_t blockSize, q63\_t \*result)

void **riscv\_dot\_prod\_q31** (**const** q31\_t \*pSrcA, **const** q31\_t \*pSrcB, uint32\_t blockSize, q63\_t \*result)

void **riscv\_dot\_prod\_q7** (**const** q7\_t \*pSrcA, **const** q7\_t \*pSrcB, uint32\_t blockSize, q31\_t \*result)

*group* **BasicDotProd**

Computes the dot product of two vectors. The vectors are multiplied element-by-element and then summed.

There are separate functions for floating-point, Q7, Q15, and Q31 data types.

**Functions**

void **riscv\_dot\_prod\_f16**(**const** float16\_t \*pSrcA, **const** float16\_t \*pSrcB, uint32\_t blockSize, float16\_t \*result)  
Dot product of floating-point vectors.

**Return** none

**Parameters**

- [in] pSrcA: points to the first input vector.
- [in] pSrcB: points to the second input vector.
- [in] blockSize: number of samples in each vector.
- [out] result: output result returned here.

void **riscv\_dot\_prod\_f32**(**const** float32\_t \*pSrcA, **const** float32\_t \*pSrcB, uint32\_t blockSize, float32\_t \*result)  
Dot product of floating-point vectors.

**Return** none

**Parameters**

- [in] pSrcA: points to the first input vector.
- [in] pSrcB: points to the second input vector.
- [in] blockSize: number of samples in each vector.
- [out] result: output result returned here.

void **riscv\_dot\_prod\_q15**(**const** q15\_t \*pSrcA, **const** q15\_t \*pSrcB, uint32\_t blockSize, q63\_t \*result)  
Dot product of Q15 vectors.

**Return** none

**Scaling and Overflow Behavior** The intermediate multiplications are in  $1.15 \times 1.15 = 2.30$  format and these results are added to a 64-bit accumulator in 34.30 format. Nonsaturating additions are used and given that there are 33 guard bits in the accumulator there is no risk of overflow. The return result is in 34.30 format.

**Parameters**

- [in] pSrcA: points to the first input vector
- [in] pSrcB: points to the second input vector
- [in] blockSize: number of samples in each vector
- [out] result: output result returned here

void **riscv\_dot\_prod\_q31**(**const** q31\_t \*pSrcA, **const** q31\_t \*pSrcB, uint32\_t blockSize, q63\_t \*result)  
Dot product of Q31 vectors.

**Return** none

**Scaling and Overflow Behavior** The intermediate multiplications are in  $1.31 \times 1.31 = 2.62$  format and these are truncated to 2.48 format by discarding the lower 14 bits. The 2.48 result is then added without saturation to a 64-bit accumulator in 16.48 format. There are 15 guard bits in the accumulator and there is no risk of overflow as long as the length of the vectors is less than  $2^{16}$  elements. The return result is in 16.48 format.

**Parameters**

- [in] pSrcA: points to the first input vector.
- [in] pSrcB: points to the second input vector.
- [in] blockSize: number of samples in each vector.
- [out] result: output result returned here.

void **riscv\_dot\_prod\_q7** (**const** q7\_t \*pSrcA, **const** q7\_t \*pSrcB, uint32\_t blockSize, q31\_t \*result)  
Dot product of Q7 vectors.

**Return** none

**Scaling and Overflow Behavior** The intermediate multiplications are in  $1.7 \times 1.7 = 2.14$  format and these results are added to an accumulator in 18.14 format. Nonsaturating additions are used and there is no danger of wrap around as long as the vectors are less than  $2^{18}$  elements long. The return result is in 18.14 format.

**Parameters**

- [in] pSrcA: points to the first input vector
- [in] pSrcB: points to the second input vector
- [in] blockSize: number of samples in each vector
- [out] result: output result returned here

## Vector Multiplication

void **riscv\_mult\_f16** (**const** float16\_t \*pSrcA, **const** float16\_t \*pSrcB, float16\_t \*pDst, uint32\_t blockSize)

void **riscv\_mult\_f32** (**const** float32\_t \*pSrcA, **const** float32\_t \*pSrcB, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_mult\_q15** (**const** q15\_t \*pSrcA, **const** q15\_t \*pSrcB, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_mult\_q31** (**const** q31\_t \*pSrcA, **const** q31\_t \*pSrcB, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_mult\_q7** (**const** q7\_t \*pSrcA, **const** q7\_t \*pSrcB, q7\_t \*pDst, uint32\_t blockSize)

*group* **BasicMult**

Element-by-element multiplication of two vectors.

There are separate functions for floating-point, Q7, Q15, and Q31 data types.

## Functions

void **riscv\_mult\_f16** (**const** float16\_t \*pSrcA, **const** float16\_t \*pSrcB, float16\_t \*pDst, uint32\_t blockSize)  
Floating-point vector multiplication.

**Return** none

**Parameters**

- [in] pSrcA: points to the first input vector.
- [in] pSrcB: points to the second input vector.
- [out] pDst: points to the output vector.
- [in] blockSize: number of samples in each vector.

void **riscv\_mult\_f32** (**const** float32\_t \*pSrcA, **const** float32\_t \*pSrcB, float32\_t \*pDst, uint32\_t *blockSize*)  
Floating-point vector multiplication.

**Return** none

**Parameters**

- [in] pSrcA: points to the first input vector.
- [in] pSrcB: points to the second input vector.
- [out] pDst: points to the output vector.
- [in] blockSize: number of samples in each vector.

void **riscv\_mult\_q15** (**const** q15\_t \*pSrcA, **const** q15\_t \*pSrcB, q15\_t \*pDst, uint32\_t *blockSize*)  
Q15 vector multiplication.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

**Parameters**

- [in] pSrcA: points to first input vector
- [in] pSrcB: points to second input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_mult\_q31** (**const** q31\_t \*pSrcA, **const** q31\_t \*pSrcB, q31\_t \*pDst, uint32\_t *blockSize*)  
Q31 vector multiplication.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range [0x80000000 0x7FFFFFFF] are saturated.

**Parameters**

- [in] pSrcA: points to the first input vector.
- [in] pSrcB: points to the second input vector.
- [out] pDst: points to the output vector.
- [in] blockSize: number of samples in each vector.

void **riscv\_mult\_q7** (const q7\_t \*pSrcA, const q7\_t \*pSrcB, q7\_t \*pDst, uint32\_t blockSize)  
Q7 vector multiplication.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q7 range [0x80 0x7F] are saturated.

**Parameters**

- [in] pSrcA: points to the first input vector
- [in] pSrcB: points to the second input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

## Vector Negate

void **riscv\_negate\_f16** (const float16\_t \*pSrc, float16\_t \*pDst, uint32\_t blockSize)

void **riscv\_negate\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_negate\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_negate\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_negate\_q7** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

group **BasicNegate**

Negates the elements of a vector.

The functions support in-place computation allowing the source and destination pointers to reference the same memory buffer. There are separate functions for floating-point, Q7, Q15, and Q31 data types.

## Functions

void **riscv\_negate\_f16** (const float16\_t \*pSrc, float16\_t \*pDst, uint32\_t blockSize)  
Negates the elements of a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector.
- [out] pDst: points to output vector.
- [in] blockSize: number of samples in each vector.

void **riscv\_negate\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)  
Negates the elements of a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector.
- [out] pDst: points to output vector.

- [in] blockSize: number of samples in each vector.

void **riscv\_negate\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

Negates the elements of a Q15 vector.

**Return** none

**Conditions for optimum performance** Input and output buffers should be aligned by 32-bit

**Scaling and Overflow Behavior** The function uses saturating arithmetic. The Q15 value -1 (0x8000) is saturated to the maximum allowable positive value 0x7FFF.

**Parameters**

- [in] pSrc: points to the input vector.
- [out] pDst: points to the output vector.
- [in] blockSize: number of samples in each vector.

void **riscv\_negate\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t blockSize)

Negates the elements of a Q31 vector.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. The Q31 value -1 (0x80000000) is saturated to the maximum allowable positive value 0x7FFFFFFF.

**Parameters**

- [in] pSrc: points to the input vector.
- [out] pDst: points to the output vector.
- [in] blockSize: number of samples in each vector.

void **riscv\_negate\_q7** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

Negates the elements of a Q7 vector.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. The Q7 value -1 (0x80) is saturated to the maximum allowable positive value 0x7F.

**Parameters**

- [in] pSrc: points to the input vector.
- [out] pDst: points to the output vector.
- [in] blockSize: number of samples in each vector.

### Vector bitwise NOT

void **riscv\_not\_u16** (const uint16\_t \*pSrc, uint16\_t \*pDst, uint32\_t blockSize)

void **riscv\_not\_u32** (const uint32\_t \*pSrc, uint32\_t \*pDst, uint32\_t blockSize)

void **riscv\_not\_u8** (const uint8\_t \*pSrc, uint8\_t \*pDst, uint32\_t blockSize)

**group Not**

Compute the logical bitwise NOT.

There are separate functions for uint32\_t, uint16\_t, and uint8\_t data types.

**Functions**

void **riscv\_not\_u16** (**const** uint16\_t \*pSrc, uint16\_t \*pDst, uint32\_t blockSize)

Compute the logical bitwise NOT of a fixed-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_not\_u32** (**const** uint32\_t \*pSrc, uint32\_t \*pDst, uint32\_t blockSize)

Compute the logical bitwise NOT of a fixed-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_not\_u8** (**const** uint8\_t \*pSrc, uint8\_t \*pDst, uint32\_t blockSize)

Compute the logical bitwise NOT of a fixed-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

**Vector Offset**

void **riscv\_offset\_f16** (**const** float16\_t \*pSrc, float16\_t offset, float16\_t \*pDst, uint32\_t blockSize)

void **riscv\_offset\_f32** (**const** float32\_t \*pSrc, float32\_t offset, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_offset\_q15** (**const** q15\_t \*pSrc, q15\_t offset, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_offset\_q31** (**const** q31\_t \*pSrc, q31\_t offset, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_offset\_q7** (**const** q7\_t \*pSrc, q7\_t offset, q7\_t \*pDst, uint32\_t blockSize)



**group BasicOffset**

Adds a constant offset to each element of a vector.

The functions support in-place computation allowing the source and destination pointers to reference the same memory buffer. There are separate functions for floating-point, Q7, Q15, and Q31 data types.

**Functions**

void **riscv\_offset\_f16**(**const** float16\_t \*pSrc, float16\_t offset, float16\_t \*pDst, uint32\_t block-  
Size)

Adds a constant offset to a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [in] offset: is the offset to be added
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_offset\_f32**(**const** float32\_t \*pSrc, float32\_t offset, float32\_t \*pDst, uint32\_t block-  
Size)

Adds a constant offset to a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [in] offset: is the offset to be added
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_offset\_q15**(**const** q15\_t \*pSrc, q15\_t offset, q15\_t \*pDst, uint32\_t blockSize)

Adds a constant offset to a Q15 vector.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

**Parameters**

- [in] pSrc: points to the input vector
- [in] offset: is the offset to be added
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_offset\_q31**(**const** q31\_t \*pSrc, q31\_t offset, q31\_t \*pDst, uint32\_t blockSize)

Adds a constant offset to a Q31 vector.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range [0x80000000 0x7FFFFFFF] are saturated.

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *offset*: is the offset to be added
- [out] *pDst*: points to the output vector
- [in] *blockSize*: number of samples in each vector

void **riscv\_offset\_q7** (const q7\_t \**pSrc*, q7\_t *offset*, q7\_t \**pDst*, uint32\_t *blockSize*)

Adds a constant offset to a Q7 vector.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q7 range [0x80 0x7F] are saturated.

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *offset*: is the offset to be added
- [out] *pDst*: points to the output vector
- [in] *blockSize*: number of samples in each vector

## Vector bitwise inclusive OR

void **riscv\_or\_u16** (const uint16\_t \**pSrcA*, const uint16\_t \**pSrcB*, uint16\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_or\_u32** (const uint32\_t \**pSrcA*, const uint32\_t \**pSrcB*, uint32\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_or\_u8** (const uint8\_t \**pSrcA*, const uint8\_t \**pSrcB*, uint8\_t \**pDst*, uint32\_t *blockSize*)

*group Or*

Compute the logical bitwise OR.

There are separate functions for uint32\_t, uint16\_t, and uint8\_t data types.

## Functions

void **riscv\_or\_u16** (const uint16\_t \**pSrcA*, const uint16\_t \**pSrcB*, uint16\_t \**pDst*, uint32\_t *blockSize*)

Compute the logical bitwise OR of two fixed-point vectors.

**Return** none

**Parameters**

- [in] *pSrcA*: points to input vector A
- [in] *pSrcB*: points to input vector B
- [out] *pDst*: points to output vector
- [in] *blockSize*: number of samples in each vector

void **riscv\_or\_u32** (const uint32\_t \*pSrcA, const uint32\_t \*pSrcB, uint32\_t \*pDst, uint32\_t blockSize)

Compute the logical bitwise OR of two fixed-point vectors.

**Return** none

**Parameters**

- [in] pSrcA: points to input vector A
- [in] pSrcB: points to input vector B
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_or\_u8** (const uint8\_t \*pSrcA, const uint8\_t \*pSrcB, uint8\_t \*pDst, uint32\_t blockSize)

Compute the logical bitwise OR of two fixed-point vectors.

**Return** none

**Parameters**

- [in] pSrcA: points to input vector A
- [in] pSrcB: points to input vector B
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

## Vector Scale

void **riscv\_scale\_f16** (const float16\_t \*pSrc, float16\_t scale, float16\_t \*pDst, uint32\_t blockSize)

void **riscv\_scale\_f32** (const float32\_t \*pSrc, float32\_t scale, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_scale\_q15** (const q15\_t \*pSrc, q15\_t scaleFract, int8\_t shift, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_scale\_q31** (const q31\_t \*pSrc, q31\_t scaleFract, int8\_t shift, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_scale\_q7** (const q7\_t \*pSrc, q7\_t scaleFract, int8\_t shift, q7\_t \*pDst, uint32\_t blockSize)

group **BasicScale**

Multiply a vector by a scalar value. For floating-point data, the algorithm used is:

In the fixed-point Q7, Q15, and Q31 functions, `scale` is represented by a fractional multiplication `scaleFract` and an arithmetic shift `shift`. The shift allows the gain of the scaling operation to exceed 1.0. The algorithm used with fixed-point data is:

The overall scale factor applied to the fixed-point data is

The functions support in-place computation allowing the source and destination pointers to reference the same memory buffer.

## Functions

void **riscv\_scale\_f16** (const float16\_t \*pSrc, float16\_t scale, float16\_t \*pDst, uint32\_t blockSize)

Multiplies a floating-point vector by a scalar.

**Return** none

**Parameters**

- [in] `pSrc`: points to the input vector
- [in] `scale`: scale factor to be applied
- [out] `pDst`: points to the output vector
- [in] `blockSize`: number of samples in each vector

void **riscv\_scale\_f32** (**const** float32\_t \**pSrc*, float32\_t *scale*, float32\_t \**pDst*, uint32\_t *blockSize*)  
Multiplies a floating-point vector by a scalar.

**Return** none

**Parameters**

- [in] `pSrc`: points to the input vector
- [in] `scale`: scale factor to be applied
- [out] `pDst`: points to the output vector
- [in] `blockSize`: number of samples in each vector

void **riscv\_scale\_q15** (**const** q15\_t \**pSrc*, q15\_t *scaleFract*, int8\_t *shift*, q15\_t \**pDst*, uint32\_t *blockSize*)  
Multiplies a Q15 vector by a scalar.

**Return** none

**Scaling and Overflow Behavior** The input data \**pSrc* and *scaleFract* are in 1.15 format. These are multiplied to yield a 2.30 intermediate result and this is shifted with saturation to 1.15 format.

**Parameters**

- [in] `pSrc`: points to the input vector
- [in] `scaleFract`: fractional portion of the scale value
- [in] `shift`: number of bits to shift the result by
- [out] `pDst`: points to the output vector
- [in] `blockSize`: number of samples in each vector

void **riscv\_scale\_q31** (**const** q31\_t \**pSrc*, q31\_t *scaleFract*, int8\_t *shift*, q31\_t \**pDst*, uint32\_t *blockSize*)  
Multiplies a Q31 vector by a scalar.

**Return** none

**Scaling and Overflow Behavior** The input data \**pSrc* and *scaleFract* are in 1.31 format. These are multiplied to yield a 2.62 intermediate result and this is shifted with saturation to 1.31 format.

**Parameters**

- [in] `pSrc`: points to the input vector
- [in] `scaleFract`: fractional portion of the scale value
- [in] `shift`: number of bits to shift the result by
- [out] `pDst`: points to the output vector

- [in] blockSize: number of samples in each vector

void **riscv\_scale\_q7** (const q7\_t \*pSrc, q7\_t scaleFract, int8\_t shift, q7\_t \*pDst, uint32\_t blockSize)

Multiplies a Q7 vector by a scalar.

**Return** none

**Scaling and Overflow Behavior** The input data \*pSrc and scaleFract are in 1.7 format. These are multiplied to yield a 2.14 intermediate result and this is shifted with saturation to 1.7 format.

#### Parameters

- [in] pSrc: points to the input vector
- [in] scaleFract: fractional portion of the scale value
- [in] shift: number of bits to shift the result by
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

## Vector Shift

void **riscv\_shift\_q15** (const q15\_t \*pSrc, int8\_t shiftBits, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_shift\_q31** (const q31\_t \*pSrc, int8\_t shiftBits, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_shift\_q7** (const q7\_t \*pSrc, int8\_t shiftBits, q7\_t \*pDst, uint32\_t blockSize)

#### group BasicShift

Shifts the elements of a fixed-point vector by a specified number of bits. There are separate functions for Q7, Q15, and Q31 data types. The underlying algorithm used is:

If shift is positive then the elements of the vector are shifted to the left. If shift is negative then the elements of the vector are shifted to the right.

The functions support in-place computation allowing the source and destination pointers to reference the same memory buffer.

## Functions

void **riscv\_shift\_q15** (const q15\_t \*pSrc, int8\_t shiftBits, q15\_t \*pDst, uint32\_t blockSize)

Shifts the elements of a Q15 vector a specified number of bits.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

#### Parameters

- [in] pSrc: points to the input vector
- [in] shiftBits: number of bits to shift. A positive value shifts left; a negative value shifts right.
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_shift\_q31** (**const** q31\_t \*pSrc, int8\_t shiftBits, q31\_t \*pDst, uint32\_t blockSize)  
Shifts the elements of a Q31 vector a specified number of bits.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range [0x80000000 0x7FFFFFFF] are saturated.

**Parameters**

- [in] pSrc: points to the input vector
- [in] shiftBits: number of bits to shift. A positive value shifts left; a negative value shifts right.
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in the vector

void **riscv\_shift\_q7** (**const** q7\_t \*pSrc, int8\_t shiftBits, q7\_t \*pDst, uint32\_t blockSize)  
Shifts the elements of a Q7 vector a specified number of bits.

**Return** none

**Conditions for optimum performance** Input and output buffers should be aligned by 32-bit

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q7 range [0x80 0x7F] are saturated.

**Parameters**

- [in] pSrc: points to the input vector
- [in] shiftBits: number of bits to shift. A positive value shifts left; a negative value shifts right.
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

## Vector Subtraction

void **riscv\_sub\_f16** (**const** float16\_t \*pSrcA, **const** float16\_t \*pSrcB, float16\_t \*pDst, uint32\_t blockSize)

void **riscv\_sub\_f32** (**const** float32\_t \*pSrcA, **const** float32\_t \*pSrcB, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_sub\_q15** (**const** q15\_t \*pSrcA, **const** q15\_t \*pSrcB, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_sub\_q31** (**const** q31\_t \*pSrcA, **const** q31\_t \*pSrcB, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_sub\_q7** (**const** q7\_t \*pSrcA, **const** q7\_t \*pSrcB, q7\_t \*pDst, uint32\_t blockSize)

*group* **BasicSub**

Element-by-element subtraction of two vectors.

There are separate functions for floating-point, Q7, Q15, and Q31 data types.

## Functions

void **riscv\_sub\_f16**(**const** float16\_t \*pSrcA, **const** float16\_t \*pSrcB, float16\_t \*pDst, uint32\_t blockSize)  
Floating-point vector subtraction.

**Return** none

**Parameters**

- [in] pSrcA: points to the first input vector
- [in] pSrcB: points to the second input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_sub\_f32**(**const** float32\_t \*pSrcA, **const** float32\_t \*pSrcB, float32\_t \*pDst, uint32\_t blockSize)  
Floating-point vector subtraction.

**Return** none

**Parameters**

- [in] pSrcA: points to the first input vector
- [in] pSrcB: points to the second input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_sub\_q15**(**const** q15\_t \*pSrcA, **const** q15\_t \*pSrcB, q15\_t \*pDst, uint32\_t blockSize)  
Q15 vector subtraction.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

**Parameters**

- [in] pSrcA: points to the first input vector
- [in] pSrcB: points to the second input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_sub\_q31**(**const** q31\_t \*pSrcA, **const** q31\_t \*pSrcB, q31\_t \*pDst, uint32\_t blockSize)  
Q31 vector subtraction.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range [0x80000000 0x7FFFFFFF] are saturated.

**Parameters**

- [in] pSrcA: points to the first input vector

- [in] pSrcB: points to the second input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

void **riscv\_sub\_q7**(const q7\_t \*pSrcA, const q7\_t \*pSrcB, q7\_t \*pDst, uint32\_t blockSize)  
Q7 vector subtraction.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q7 range [0x80 0x7F] will be saturated.

**Parameters**

- [in] pSrcA: points to the first input vector
- [in] pSrcB: points to the second input vector
- [out] pDst: points to the output vector
- [in] blockSize: number of samples in each vector

### Vector bitwise exclusive OR

void **riscv\_xor\_u16**(const uint16\_t \*pSrcA, const uint16\_t \*pSrcB, uint16\_t \*pDst, uint32\_t blockSize)  
Size)

void **riscv\_xor\_u32**(const uint32\_t \*pSrcA, const uint32\_t \*pSrcB, uint32\_t \*pDst, uint32\_t blockSize)  
Size)

void **riscv\_xor\_u8**(const uint8\_t \*pSrcA, const uint8\_t \*pSrcB, uint8\_t \*pDst, uint32\_t blockSize)

group **Xor**

Compute the logical bitwise XOR.

There are separate functions for uint32\_t, uint16\_t, and uint8\_t data types.

### Functions

void **riscv\_xor\_u16**(const uint16\_t \*pSrcA, const uint16\_t \*pSrcB, uint16\_t \*pDst, uint32\_t blockSize)  
blockSize)

Compute the logical bitwise XOR of two fixed-point vectors.

**Return** none

**Parameters**

- [in] pSrcA: points to input vector A
- [in] pSrcB: points to input vector B
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_xor\_u32**(const uint32\_t \*pSrcA, const uint32\_t \*pSrcB, uint32\_t \*pDst, uint32\_t blockSize)  
blockSize)

Compute the logical bitwise XOR of two fixed-point vectors.

**Return** none



**Parameters**

- [in] pSrcA: points to input vector A
- [in] pSrcB: points to input vector B
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_xor\_u8**(const uint8\_t \*pSrcA, const uint8\_t \*pSrcB, uint8\_t \*pDst, uint32\_t blockSize)

Compute the logical bitwise XOR of two fixed-point vectors.

**Return** none

**Parameters**

- [in] pSrcA: points to input vector A
- [in] pSrcB: points to input vector B
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

group **groupMath**

### 3.3.3 Bayesian estimators

uint32\_t **riscv\_gaussian\_naive\_bayes\_predict\_f16**(const riscv\_gaussian\_naive\_bayes\_instance\_f16 \*S, const float16\_t \*in, float16\_t \*pOutputProbabilities, float16\_t \*pBufferB)

uint32\_t **riscv\_gaussian\_naive\_bayes\_predict\_f32**(const riscv\_gaussian\_naive\_bayes\_instance\_f32 \*S, const float32\_t \*in, float32\_t \*pOutputProbabilities, float32\_t \*pBufferB)

group **groupBayes**

Implement the naive gaussian Bayes estimator. The training must be done from scikit-learn.

The parameters can be easily generated from the scikit-learn object. Some examples are given in DSP/Testing/PatternGeneration/Bayes.py

**Functions**

uint32\_t **riscv\_gaussian\_naive\_bayes\_predict\_f16**(const riscv\_gaussian\_naive\_bayes\_instance\_f16 \*S, const float16\_t \*in, float16\_t \*pOutputProbabilities, float16\_t \*pBufferB)

Naive Gaussian Bayesian Estimator.

**Return** The predicted class

**Parameters**

- [in] \*S: points to a naive bayes instance structure
- [in] \*in: points to the elements of the input vector.

- [out] *\*pOutputProbabilities*: points to a buffer of length *numberOfClasses* containing estimated probabilities
- [out] *\*pBufferB*: points to a temporary buffer of length *numberOfClasses*

```
uint32_t riscv_gaussian_naive_bayes_predict_f32 (const
                                                riscv_gaussian_naive_bayes_instance_f32
                                                *S, const float32_t *in, float32_t
                                                *pOutputProbabilities, float32_t
                                                *pBufferB)
```

Naive Gaussian Bayesian Estimator.

**Return** The predicted class

**Parameters**

- [in] *\*S*: points to a naive bayes instance structure
- [in] *\*in*: points to the elements of the input vector.
- [out] *\*pOutputProbabilities*: points to a buffer of length *numberOfClasses* containing estimated probabilities
- [out] *\*pBufferB*: points to a temporary buffer of length *numberOfClasses*

### 3.3.4 Complex Math Functions

#### Complex Conjugate

```
void riscv_cmplx_conj_f16 (const float16_t *pSrc, float16_t *pDst, uint32_t numSamples)
```

```
void riscv_cmplx_conj_f32 (const float32_t *pSrc, float32_t *pDst, uint32_t numSamples)
```

```
void riscv_cmplx_conj_q15 (const q15_t *pSrc, q15_t *pDst, uint32_t numSamples)
```

```
void riscv_cmplx_conj_q31 (const q31_t *pSrc, q31_t *pDst, uint32_t numSamples)
```

*group* **cmplx\_conj**

Conjugates the elements of a complex data vector.

The *pSrc* points to the source data and *pDst* points to the destination data where the result should be written. *numSamples* specifies the number of complex samples and the data in each array is stored in an interleaved fashion (real, imag, real, imag, ...). Each array has a total of  $2 * \text{numSamples}$  values.

The underlying algorithm is used:

There are separate functions for floating-point, Q15, and Q31 data types.

#### Functions

```
void riscv_cmplx_conj_f16 (const float16_t *pSrc, float16_t *pDst, uint32_t numSamples)
```

Floating-point complex conjugate.

**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [out] *pDst*: points to the output vector

- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_conj\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t numSamples)  
Floating-point complex conjugate.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [out] pDst: points to the output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_conj\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t numSamples)  
Q15 complex conjugate.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. The Q15 value -1 (0x8000) is saturated to the maximum allowable positive value 0x7FFF.

**Parameters**

- [in] pSrc: points to the input vector
- [out] pDst: points to the output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_conj\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t numSamples)  
Q31 complex conjugate.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. The Q31 value -1 (0x80000000) is saturated to the maximum allowable positive value 0x7FFFFFFF.

**Parameters**

- [in] pSrc: points to the input vector
- [out] pDst: points to the output vector
- [in] numSamples: number of samples in each vector

## Complex Dot Product

void **riscv\_cmplx\_dot\_prod\_f16** (const float16\_t \*pSrcA, const float16\_t \*pSrcB, uint32\_t numSamples, float16\_t \*realResult, float16\_t \*imagResult)

void **riscv\_cmplx\_dot\_prod\_f32** (const float32\_t \*pSrcA, const float32\_t \*pSrcB, uint32\_t numSamples, float32\_t \*realResult, float32\_t \*imagResult)

void **riscv\_cmplx\_dot\_prod\_q15** (const q15\_t \*pSrcA, const q15\_t \*pSrcB, uint32\_t numSamples, q15\_t \*realResult, q15\_t \*imagResult)

void **riscv\_cmplx\_dot\_prod\_q31** (const q31\_t \*pSrcA, const q31\_t \*pSrcB, uint32\_t numSamples, q31\_t \*realResult, q31\_t \*imagResult)

**group `cmplx_dot_prod`**

Computes the dot product of two complex vectors. The vectors are multiplied element-by-element and then summed.

The `pSrcA` points to the first complex input vector and `pSrcB` points to the second complex input vector. `numSamples` specifies the number of complex samples and the data in each array is stored in an interleaved fashion (real, imag, real, imag, ...). Each array has a total of  $2 \times \text{numSamples}$  values.

The underlying algorithm is used:

There are separate functions for floating-point, Q15, and Q31 data types.

**Functions**

void **riscv\_cmplx\_dot\_prod\_f16**(const float16\_t \*pSrcA, const float16\_t \*pSrcB, uint32\_t numSamples, float16\_t \*realResult, float16\_t \*imagResult)  
Floating-point complex dot product.

**Return** none

**Parameters**

- [in] `pSrcA`: points to the first input vector
- [in] `pSrcB`: points to the second input vector
- [in] `numSamples`: number of samples in each vector
- [out] `realResult`: real part of the result returned here
- [out] `imagResult`: imaginary part of the result returned here

void **riscv\_cmplx\_dot\_prod\_f32**(const float32\_t \*pSrcA, const float32\_t \*pSrcB, uint32\_t numSamples, float32\_t \*realResult, float32\_t \*imagResult)  
Floating-point complex dot product.

**Return** none

**Parameters**

- [in] `pSrcA`: points to the first input vector
- [in] `pSrcB`: points to the second input vector
- [in] `numSamples`: number of samples in each vector
- [out] `realResult`: real part of the result returned here
- [out] `imagResult`: imaginary part of the result returned here

void **riscv\_cmplx\_dot\_prod\_q15**(const q15\_t \*pSrcA, const q15\_t \*pSrcB, uint32\_t numSamples, q31\_t \*realResult, q31\_t \*imagResult)  
Q15 complex dot product.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The intermediate 1.15 by 1.15 multiplications are performed with full precision and yield a 2.30 result. These are accumulated in a 64-bit accumulator with 34.30 precision. As a final step, the accumulators are converted to 8.24 format. The return results `realResult` and `imagResult` are in 8.24 format.

**Parameters**

- [in] `pSrcA`: points to the first input vector
- [in] `pSrcB`: points to the second input vector
- [in] `numSamples`: number of samples in each vector
- [out] `realResult`: real part of the result returned here
- [out] `imagResult`: imaginary part of the result returned here

void **riscv\_cmplx\_dot\_prod\_q31**(const q31\_t \**pSrcA*, const q31\_t \**pSrcB*, uint32\_t *numSamples*, q63\_t \**realResult*, q63\_t \**imagResult*)

Q31 complex dot product.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The intermediate 1.31 by 1.31 multiplications are performed with 64-bit precision and then shifted to 16.48 format. The internal real and imaginary accumulators are in 16.48 format and provide 15 guard bits. Additions are nonsaturating and no overflow will occur as long as `numSamples` is less than 32768. The return results `realResult` and `imagResult` are in 16.48 format. Input down scaling is not required.

#### Parameters

- [in] `pSrcA`: points to the first input vector
- [in] `pSrcB`: points to the second input vector
- [in] `numSamples`: number of samples in each vector
- [out] `realResult`: real part of the result returned here
- [out] `imagResult`: imaginary part of the result returned here

## Complex Magnitude

void **riscv\_cmplx\_mag\_f16**(const float16\_t \**pSrc*, float16\_t \**pDst*, uint32\_t *numSamples*)

void **riscv\_cmplx\_mag\_f32**(const float32\_t \**pSrc*, float32\_t \**pDst*, uint32\_t *numSamples*)

void **riscv\_cmplx\_mag\_q15**(const q15\_t \**pSrc*, q15\_t \**pDst*, uint32\_t *numSamples*)

void **riscv\_cmplx\_mag\_q31**(const q31\_t \**pSrc*, q31\_t \**pDst*, uint32\_t *numSamples*)

group **cmplx\_mag**

Computes the magnitude of the elements of a complex data vector.

The `pSrc` points to the source data and `pDst` points to the where the result should be written. `numSamples` specifies the number of complex samples in the input array and the data is stored in an interleaved fashion (real, imag, real, imag, ...). The input array has a total of  $2 * \text{numSamples}$  values; the output array has a total of `numSamples` values.

The underlying algorithm is used:

There are separate functions for floating-point, Q15, and Q31 data types.

## Functions

void **riscv\_cmplx\_mag\_f16**(const float16\_t \**pSrc*, float16\_t \**pDst*, uint32\_t *numSamples*)

Floating-point complex magnitude.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mag\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t numSamples)  
Floating-point complex magnitude.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mag\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t numSamples)  
Q15 complex magnitude.

**Return** none

**Scaling and Overflow Behavior** The function implements 1.15 by 1.15 multiplications and finally output is converted into 2.14 format.

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mag\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t numSamples)  
Q31 complex magnitude.

**Return** none

**Scaling and Overflow Behavior** The function implements 1.31 by 1.31 multiplications and finally output is converted into 2.30 format. Input down scaling is not required.

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

## Complex Magnitude Squared

void **riscv\_cmplx\_mag\_squared\_f16** (const float16\_t \*pSrc, float16\_t \*pDst, uint32\_t numSamples)

void **riscv\_cmplx\_mag\_squared\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t numSamples)

void **riscv\_cmplx\_mag\_squared\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t numSamples)

void **riscv\_cmplx\_mag\_squared\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t numSamples)

*group* **cmplx\_mag\_squared**

Computes the magnitude squared of the elements of a complex data vector.

The `pSrc` points to the source data and `pDst` points to the where the result should be written. `numSamples` specifies the number of complex samples in the input array and the data is stored in an interleaved fashion (real, imag, real, imag, ...). The input array has a total of  $2 * \text{numSamples}$  values; the output array has a total of `numSamples` values.

The underlying algorithm is used:

There are separate functions for floating-point, Q15, and Q31 data types.

**Functions**

void **riscv\_cmplx\_mag\_squared\_f16**(const float16\_t \*pSrc, float16\_t \*pDst, uint32\_t numSamples)

Floating-point complex magnitude squared.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mag\_squared\_f32**(const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t numSamples)

Floating-point complex magnitude squared.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mag\_squared\_q15**(const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t numSamples)

Q15 complex magnitude squared.

**Return** none

**Scaling and Overflow Behavior** The function implements 1.15 by 1.15 multiplications and finally output is converted into 3.13 format.

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mag\_squared\_q31**(const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t numSamples)

Q31 complex magnitude squared.

**Return** none

**Scaling and Overflow Behavior** The function implements 1.31 by 1.31 multiplications and finally output is converted into 3.29 format. Input down scaling is not required.

**Parameters**

- [in] `pSrc`: points to input vector
- [out] `pDst`: points to output vector
- [in] `numSamples`: number of samples in each vector

## Complex-by-Complex Multiplication

```
void riscv_cmplx_mult_cmplx_f16(const float16_t *pSrcA, const float16_t *pSrcB, float16_t *pDst, uint32_t numSamples)
```

```
void riscv_cmplx_mult_cmplx_f32(const float32_t *pSrcA, const float32_t *pSrcB, float32_t *pDst, uint32_t numSamples)
```

```
void riscv_cmplx_mult_cmplx_q15(const q15_t *pSrcA, const q15_t *pSrcB, q15_t *pDst, uint32_t numSamples)
```

```
void riscv_cmplx_mult_cmplx_q31(const q31_t *pSrcA, const q31_t *pSrcB, q31_t *pDst, uint32_t numSamples)
```

*group* **CmplxByCmplxMult**

Multiplies a complex vector by another complex vector and generates a complex result. The data in the complex arrays is stored in an interleaved fashion (real, imag, real, imag, ...). The parameter `numSamples` represents the number of complex samples processed. The complex arrays have a total of  $2 * \text{numSamples}$  real values.

The underlying algorithm is used:

There are separate functions for floating-point, Q15, and Q31 data types.

## Functions

```
void riscv_cmplx_mult_cmplx_f16(const float16_t *pSrcA, const float16_t *pSrcB, float16_t *pDst, uint32_t numSamples)
```

Floating-point complex-by-complex multiplication.

**Return** none

**Parameters**

- [in] `pSrcA`: points to first input vector
- [in] `pSrcB`: points to second input vector
- [out] `pDst`: points to output vector
- [in] `numSamples`: number of samples in each vector

```
void riscv_cmplx_mult_cmplx_f32(const float32_t *pSrcA, const float32_t *pSrcB, float32_t *pDst, uint32_t numSamples)
```

Floating-point complex-by-complex multiplication.

**Return** none

**Parameters**

- [in] `pSrcA`: points to first input vector
- [in] `pSrcB`: points to second input vector



- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mult\_cmplx\_q15** (const q15\_t \*pSrcA, const q15\_t \*pSrcB, q15\_t \*pDst, uint32\_t numSamples)

Q15 complex-by-complex multiplication.

**Return** none

**Scaling and Overflow Behavior** The function implements 1.15 by 1.15 multiplications and finally output is converted into 3.13 format.

**Parameters**

- [in] pSrcA: points to first input vector
- [in] pSrcB: points to second input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mult\_cmplx\_q31** (const q31\_t \*pSrcA, const q31\_t \*pSrcB, q31\_t \*pDst, uint32\_t numSamples)

Q31 complex-by-complex multiplication.

**Return** none

**Scaling and Overflow Behavior** The function implements 1.31 by 1.31 multiplications and finally output is converted into 3.29 format. Input down scaling is not required.

**Parameters**

- [in] pSrcA: points to first input vector
- [in] pSrcB: points to second input vector
- [out] pDst: points to output vector
- [in] numSamples: number of samples in each vector

## Complex-by-Real Multiplication

void **riscv\_cmplx\_mult\_real\_f16** (const float16\_t \*pSrcCmplx, const float16\_t \*pSrcReal, float16\_t \*pCmplxDst, uint32\_t numSamples)

void **riscv\_cmplx\_mult\_real\_f32** (const float32\_t \*pSrcCmplx, const float32\_t \*pSrcReal, float32\_t \*pCmplxDst, uint32\_t numSamples)

void **riscv\_cmplx\_mult\_real\_q15** (const q15\_t \*pSrcCmplx, const q15\_t \*pSrcReal, q15\_t \*pCmplxDst, uint32\_t numSamples)

void **riscv\_cmplx\_mult\_real\_q31** (const q31\_t \*pSrcCmplx, const q31\_t \*pSrcReal, q31\_t \*pCmplxDst, uint32\_t numSamples)

group **CmplxByRealMult**

Multiplies a complex vector by a real vector and generates a complex result. The data in the complex arrays is stored in an interleaved fashion (real, imag, real, imag, ...). The parameter numSamples represents the number of complex samples processed. The complex arrays have a total of 2\*numSamples real values while the real array has a total of numSamples real values.

The underlying algorithm is used:

There are separate functions for floating-point, Q15, and Q31 data types.

## Functions

void **riscv\_cmplx\_mult\_real\_f16**(**const** float16\_t \*pSrcCmplx, **const** float16\_t \*pSrcReal, float16\_t \*pCmplxDst, uint32\_t numSamples)

Floating-point complex-by-real multiplication.

**Return** none

### Parameters

- [in] pSrcCmplx: points to complex input vector
- [in] pSrcReal: points to real input vector
- [out] pCmplxDst: points to complex output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mult\_real\_f32**(**const** float32\_t \*pSrcCmplx, **const** float32\_t \*pSrcReal, float32\_t \*pCmplxDst, uint32\_t numSamples)

Floating-point complex-by-real multiplication.

**Return** none

### Parameters

- [in] pSrcCmplx: points to complex input vector
- [in] pSrcReal: points to real input vector
- [out] pCmplxDst: points to complex output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mult\_real\_q15**(**const** q15\_t \*pSrcCmplx, **const** q15\_t \*pSrcReal, q15\_t \*pCmplxDst, uint32\_t numSamples)

Q15 complex-by-real multiplication.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

### Parameters

- [in] pSrcCmplx: points to complex input vector
- [in] pSrcReal: points to real input vector
- [out] pCmplxDst: points to complex output vector
- [in] numSamples: number of samples in each vector

void **riscv\_cmplx\_mult\_real\_q31**(**const** q31\_t \*pSrcCmplx, **const** q31\_t \*pSrcReal, q31\_t \*pCmplxDst, uint32\_t numSamples)

Q31 complex-by-real multiplication.

**Return** none

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range[0x80000000 0x7FFFFFFF] are saturated.

#### Parameters

- [in] `pSrcCmplx`: points to complex input vector
- [in] `pSrcReal`: points to real input vector
- [out] `pCmplxDst`: points to complex output vector
- [in] `numSamples`: number of samples in each vector

#### group **groupCmplxMath**

This set of functions operates on complex data vectors. The data in the complex arrays is stored in an interleaved fashion (real, imag, real, imag, ...). In the API functions, the number of samples in a complex array refers to the number of complex values; the array contains twice this number of real values.

### 3.3.5 Controller Functions

#### PID Motor Control

```
__STATIC_FORCEINLINE float32_t riscv_pid_f32(riscv_pid_instance_f32 *S, float32_t in)
__STATIC_FORCEINLINE q31_t riscv_pid_q31(riscv_pid_instance_q31 *S, q31_t in)
__STATIC_FORCEINLINE q15_t riscv_pid_q15(riscv_pid_instance_q15 *S, q15_t in)

void riscv_pid_init_f32(riscv_pid_instance_f32 *S, int32_t resetStateFlag)
void riscv_pid_init_q15(riscv_pid_instance_q15 *S, int32_t resetStateFlag)
void riscv_pid_init_q31(riscv_pid_instance_q31 *S, int32_t resetStateFlag)
void riscv_pid_reset_f32(riscv_pid_instance_f32 *S)
void riscv_pid_reset_q15(riscv_pid_instance_q15 *S)
void riscv_pid_reset_q31(riscv_pid_instance_q31 *S)
```

#### group **PID**

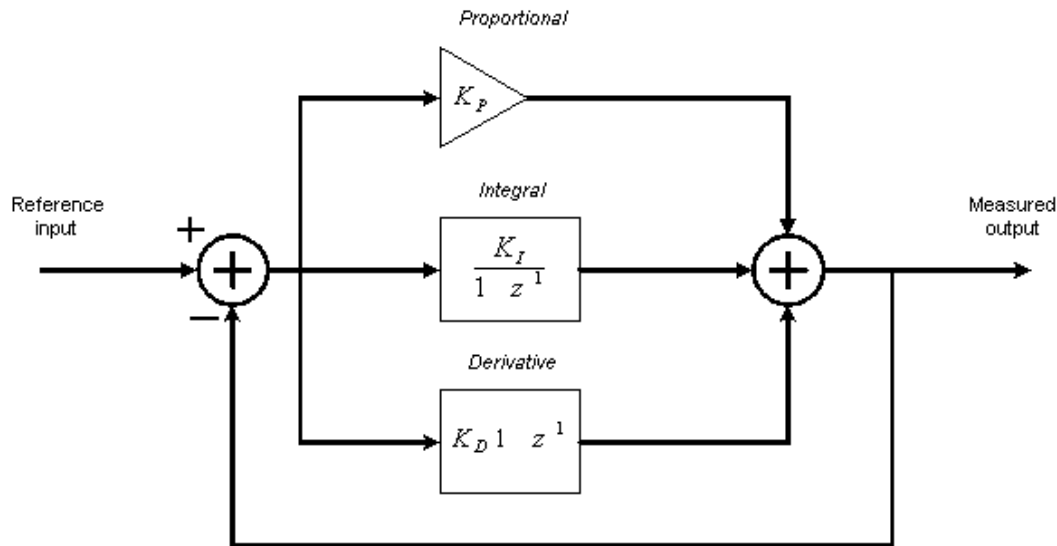
end of SinCos group

A Proportional Integral Derivative (PID) controller is a generic feedback control loop mechanism widely used in industrial control systems. A PID controller is the most commonly used type of feedback controller.

This set of functions implements (PID) controllers for Q15, Q31, and floating-point data types. The functions operate on a single sample of data and each call to the function returns a single processed value. `S` points to an instance of the PID control data structure. `in` is the input sample value. The functions return the output value.

#### Algorithm:

where  $K_p$  is proportional constant,  $K_i$  is Integral constant and  $K_d$  is Derivative constant



The PID controller calculates an “error” value as the difference between the measured output and the reference input. The controller attempts to minimize the error by adjusting the process control inputs. The proportional value determines the reaction to the current error, the integral value determines the reaction based on the sum of recent errors, and the derivative value determines the reaction based on the rate at which the error has been changing.

**Instance Structure** The Gains A0, A1, A2 and state variables for a PID controller are stored together in an instance data structure. A separate instance structure must be defined for each PID Controller. There are separate instance structure declarations for each of the 3 supported data types.

**Reset Functions** There is also an associated reset function for each data type which clears the state array.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Initializes the Gains A0, A1, A2 from Kp, Ki, Kd gains.
- Zeros out the values in the state buffer.

Instance structure cannot be placed into a const data section and it is recommended to use the initialization function.

**Fixed-Point Behavior** Care must be taken when using the fixed-point versions of the PID Controller functions. In particular, the overflow and saturation behavior of the accumulator used in each function must be considered. Refer to the function specific documentation below for usage guidelines.

## Functions

**\_\_STATIC\_FORCEINLINE float32\_t riscv\_pid\_f32(riscv\_pid\_instance\_f32 \* S, float32\_t in)**  
Process function for the floating-point PID Control.

**Return** processed output sample.

### Parameters

- [inout] S: is an instance of the floating-point PID Control structure
- [in] in: input sample to process

---

```
__STATIC_FORCEINLINE q31_t riscv_pid_q31(riscv_pid_instance_q31 * S, q31_t in)
```

Process function for the Q31 PID Control.

**Return** processed output sample.

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clip. In order to avoid overflows completely the input signal must be scaled down by 2 bits as there are four additions. After all multiply-accumulates are performed, the 2.62 accumulator is truncated to 1.32 format and then saturated to 1.31 format.

**Parameters**

- [inout] S: points to an instance of the Q31 PID Control structure
- [in] in: input sample to process

```
__STATIC_FORCEINLINE q15_t riscv_pid_q15(riscv_pid_instance_q15 * S, q15_t in)
```

Process function for the Q15 PID Control.

**Return** processed output sample.

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both Gains and state variables are represented in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. After all additions have been performed, the accumulator is truncated to 34.15 format by discarding low 15 bits. Lastly, the accumulator is saturated to yield a result in 1.15 format.

**Parameters**

- [inout] S: points to an instance of the Q15 PID Control structure
- [in] in: input sample to process

```
void riscv_pid_init_f32(riscv_pid_instance_f32 *S, int32_t resetStateFlag)
```

Initialization function for the floating-point PID Control.

**Return** none

**Details** The `resetStateFlag` specifies whether to set state to zero or not. The function computes the structure fields: A0, A1 A2 using the proportional gain(  $K_p$ ), integral gain(  $K_i$ ) and derivative gain(  $K_d$ ) also sets the state variables to all zeros.

**Parameters**

- [inout] S: points to an instance of the PID structure
- [in] resetStateFlag:
  - value = 0: no change in state
  - value = 1: reset state

```
void riscv_pid_init_q15(riscv_pid_instance_q15 *S, int32_t resetStateFlag)
```

Initialization function for the Q15 PID Control.

**Return** none

**Details** The `resetStateFlag` specifies whether to set state to zero or not. The function computes the structure fields: A0, A1 A2 using the proportional gain(  $K_p$ ), integral gain(  $K_i$ ) and derivative gain(  $K_d$ ) also sets the state variables to all zeros.

**Parameters**

- [inout] `S`: points to an instance of the Q15 PID structure
- [in] `resetStateFlag`:
  - value = 0: no change in state
  - value = 1: reset state

void **riscv\_pid\_init\_q31** (riscv\_pid\_instance\_q31 \**S*, int32\_t *resetStateFlag*)  
Initialization function for the Q31 PID Control.

**Return** none

**Details** The `resetStateFlag` specifies whether to set state to zero or not. The function computes the structure fields: A0, A1 A2 using the proportional gain(  $K_p$ ), integral gain(  $K_i$ ) and derivative gain(  $K_d$ ) also sets the state variables to all zeros.

**Parameters**

- [inout] `S`: points to an instance of the Q31 PID structure
- [in] `resetStateFlag`:
  - value = 0: no change in state
  - value = 1: reset state

void **riscv\_pid\_reset\_f32** (riscv\_pid\_instance\_f32 \**S*)  
Reset function for the floating-point PID Control.

**Return** none

**Details** The function resets the state buffer to zeros.

**Parameters**

- [inout] `S`: points to an instance of the floating-point PID structure

void **riscv\_pid\_reset\_q15** (riscv\_pid\_instance\_q15 \**S*)  
Reset function for the Q15 PID Control.

**Return** none

**Details** The function resets the state buffer to zeros.

**Parameters**

- [inout] `S`: points to an instance of the Q15 PID structure

void **riscv\_pid\_reset\_q31** (riscv\_pid\_instance\_q31 \**S*)  
Reset function for the Q31 PID Control.

**Return** none

**Details** The function resets the state buffer to zeros.

**Parameters**

- [inout] S: points to an instance of the Q31 PID structure

### Vector Park Transform

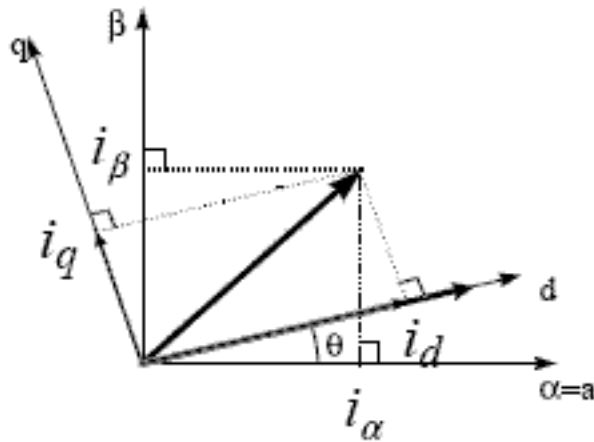
```
__STATIC_FORCEINLINE void riscv_park_f32(float32_t Ialpha, float32_t Ibeta, float32_t * pId,
```

```
__STATIC_FORCEINLINE void riscv_park_q31(q31_t Ialpha, q31_t Ibeta, q31_t * pId, q31_t * pIq,
```

```
group park
```

```
end of PID group
```

Forward Park transform converts the input two-coordinate vector to flux and torque components. The Park transform can be used to realize the transformation of the  $I_{\alpha}$  and the  $I_{\beta}$  currents from the stationary to the moving reference frame and control the spatial relationship between the stator vector current and rotor flux vector. If we consider the d axis aligned with the rotor flux, the diagram below shows the current vector and the relationship from the two reference frames: The function operates on a single sample of data and each call to the function returns the processed output. The library provides separate functions for Q31 and floating-point



data types.

#### Algorithm

where  $I_{\alpha}$  and  $I_{\beta}$  are the stator vector components,  $pId$  and  $pIq$  are rotor vector components and  $\cosVal$  and  $\sinVal$  are the cosine and sine values of  $\theta$  (rotor flux position).

$$pId = I_{\alpha} * \cosVal + I_{\beta} * \sinVal$$

$$pIq = -I_{\alpha} \sinVal + I_{\beta} * \cosVal$$

**Fixed-Point Behavior** Care must be taken when using the Q31 version of the Park transform. In particular, the overflow and saturation behavior of the accumulator used must be considered. Refer to the function specific documentation below for usage guidelines.

### Functions

```
__STATIC_FORCEINLINE void riscv_park_f32(float32_t Ialpha, float32_t Ibeta, float32_t * pId, float32_t * pIq)
```

Floating-point Park transform.

The function implements the forward Park transform.

**Return** none

**Parameters**

- [in] `Ialpha`: input two-phase vector coordinate alpha
- [in] `Ibeta`: input two-phase vector coordinate beta
- [out] `pId`: points to output rotor reference frame d
- [out] `pIq`: points to output rotor reference frame q
- [in] `sinVal`: sine value of rotation angle theta
- [in] `cosVal`: cosine value of rotation angle theta

```
__STATIC_FORCEINLINE void riscv_park_q31(q31_t Ialpha, q31_t Ibeta, q31_t * pId, q31_t * pIq, q31_t sinVal, q31_t cosVal)
    Park transform for Q31 version.
```

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 32-bit accumulator. The accumulator maintains 1.31 format by truncating lower 31 bits of the intermediate multiplication in 2.62 format. There is saturation on the addition and subtraction, hence there is no risk of overflow.

**Parameters**

- [in] `Ialpha`: input two-phase vector coordinate alpha
- [in] `Ibeta`: input two-phase vector coordinate beta
- [out] `pId`: points to output rotor reference frame d
- [out] `pIq`: points to output rotor reference frame q
- [in] `sinVal`: sine value of rotation angle theta
- [in] `cosVal`: cosine value of rotation angle theta

**Vector Inverse Park transform**

```
__STATIC_FORCEINLINE void riscv_inv_park_f32(float32_t Id, float32_t Iq, float32_t * pIalpha, float32_t * pIbeta, float32_t sinVal, float32_t cosVal)
__STATIC_FORCEINLINE void riscv_inv_park_q31(q31_t Id, q31_t Iq, q31_t * pIalpha, q31_t * pIbeta, q31_t sinVal, q31_t cosVal)
```

*group* **inv\_park**  
end of park group

Inverse Park transform converts the input flux and torque components to two-coordinate vector.

The function operates on a single sample of data and each call to the function returns the processed output. The library provides separate functions for Q31 and floating-point data types.

**Algorithm**

where `pIalpha` and `pIbeta` are the stator vector components, `Id` and `Iq` are rotor vector components and `cosVal` and `sinVal` are the cosine and sine values of theta (rotor flux position).

$$\begin{aligned} pIalpha &= Id * cosVal - Iq * sinVal \\ pIbeta &= Id * sinVal + Iq * cosVal \end{aligned}$$



**Fixed-Point Behavior** Care must be taken when using the Q31 version of the Park transform. In particular, the overflow and saturation behavior of the accumulator used must be considered. Refer to the function specific documentation below for usage guidelines.

## Functions

**\_\_STATIC\_FORCEINLINE void riscv\_inv\_park\_f32(float32\_t Id, float32\_t Iq, float32\_t \* pIalpha, float32\_t \* pIbeta)**  
Floating-point Inverse Park transform.

**Return** none

### Parameters

- [in] Id: input coordinate of rotor reference frame d
- [in] Iq: input coordinate of rotor reference frame q
- [out] pIalpha: points to output two-phase orthogonal vector axis alpha
- [out] pIbeta: points to output two-phase orthogonal vector axis beta
- [in] sinVal: sine value of rotation angle theta
- [in] cosVal: cosine value of rotation angle theta

**\_\_STATIC\_FORCEINLINE void riscv\_inv\_park\_q31(q31\_t Id, q31\_t Iq, q31\_t \* pIalpha, q31\_t \* pIbeta)**  
Inverse Park transform for Q31 version.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 32-bit accumulator. The accumulator maintains 1.31 format by truncating lower 31 bits of the intermediate multiplication in 2.62 format. There is saturation on the addition, hence there is no risk of overflow.

### Parameters

- [in] Id: input coordinate of rotor reference frame d
- [in] Iq: input coordinate of rotor reference frame q
- [out] pIalpha: points to output two-phase orthogonal vector axis alpha
- [out] pIbeta: points to output two-phase orthogonal vector axis beta
- [in] sinVal: sine value of rotation angle theta
- [in] cosVal: cosine value of rotation angle theta

## Vector Clarke Transform

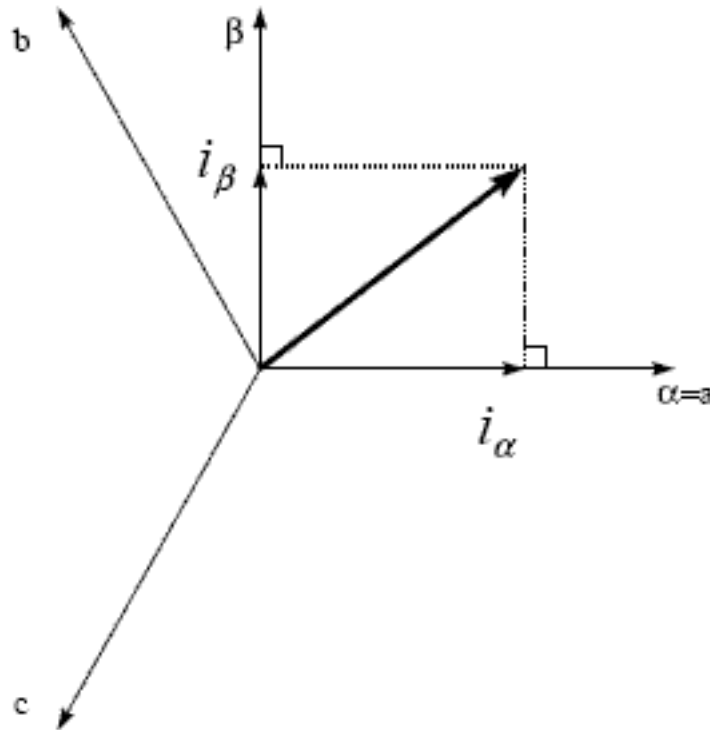
**\_\_STATIC\_FORCEINLINE void riscv\_clarke\_f32(float32\_t Ia, float32\_t Ib, float32\_t \* pIalpha, float32\_t \* pIbeta)**  
**\_\_STATIC\_FORCEINLINE void riscv\_clarke\_q31(q31\_t Ia, q31\_t Ib, q31\_t \* pIalpha, q31\_t \* pIbeta)**

*group clarke*

end of Inverse park group

Forward Clarke transform converts the instantaneous stator phases into a two-coordinate time invariant vector. Generally the Clarke transform uses three-phase currents Ia, Ib and Ic to calculate currents in the two-phase orthogonal stator axis Ialpha and Ibeta. When Ialpha is superposed with Ia as shown in the figure

below and  $I_a + I_b + I_c = 0$ , in this condition  $I_{\alpha}$  and  $I_{\beta}$  can be calculated using only  $I_a$  and



$I_b$ .

The function operates on a single sample of data and each call to the function returns the processed output. The library provides separate functions for Q31 and floating-point data types.

#### Algorithm

where  $I_a$  and  $I_b$  are the instantaneous stator phases and  $pI_{\alpha}$  and  $pI_{\beta}$  are the two coordinates

$$pI_{\alpha} = I_a$$

$$pI_{\beta} = (1/\sqrt{3}) I_a + (2/\sqrt{3}) I_b$$

of time invariant vector.

**Fixed-Point Behavior** Care must be taken when using the Q31 version of the Clarke transform. In particular, the overflow and saturation behavior of the accumulator used must be considered. Refer to the function specific documentation below for usage guidelines.

#### Functions

**\_\_STATIC\_FORCEINLINE void riscv\_clarke\_f32(float32\_t Ia, float32\_t Ib, float32\_t \* pIa)**  
Floating-point Clarke transform.

**Return** none

#### Parameters

- [in] Ia: input three-phase coordinate a
- [in] Ib: input three-phase coordinate b

- [out] pIalpha: points to output two-phase orthogonal vector axis alpha
- [out] pIbeta: points to output two-phase orthogonal vector axis beta

**\_\_STATIC\_FORCEINLINE void riscv\_clarke\_q31(q31\_t Ia, q31\_t Ib, q31\_t \* pIalpha, q31\_t \* pIbeta)**  
Clarke transform for Q31 version.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 32-bit accumulator. The accumulator maintains 1.31 format by truncating lower 31 bits of the intermediate multiplication in 2.62 format. There is saturation on the addition, hence there is no risk of overflow.

**Parameters**

- [in] Ia: input three-phase coordinate a
- [in] Ib: input three-phase coordinate b
- [out] pIalpha: points to output two-phase orthogonal vector axis alpha
- [out] pIbeta: points to output two-phase orthogonal vector axis beta

### Vector Inverse Clarke Transform

**\_\_STATIC\_FORCEINLINE void riscv\_inv\_clarke\_f32(float32\_t Ialpha, float32\_t Ibeta, float32\_t \* pIa, float32\_t \* pIb)**

**\_\_STATIC\_FORCEINLINE void riscv\_inv\_clarke\_q31(q31\_t Ialpha, q31\_t Ibeta, q31\_t \* pIa, q31\_t \* pIb)**

*group* **inv\_clarke**

end of clarke group

Inverse Clarke transform converts the two-coordinate time invariant vector into instantaneous stator phases.

The function operates on a single sample of data and each call to the function returns the processed output. The library provides separate functions for Q31 and floating-point data types.

**Algorithm**

where pIa and pIb are the instantaneous stator phases and Ialpha and Ibeta are the two coordinates

$$pIa = Ialpha$$

$$pIb = (-1/2) Ialpha + (\sqrt{3}/2) Ibeta$$

of time invariant vector.

**Fixed-Point Behavior** Care must be taken when using the Q31 version of the Clarke transform. In particular, the overflow and saturation behavior of the accumulator used must be considered. Refer to the function specific documentation below for usage guidelines.

### Functions

**\_\_STATIC\_FORCEINLINE void riscv\_inv\_clarke\_f32(float32\_t Ialpha, float32\_t Ibeta, float32\_t \* pIa, float32\_t \* pIb)**  
Floating-point Inverse Clarke transform.

**Return** none

**Parameters**

- [in] Ialpha: input two-phase orthogonal vector axis alpha

- [in] *Ibeta*: input two-phase orthogonal vector axis beta
- [out] *pIa*: points to output three-phase coordinate a
- [out] *pIb*: points to output three-phase coordinate b

**\_\_STATIC\_FORCEINLINE void riscv\_inv\_clarke\_q31**(q31\_t *Ialpha*, q31\_t *Ibeta*, q31\_t \* *pIa*,  
Inverse Clarke transform for Q31 version.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 32-bit accumulator. The accumulator maintains 1.31 format by truncating lower 31 bits of the intermediate multiplication in 2.62 format. There is saturation on the subtraction, hence there is no risk of overflow.

**Parameters**

- [in] *Ialpha*: input two-phase orthogonal vector axis alpha
- [in] *Ibeta*: input two-phase orthogonal vector axis beta
- [out] *pIa*: points to output three-phase coordinate a
- [out] *pIb*: points to output three-phase coordinate b

## Sine Cosine

void **riscv\_sin\_cos\_f32** (float32\_t *theta*, float32\_t \**pSinVal*, float32\_t \**pCosVal*)

void **riscv\_sin\_cos\_q31** (q31\_t *theta*, q31\_t \**pSinVal*, q31\_t \**pCosVal*)

*group* **SinCos**

Computes the trigonometric sine and cosine values using a combination of table lookup and linear interpolation. There are separate functions for Q31 and floating-point data types. The input to the floating-point version is in degrees while the fixed-point Q31 have a scaled input with the range [-1 0.9999] mapping to [-180 +180] degrees.

The floating point function also allows values that are out of the usual range. When this happens, the function will take extra time to adjust the input value to the range of [-180 180].

The result is accurate to 5 digits after the decimal point.

The implementation is based on table lookup using 360 values together with linear interpolation. The steps used are:

1. Calculation of the nearest integer table index.
2. Compute the fractional portion (fract) of the input.
3. Fetch the value corresponding to *index* from sine table to *y0* and also value from *index+1* to *y1*.
4. Sine value is computed as  $*psinVal = y0 + (fract * (y1 - y0))$ .
5. Fetch the value corresponding to *index* from cosine table to *y0* and also value from *index+1* to *y1*.
6. Cosine value is computed as  $*pcosVal = y0 + (fract * (y1 - y0))$ .

## Functions

void **riscv\_sin\_cos\_f32** (float32\_t *theta*, float32\_t \**pSinVal*, float32\_t \**pCosVal*)

Floating-point sin\_cos function.

**Return** none

**Parameters**

- [in] *theta*: input value in degrees
- [out] *pSinVal*: points to the processed sine output.
- [out] *pCosVal*: points to the processed cos output.
- [in] *theta*: input value in degrees
- [out] *pSinVal*: points to processed sine output
- [out] *pCosVal*: points to processed cosine output

void **riscv\_sin\_cos\_q31** (q31\_t *theta*, q31\_t \**pSinVal*, q31\_t \**pCosVal*)  
Q31 sin\_cos function.

The Q31 input value is in the range [-1 0.999999] and is mapped to a degree value in the range [-180 179].

**Return** none

**Parameters**

- [in] *theta*: scaled input value in degrees
- [out] *pSinVal*: points to the processed sine output.
- [out] *pCosVal*: points to the processed cosine output.
- [in] *theta*: scaled input value in degrees
- [out] *pSinVal*: points to processed sine output
- [out] *pCosVal*: points to processed cosine output

*group* **groupController**

### 3.3.6 Distance functions

#### Float Distances

##### Bray-Curtis distance

float16\_t **riscv\_braycurtis\_distance\_f16** (const float16\_t \**pA*, const float16\_t \**pB*, uint32\_t *blockSize*)

float32\_t **riscv\_braycurtis\_distance\_f32** (const float32\_t \**pA*, const float32\_t \**pB*, uint32\_t *blockSize*)

*group* **braycurtis**

Bray-Curtis distance between two vectors

#### Functions

float16\_t **riscv\_braycurtis\_distance\_f16** (const float16\_t \**pA*, const float16\_t \**pB*,  
uint32\_t *blockSize*)

Bray-Curtis distance between two vectors.

**Return** distance

**Parameters**

- [in] *pA*: First vector

- [in] pB: Second vector
- [in] blockSize: vector length

float32\_t **riscv\_braycurtis\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

Bray-Curtis distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

## Canberra distance

float16\_t **riscv\_canberra\_distance\_f16**(const float16\_t \*pA, const float16\_t \*pB, uint32\_t blockSize)

float32\_t **riscv\_canberra\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

*group* **Canberra**  
Canberra distance

## Functions

float16\_t **riscv\_canberra\_distance\_f16**(const float16\_t \*pA, const float16\_t \*pB, uint32\_t blockSize)

Canberra distance between two vectors.

This function may divide by zero when samples pA[i] and pB[i] are both zero. The result of the computation will be correct. So the division per zero may be ignored.

**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

float32\_t **riscv\_canberra\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

Canberra distance between two vectors.

This function may divide by zero when samples pA[i] and pB[i] are both zero. The result of the computation will be correct. So the division per zero may be ignored.

**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector

- [in] blockSize: vector length

### Chebyshev distance

float16\_t **riscv\_chebyshev\_distance\_f16**(const float16\_t \*pA, const float16\_t \*pB, uint32\_t blockSize)

float32\_t **riscv\_chebyshev\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

group **Chebyshev**  
Chebyshev distance

### Functions

float16\_t **riscv\_chebyshev\_distance\_f16**(const float16\_t \*pA, const float16\_t \*pB, uint32\_t blockSize)

Chebyshev distance between two vectors.

**Return** distance

#### Parameters

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

float32\_t **riscv\_chebyshev\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

Chebyshev distance between two vectors.

**Return** distance

#### Parameters

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

### Cityblock (Manhattan) distance

float16\_t **riscv\_cityblock\_distance\_f16**(const float16\_t \*pA, const float16\_t \*pB, uint32\_t blockSize)

float32\_t **riscv\_cityblock\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

group **Manhattan**  
Cityblock (Manhattan) distance

## Functions

`float16_t riscv_cityblock_distance_f16 (const float16_t *pA, const float16_t *pB, uint32_t blockSize)`

Cityblock (Manhattan) distance between two vectors.

**Return** distance

### Parameters

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

`float32_t riscv_cityblock_distance_f32 (const float32_t *pA, const float32_t *pB, uint32_t blockSize)`

Cityblock (Manhattan) distance between two vectors.

**Return** distance

### Parameters

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

## Correlation distance

`float16_t riscv_correlation_distance_f16 (float16_t *pA, float16_t *pB, uint32_t blockSize)`

`float32_t riscv_correlation_distance_f32 (float32_t *pA, float32_t *pB, uint32_t blockSize)`

*group* **Correlation**

Correlation distance

## Functions

`float16_t riscv_correlation_distance_f16 (float16_t *pA, float16_t *pB, uint32_t blockSize)`

Correlation distance between two vectors.

The input vectors are modified in place !

**Return** distance

### Parameters

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

`float32_t riscv_correlation_distance_f32 (float32_t *pA, float32_t *pB, uint32_t blockSize)`

Correlation distance between two vectors.

The input vectors are modified in place !



**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

### Cosine distance

float16\_t **riscv\_cosine\_distance\_f16**(const float16\_t \*pA, const float16\_t \*pB, uint32\_t blockSize)

float32\_t **riscv\_cosine\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

group **CosineDist**  
Cosine distance

### Functions

float16\_t **riscv\_cosine\_distance\_f16**(const float16\_t \*pA, const float16\_t \*pB, uint32\_t blockSize)

Cosine distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

float32\_t **riscv\_cosine\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

Cosine distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

### Euclidean distance

float16\_t **riscv\_euclidean\_distance\_f16**(const float16\_t \*pA, const float16\_t \*pB, uint32\_t blockSize)

float32\_t **riscv\_euclidean\_distance\_f32**(const float32\_t \*pA, const float32\_t \*pB, uint32\_t blockSize)

group **Euclidean**  
Euclidean distance

## Functions

`float16_t riscv_euclidean_distance_f16(const float16_t *pA, const float16_t *pB, uint32_t blockSize)`

Euclidean distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

`float32_t riscv_euclidean_distance_f32(const float32_t *pA, const float32_t *pB, uint32_t blockSize)`

Euclidean distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

## Jensen-Shannon distance

`float16_t riscv_jensenshannon_distance_f16(const float16_t *pA, const float16_t *pB, uint32_t blockSize)`

`float32_t riscv_jensenshannon_distance_f32(const float32_t *pA, const float32_t *pB, uint32_t blockSize)`

*group* **JensenShannon**

Jensen-Shannon distance

## Functions

`__STATIC_INLINE float16_t rel ENTR(float16_t x, float16_t y)`

`float16_t riscv_jensenshannon_distance_f16(const float16_t *pA, const float16_t *pB, uint32_t blockSize)`

Jensen-Shannon distance between two vectors.

This function is assuming that elements of second vector are  $> 0$  and 0 only when the corresponding element of first vector is 0. Otherwise the result of the computation does not make sense and for speed reasons, the cases returning NaN or Infinity are not managed.

When the function is computing  $x \log(x / y)$  with  $x == 0$  and  $y == 0$ , it will compute the right result (0) but a division by zero will occur and should be ignored in client code.

**Return** distance

**Parameters**

- [in] pA: First vector

- [in] pB: Second vector
- [in] blockSize: vector length

**\_\_STATIC\_INLINE float32\_t rel\_entr(float32\_t x, float32\_t y)**

**float32\_t riscv\_jensenshannon\_distance\_f32** (**const** float32\_t \*pA, **const** float32\_t \*pB, uint32\_t blockSize)

Jensen-Shannon distance between two vectors.

This function is assuming that elements of second vector are > 0 and 0 only when the corresponding element of first vector is 0. Otherwise the result of the computation does not make sense and for speed reasons, the cases returning NaN or Infinity are not managed.

When the function is computing  $x \log(x/y)$  with  $x == 0$  and  $y == 0$ , it will compute the right result (0) but a division by zero will occur and should be ignored in client code.

**Return** distance

#### Parameters

- [in] pA: First vector
- [in] pB: Second vector
- [in] blockSize: vector length

### Minkowski distance

**float16\_t riscv\_minkowski\_distance\_f16** (**const** float16\_t \*pA, **const** float16\_t \*pB, int32\_t order, uint32\_t blockSize)

**float32\_t riscv\_minkowski\_distance\_f32** (**const** float32\_t \*pA, **const** float32\_t \*pB, int32\_t order, uint32\_t blockSize)

*group* **Minkowski**  
Minkowski distance

#### Functions

**float16\_t riscv\_minkowski\_distance\_f16** (**const** float16\_t \*pA, **const** float16\_t \*pB, int32\_t order, uint32\_t blockSize)

Minkowski distance between two vectors.

**Return** distance

#### Parameters

- [in] pA: First vector
- [in] pB: Second vector
- [in] order: Distance order
- [in] blockSize: Number of samples

**float32\_t riscv\_minkowski\_distance\_f32** (**const** float32\_t \*pA, **const** float32\_t \*pB, int32\_t order, uint32\_t blockSize)

Minkowski distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector
- [in] pB: Second vector
- [in] order: Distance order
- [in] blockSize: Number of samples

**group FloatDist**

Distances between two vectors of float values.

**Boolean Distances**

float32\_t **riscv\_dice\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

float32\_t **riscv\_hamming\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

float32\_t **riscv\_jaccard\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

float32\_t **riscv\_kulsinski\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

float32\_t **riscv\_rogerstanimoto\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

float32\_t **riscv\_russellrao\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

float32\_t **riscv\_sokalmichener\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

float32\_t **riscv\_sokalsneath\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

float32\_t **riscv\_yule\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

**group BoolDist**

Distances between two vectors of boolean values.

Booleans are packed in 32 bit words. numberOfBooleans argument is the number of booleans and not the number of words.

Bits are packed in big-endian mode (because of behavior of numpy packbits in in version < 1.17)

**Unnamed Group**

float32\_t **riscv\_dice\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

Dice distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

## Functions

float32\_t **riscv\_hamming\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t *numberOfBools*)

Hamming distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

float32\_t **riscv\_jaccard\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t *numberOfBools*)

Jaccard distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

float32\_t **riscv\_kulsinski\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t *numberOfBools*)

Kulsinski distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

float32\_t **riscv\_rogerstanimoto\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t *numberOfBools*)

Rogers Tanimoto distance between two vectors.

Roger Stanimoto distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

float32\_t **riscv\_russellrao\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t *numberOfBools*)

Russell-Rao distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

float32\_t **riscv\_sokalmichener\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

Sokal-Michener distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

float32\_t **riscv\_sokalsneath\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

Sokal-Sneath distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

float32\_t **riscv\_yule\_distance** (const uint32\_t \*pA, const uint32\_t \*pB, uint32\_t numberOfBools)

Yule distance between two vectors.

**Return** distance

**Parameters**

- [in] pA: First vector of packed booleans
- [in] pB: Second vector of packed booleans
- [in] numberOfBools: Number of booleans

*group* **groupDistance**

Distance functions for use with clustering algorithms. There are distance functions for float vectors and boolean vectors.

### 3.3.7 Fast Math Functions

#### Cosine

float32\_t **riscv\_cos\_f32** (float32\_t x)

q31\_t **riscv\_cos\_q31** (q31\_t x)

q15\_t **riscv\_cos\_q15** (q15\_t x)

*group* **cos**

Computes the trigonometric cosine function using a combination of table lookup and linear interpolation. There are separate functions for Q15, Q31, and floating-point data types. The input to the floating-point version is in radians while the fixed-point Q15 and Q31 have a scaled input with the range [0 +0.9999] mapping to [0 2\*pi). The fixed-point range is chosen so that a value of 2\*pi wraps around to 0.

The implementation is based on table lookup using 512 values together with linear interpolation. The steps used are:

1. Calculation of the nearest integer table index
2. Compute the fractional portion (fract) of the table index.
3. The final result equals  $(1.0f - \text{fract}) * a + \text{fract} * b;$

where

end of sin group

## Functions

float32\_t **riscv\_cos\_f32** (float32\_t x)

Fast approximation to the trigonometric cosine function for floating-point data.

**Return** cos(x).

**Return** cos(x)

**Parameters**

- [in] x: input value in radians.

**Parameters**

- [in] x: input value in radians

q31\_t **riscv\_cos\_q31** (q31\_t x)

Fast approximation to the trigonometric cosine function for Q31 data.

The Q31 input value is in the range [0 +0.9999] and is mapped to a radian value in the range [0 2\*PI).

**Return** cos(x).

**Return** cos(x)

**Parameters**

- [in] x: Scaled input value in radians.

**Parameters**

- [in] x: Scaled input value in radians

q15\_t **riscv\_cos\_q15** (q15\_t x)

Fast approximation to the trigonometric cosine function for Q15 data.

The Q15 input value is in the range [0 +0.9999] and is mapped to a radian value in the range [0 2\*PI).

**Return** cos(x).

**Return** cos(x)

**Parameters**

- [in] *x*: Scaled input value in radians.

**Parameters**

- [in] *x*: Scaled input value in radians

**Fixed point division**

riscv\_status **riscv\_divide\_q15** (q15\_t *numerator*, q15\_t *denominator*, q15\_t \**quotient*, int16\_t \**shift*)

*group divide*

**Functions**

riscv\_status **riscv\_divide\_q15** (q15\_t *numerator*, q15\_t *denominator*, q15\_t \**quotient*, int16\_t \**shift*)

Fixed point division.

When dividing by 0, an error RISCVC\_MATH\_NANINF is returned. And the quotient is forced to the saturated negative or positive value.

**Return** error status

**Parameters**

- [in] *numerator*: Numerator
- [in] *denominator*: Denominator
- [out] *quotient*: Quotient value normalized between -1.0 and 1.0
- [out] *shift*: Shift left value to get the unnormalized quotient

**Sine**

float32\_t **riscv\_sin\_f32** (float32\_t *x*)

q31\_t **riscv\_sin\_q31** (q31\_t *x*)

q15\_t **riscv\_sin\_q15** (q15\_t *x*)

*group sin*

Computes the trigonometric sine function using a combination of table lookup and linear interpolation. There are separate functions for Q15, Q31, and floating-point data types. The input to the floating-point version is in radians while the fixed-point Q15 and Q31 have a scaled input with the range [0 +0.9999] mapping to [0 2\*pi). The fixed-point range is chosen so that a value of 2\*pi wraps around to 0.

The implementation is based on table lookup using 512 values together with linear interpolation. The steps used are:

1. Calculation of the nearest integer table index
2. Compute the fractional portion (fract) of the table index.
3. The final result equals  $(1.0f - \text{fract}) * a + \text{fract} * b;$

where



## Functions

`float32_t riscv_sin_f32 (float32_t x)`

Fast approximation to the trigonometric sine function for floating-point data.

**Return**  $\sin(x)$ .

**Return**  $\sin(x)$

**Parameters**

- [in] `x`: input value in radians.

**Parameters**

- [in] `x`: input value in radians.

`q31_t riscv_sin_q31 (q31_t x)`

Fast approximation to the trigonometric sine function for Q31 data.

The Q31 input value is in the range  $[0 +0.9999]$  and is mapped to a radian value in the range  $[0 \ 2\pi]$ .

**Return**  $\sin(x)$ .

**Return**  $\sin(x)$

**Parameters**

- [in] `x`: Scaled input value in radians.

**Parameters**

- [in] `x`: Scaled input value in radians

`q15_t riscv_sin_q15 (q15_t x)`

Fast approximation to the trigonometric sine function for Q15 data.

The Q15 input value is in the range  $[0 +0.9999]$  and is mapped to a radian value in the range  $[0 \ 2\pi]$ .

**Return**  $\sin(x)$ .

**Return**  $\sin(x)$

**Parameters**

- [in] `x`: Scaled input value in radians.

**Parameters**

- [in] `x`: Scaled input value in radians

## Square Root

```
__STATIC_FORCEINLINE riscv_status riscv_sqrt_f32(float32_t in, float32_t * pOut)
```

```
riscv_status riscv_sqrt_q31(q31_t in, q31_t *pOut)
```

```
riscv_status riscv_sqrt_q15(q15_t in, q15_t *pOut)
```

```
void riscv_vsqrtd_f32(float32_t *pIn, float32_t *pOut, uint16_t len)
```

```
void riscv_vsqrtd_q31(q31_t *pIn, q31_t *pOut, uint16_t len)
```

```
void riscv_vsqrtd_q15(q15_t *pIn, q15_t *pOut, uint16_t len)
```

```
__STATIC_FORCEINLINE riscv_status riscv_sqrt_f16(float16_t in, float16_t * pOut)
```

*group* **SQRT**

Computes the square root of a number. There are separate functions for Q15, Q31, and floating-point data types. The square root function is computed using the Newton-Raphson algorithm. This is an iterative algorithm of the form: where  $x_1$  is the current estimate,  $x_0$  is the previous estimate, and  $f'(x_0)$  is the derivative of  $f()$  evaluated at  $x_0$ . For the square root function, the algorithm reduces to:

**Functions**

**\_\_STATIC\_FORCEINLINE riscv\_status riscv\_sqrt\_f32(float32\_t in, float32\_t \* pOut)**  
Floating-point square root function.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : input value is positive
- RISC\_V\_MATH\_ARGUMENT\_ERROR : input value is negative; \*pOut is set to 0

**Parameters**

- [in] in: input value
- [out] pOut: square root of input value

riscv\_status **riscv\_sqrt\_q31**(q31\_t in, q31\_t \*pOut)  
Q31 square root function.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : input value is positive
- RISC\_V\_MATH\_ARGUMENT\_ERROR : input value is negative; \*pOut is set to 0

**Parameters**

- [in] in: input value. The range of the input value is [0 +1) or 0x00000000 to 0x7FFFFFFF
- [out] pOut: points to square root of input value

riscv\_status **riscv\_sqrt\_q15**(q15\_t in, q15\_t \*pOut)  
Q15 square root function.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : input value is positive
- RISC\_V\_MATH\_ARGUMENT\_ERROR : input value is negative; \*pOut is set to 0

**Parameters**

- [in] in: input value. The range of the input value is [0 +1) or 0x0000 to 0x7FFF
- [out] pOut: points to square root of input value

void **riscv\_vsqrt\_f32**(float32\_t \*pIn, float32\_t \*pOut, uint16\_t len)  
Vector Floating-point square root function.

**Return** The function returns RISC\_V\_MATH\_SUCCESS if input value is positive value or RISC\_V\_MATH\_ARGUMENT\_ERROR if in is negative value and returns zero output for negative values.

**Parameters**

- [in] `pIn`: input vector.
- [out] `pOut`: vector of square roots of input elements.
- [in] `len`: length of input vector.

void **riscv\_vsqr\_t\_q31** (q31\_t \**pIn*, q31\_t \**pOut*, uint16\_t *len*)

void **riscv\_vsqr\_t\_q15** (q15\_t \**pIn*, q15\_t \**pOut*, uint16\_t *len*)

**\_\_STATIC\_FORCEINLINE riscv\_status riscv\_sqrt\_f16**(float16\_t *in*, float16\_t \* *pOut*)  
Floating-point square root function.

**Return** execution status

- **RISCV\_MATH\_SUCCESS** : input value is positive
- **RISCV\_MATH\_ARGUMENT\_ERROR** : input value is negative; \**pOut* is set to 0

**Parameters**

- [in] *in*: input value
- [out] *pOut*: square root of input value

#### group **groupFastMath**

This set of functions provides a fast approximation to sine, cosine, and square root. As compared to most of the other functions in the NMSIS math library, the fast math functions operate on individual values and not arrays. There are separate functions for Q15, Q31, and floating-point data.

### 3.3.8 Filtering Functions

#### High Precision Q31 Biquad Cascade Filter

void **riscv\_biquad\_cas\_df1\_32x64\_init\_q31** (riscv\_biquad\_cas\_df1\_32x64\_ins\_q31 \**S*, uint8\_t *numStages*, **const** q31\_t \**pCoeffs*, q63\_t \**pState*, uint8\_t *postShift*)

void **riscv\_biquad\_cas\_df1\_32x64\_q31** (**const** riscv\_biquad\_cas\_df1\_32x64\_ins\_q31 \**S*, **const** q31\_t \**pSrc*, q31\_t \**pDst*, uint32\_t *blockSize*)

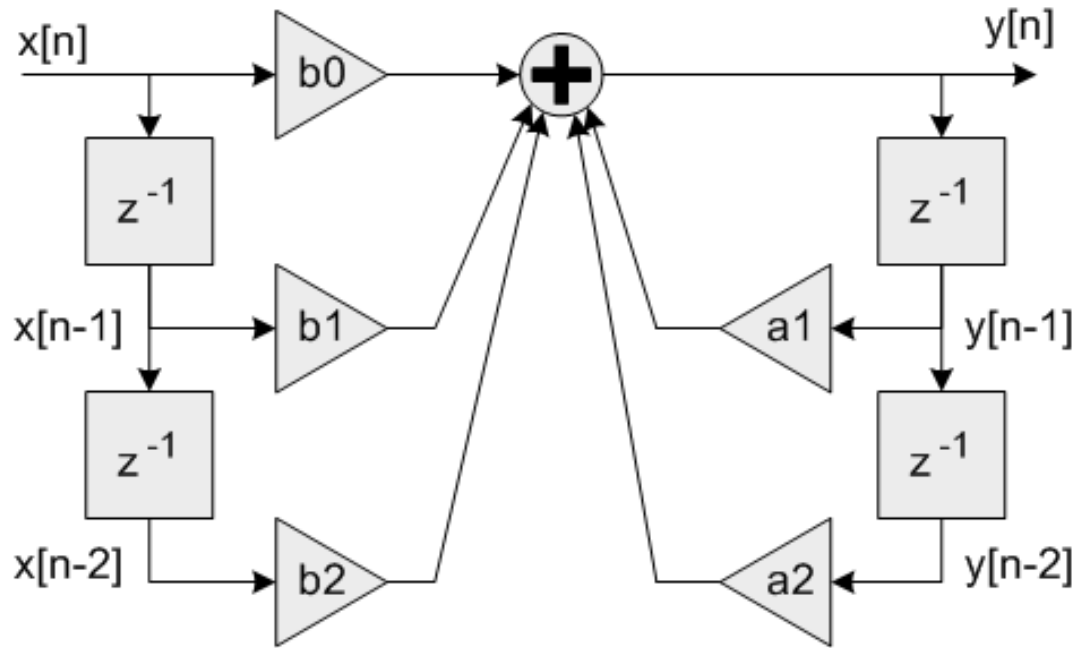
#### group **BiquadCascadeDF1\_32x64**

This function implements a high precision Biquad cascade filter which operates on Q31 data values. The filter coefficients are in 1.31 format and the state variables are in 1.63 format. The double precision state variables reduce quantization noise in the filter and provide a cleaner output. These filters are particularly useful when implementing filters in which the singularities are close to the unit circle. This is common for low pass or high pass filters with very low cutoff frequencies.

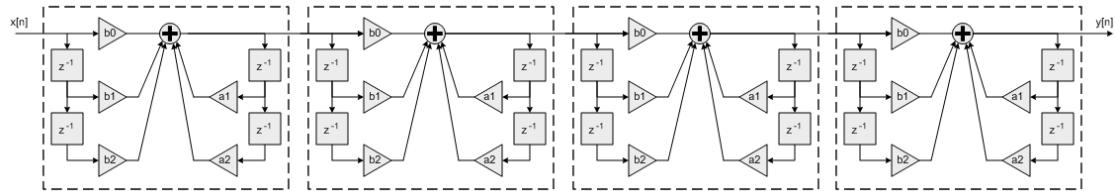
The function operates on blocks of input and output data and each call to the function processes `blockSize` samples through the filter. `pSrc` and `pDst` points to input and output arrays containing `blockSize` Q31 values.

**Algorithm** Each Biquad stage implements a second order filter using the difference equation: A Direct Form I algorithm is used with 5 coefficients and 4 state variables per stage. Coefficients `b0`, `b1` and `b2` multiply the input signal  $x[n]$  and are referred to as the feedforward coefficients. Coefficients `a1` and `a2` multiply the output signal  $y[n]$  and are referred to as the feedback coefficients. Pay careful attention to the sign of the feedback coefficients. Some design tools use the difference equation In this

case the feedback coefficients  $a1$  and  $a2$  must be negated when used with the NMSIS DSP Library.



Higher order filters are realized as a cascade of second order sections. `numStages` refers to the number of second order stages used. For example, an 8th order filter would be realized with `numStages=4` second order stages. A 9th order filter would be realized with `numStages=5` second order stages with the coefficients for one of the stages configured as a first order filter ( $b2=0$  and  $a2=0$ ).



The `pState` points to state variables array. Each Biquad stage has 4 state variables  $x[n-1]$ ,  $x[n-2]$ ,  $y[n-1]$ , and  $y[n-2]$  and each state variable in 1.63 format to improve precision. The state variables are arranged in the array as:

The 4 state variables for stage 1 are first, then the 4 state variables for stage 2, and so on. The state array has a total length of  $4 * \text{numStages}$  values of data in 1.63 format. The state variables are updated after each block of data is processed, the coefficients are untouched.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient arrays may be shared among several instances while state variable arrays cannot be shared.

**Init Function** There is also an associated initialization function which performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numStages`, `pCoeffs`, `postShift`, `pState`. Also set all of the values in `pState` to zero.

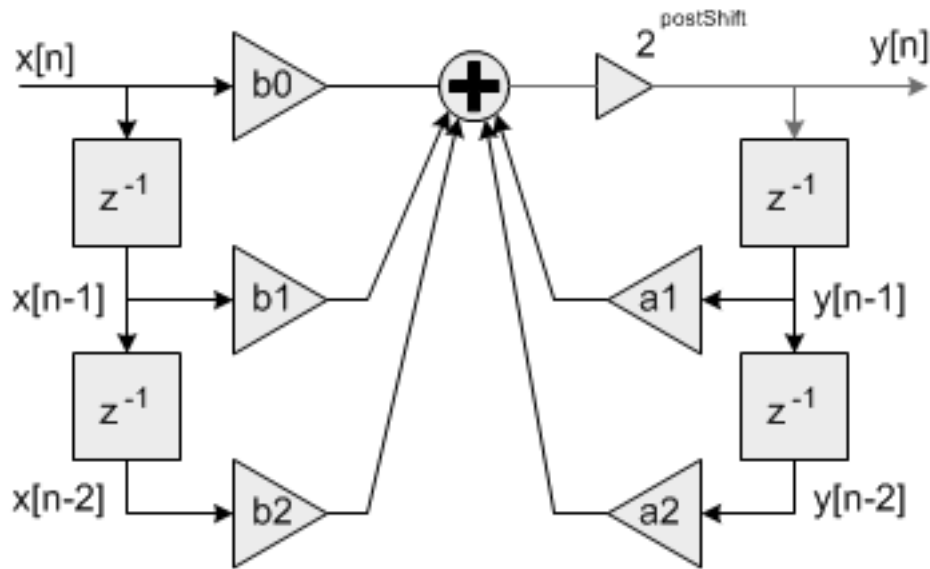
Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section,

the instance structure must be manually initialized. Set the values in the state buffer to zeros before static initialization. For example, to statically initialize the filter instance structure use where `numStages` is the number of Biquad stages in the filter; `pState` is the address of the state buffer; `pCoeffs` is the address of the coefficient buffer; `postShift` shift to be applied which is described in detail below.

**Fixed-Point Behavior** Care must be taken while using Biquad Cascade 32x64 filter function. Following issues must be considered:

- Scaling of coefficients
- Filter gain
- Overflow and saturation

Filter coefficients are represented as fractional values and restricted to lie in the range  $[-1 \text{ } +1)$ . The processing function has an additional scaling parameter `postShift` which allows the filter coefficients to exceed the range  $[-1 \text{ } +1)$ . At the output of the filter's accumulator is a shift register which shifts the result by `postShift` bits. This essentially scales the filter coefficients by  $2^{\text{postShift}}$ . For example, to realize the coefficients set the Coefficient array to: and set `postShift=1`



The second thing to keep in mind is the gain through the filter. The frequency response of a Biquad filter is a function of its coefficients. It is possible for the gain through the filter to exceed 1.0 meaning that the filter increases the amplitude of certain frequencies. This means that an input signal with amplitude  $< 1.0$  may result in an output  $> 1.0$  and these are saturated or overflowed based on the implementation of the filter. To avoid this behavior the filter needs to be scaled down such that its peak gain  $< 1.0$  or the input signal must be scaled down so that the combination of input and filter are never overflowed.

The third item to consider is the overflow and saturation behavior of the fixed-point Q31 version. This is described in the function specific documentation below.

## Functions

```
void riscv_biquad_cas_df1_32x64_init_q31 (riscv_biquad_cas_df1_32x64_ins_q31 *S,
                                         uint8_t numStages, const q31_t *pCoeffs,
                                         q63_t *pState, uint8_t postShift)
```

Initialization function for the Q31 Biquad cascade 32x64 filter.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array `pCoeffs` in the following order: where `b1x` and `a1x` are the coefficients for the first stage, `b2x` and `a2x` are the coefficients for the second stage, and so on. The `pCoeffs` array contains a total of `5*numStages` values.

The `pState` points to state variables array and size of each state variable is 1.63 format. Each Biquad stage has 4 state variables `x[n-1]`, `x[n-2]`, `y[n-1]`, and `y[n-2]`. The state variables are arranged in the state array as: The 4 state variables for stage 1 are first, then the 4 state variables for stage 2, and so on. The state array has a total length of `4*numStages` values. The state variables are updated after each block of data is processed; the coefficients are untouched.

#### Parameters

- `[inout]` `S`: points to an instance of the high precision Q31 Biquad cascade filter structure
- `[in]` `numStages`: number of 2nd order stages in the filter
- `[in]` `pCoeffs`: points to the filter coefficients
- `[in]` `pState`: points to the state buffer
- `[in]` `postShift`: Shift to be applied after the accumulator. Varies according to the coefficients format

```
void riscv_biquad_cas_df1_32x64_q31 (const riscv_biquad_cas_df1_32x64_ins_q31 *S,
                                     const q31_t *pSrc, q31_t *pDst, uint32_t blockSize)
```

Processing function for the Q31 Biquad cascade 32x64 filter.

**Return** none

**Details** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clip. In order to avoid overflows completely the input signal must be scaled down by 2 bits and lie in the range  $[-0.25 + 0.25]$ . After all 5 multiply-accumulates are performed, the 2.62 accumulator is shifted by `postShift` bits and the result truncated to 1.31 format by discarding the low 32 bits.

Two related functions are provided in the NMSIS DSP library.

- `riscv_biquad_cascade_df1_q31()` implements a Biquad cascade with 32-bit coefficients and state variables with a Q63 accumulator.
- `riscv_biquad_cascade_df1_fast_q31()` implements a Biquad cascade with 32-bit coefficients and state variables with a Q31 accumulator.

#### Parameters

- `[in]` `S`: points to an instance of the high precision Q31 Biquad cascade filter
- `[in]` `pSrc`: points to the block of input data
- `[out]` `pDst`: points to the block of output data
- `[in]` `blockSize`: number of samples to process

### Biquad Cascade IIR Filters Using Direct Form I Structure

```
void riscv_biquad_cascade_df1_f16 (const riscv_biquad_casd_df1_inst_f16 *S, const float16_t
                                   *pSrc, float16_t *pDst, uint32_t blockSize)
```

```
void riscv_biquad_cascade_df1_f32 (const riscv_biquad_casd_df1_inst_f32 *S, const float32_t
                                   *pSrc, float32_t *pDst, uint32_t blockSize)
```

```

void riscv_biquad_cascade_df1_fast_q15 (const riscv_biquad_casd_df1_inst_q15 *S, const
q15_t *pSrc, q15_t *pDst, uint32_t blockSize)
void riscv_biquad_cascade_df1_fast_q31 (const riscv_biquad_casd_df1_inst_q31 *S, const
q31_t *pSrc, q31_t *pDst, uint32_t blockSize)
void riscv_biquad_cascade_df1_init_f16 (riscv_biquad_casd_df1_inst_f16 *S, uint8_t num-
Stages, const float16_t *pCoeffs, float16_t *pState)
void riscv_biquad_cascade_df1_init_f32 (riscv_biquad_casd_df1_inst_f32 *S, uint8_t num-
Stages, const float32_t *pCoeffs, float32_t *pState)
void riscv_biquad_cascade_df1_init_q15 (riscv_biquad_casd_df1_inst_q15 *S, uint8_t num-
Stages, const q15_t *pCoeffs, q15_t *pState, int8_t
postShift)
void riscv_biquad_cascade_df1_init_q31 (riscv_biquad_casd_df1_inst_q31 *S, uint8_t num-
Stages, const q31_t *pCoeffs, q31_t *pState, int8_t
postShift)
void riscv_biquad_cascade_df1_q15 (const riscv_biquad_casd_df1_inst_q15 *S, const q15_t
*pSrc, q15_t *pDst, uint32_t blockSize)
void riscv_biquad_cascade_df1_q31 (const riscv_biquad_casd_df1_inst_q31 *S, const q31_t
*pSrc, q31_t *pDst, uint32_t blockSize)

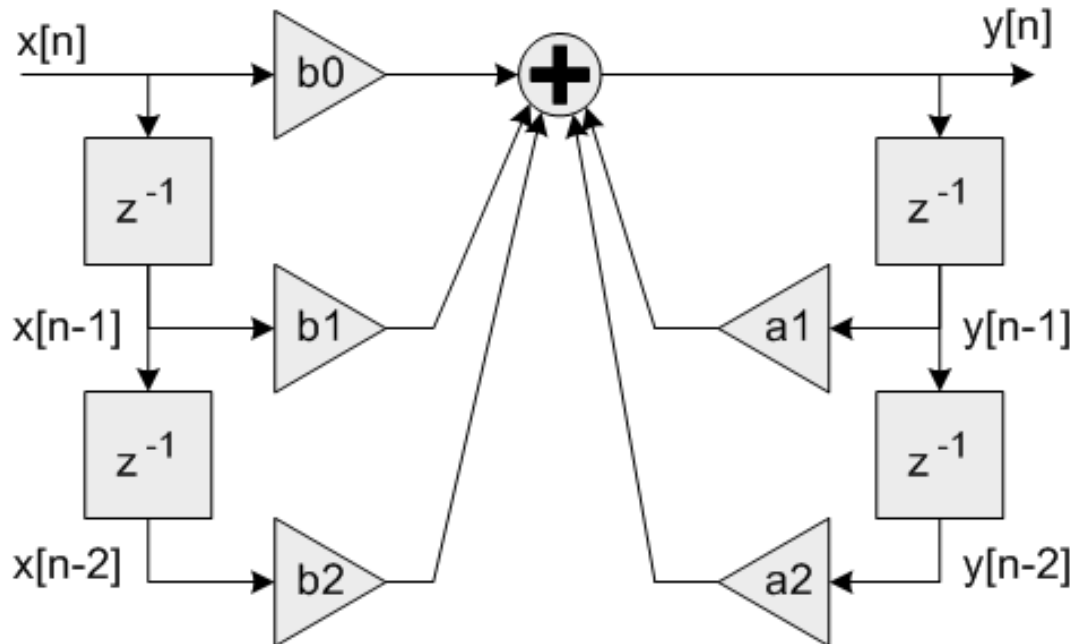
```

*group* **BiquadCascadeDF1**

This set of functions implements arbitrary order recursive (IIR) filters. The filters are implemented as a cascade of second order Biquad sections. The functions support Q15, Q31 and floating-point data types. Fast version of Q15 and Q31 also available.

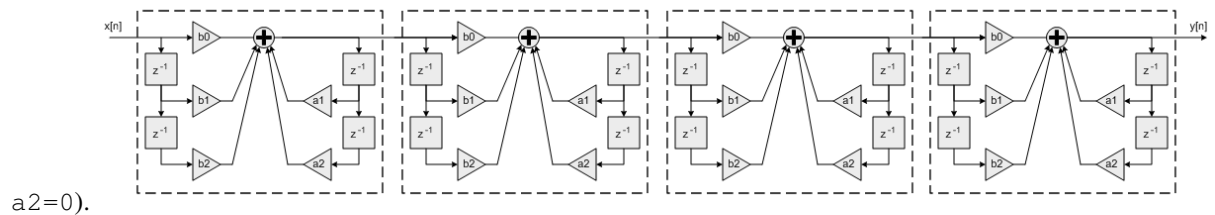
The functions operate on blocks of input and output data and each call to the function processes `blockSize` samples through the filter. `pSrc` points to the array of input data and `pDst` points to the array of output data. Both arrays contain `blockSize` values.

**Algorithm** Each Biquad stage implements a second order filter using the difference equation: A Direct Form I algorithm is used with 5 coefficients and 4 state variables per stage. Coefficients `b0`, `b1` and `b2` multiply the input signal  $x[n]$  and are referred to as the feedforward coefficients. Coefficients `a1` and `a2` multiply the output signal  $y[n]$  and are referred to as the feedback coefficients. Pay careful attention to the sign of the feedback coefficients. Some design tools use the difference equation  $y[n] = b_0x[n] + b_1x[n-1] + b_2x[n-2] - a_1y[n-1] - a_2y[n-2]$ . In this case the feedback coefficients `a1` and `a2` must be negated when used with the NMSIS DSP



Library.

Higher order filters are realized as a cascade of second order sections. `numStages` refers to the number of second order stages used. For example, an 8th order filter would be realized with `numStages=4` second order stages. A 9th order filter would be realized with `numStages=5` second order stages with the coefficients for one of the stages configured as a first order filter (`b2=0` and `a2=0`).



The `pState` points to state variables array. Each Biquad stage has 4 state variables  $x[n-1]$ ,  $x[n-2]$ ,  $y[n-1]$ , and  $y[n-2]$ . The state variables are arranged in the `pState` array as:

The 4 state variables for stage 1 are first, then the 4 state variables for stage 2, and so on. The state array has a total length of  $4 * \text{numStages}$  values. The state variables are updated after each block of data is processed, the coefficients are untouched.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient arrays may be shared among several instances while state variable arrays cannot be shared. There are separate instance structure declarations for each of the 3 supported data types.

**Init Function** There is also an associated initialization function for each data type. The initialization function performs following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numStages`, `pCoeffs`, `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section,

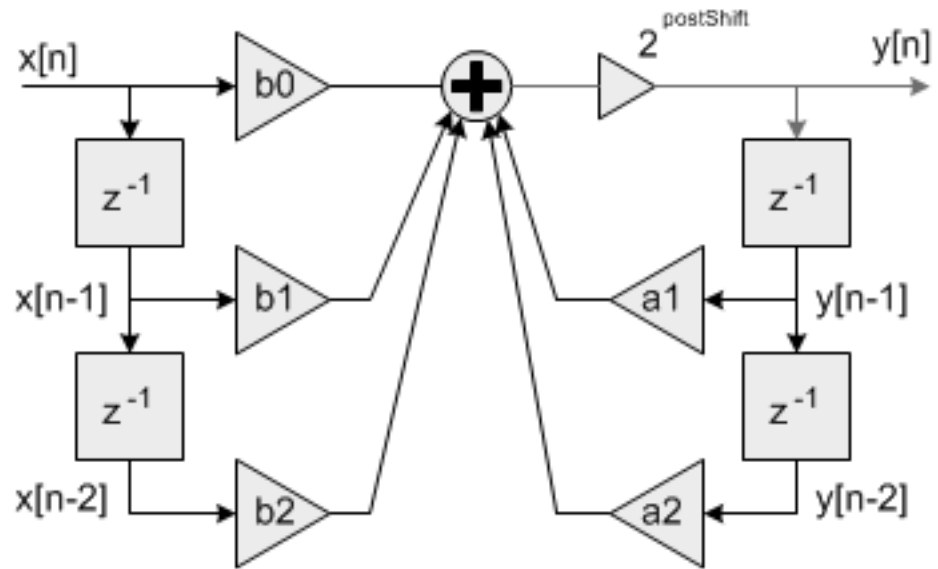


the instance structure must be manually initialized. Set the values in the state buffer to zeros before static initialization. The code below statically initializes each of the 3 different data type filter instance structures where `numStages` is the number of Biquad stages in the filter; `pState` is the address of the state buffer; `pCoeffs` is the address of the coefficient buffer; `postShift` shift to be applied.

**Fixed-Point Behavior** Care must be taken when using the Q15 and Q31 versions of the Biquad Cascade filter functions. Following issues must be considered:

- Scaling of coefficients
- Filter gain
- Overflow and saturation

**Scaling of coefficients** Filter coefficients are represented as fractional values and coefficients are restricted to lie in the range  $[-1 \ +1)$ . The fixed-point functions have an additional scaling parameter `postShift` which allow the filter coefficients to exceed the range  $[+1 \ -1)$ . At the output of the filter's accumulator is a shift register which shifts the result by `postShift` bits. This essentially scales the filter coefficients by  $2^{\text{postShift}}$ . For example, to realize the coefficients set the `pCoeffs` array to: and set



`postShift=1`

**Filter gain** The frequency response of a Biquad filter is a function of its coefficients. It is possible for the gain through the filter to exceed 1.0 meaning that the filter increases the amplitude of certain frequencies. This means that an input signal with amplitude  $< 1.0$  may result in an output  $> 1.0$  and these are saturated or overflowed based on the implementation of the filter. To avoid this behavior the filter needs to be scaled down such that its peak gain  $< 1.0$  or the input signal must be scaled down so that the combination of input and filter are never overflowed.

**Overflow and saturation** For Q15 and Q31 versions, it is described separately as part of the function specific documentation below.

## Functions

```
void riscv_biquad_cascade_df1_f16(const riscv_biquad_casd_df1_inst_f16 *S, const
                                float16_t *pSrc, float16_t *pDst, uint32_t blockSize)
    Processing function for the floating-point Biquad cascade filter.
```

**Return** none

**Parameters**

- [in] *S*: points to an instance of the floating-point Biquad cascade structure
- [in] *pSrc*: points to the block of input data
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process

```
void riscv_biquad_cascade_df1_f32 (const riscv_biquad_casd_df1_inst_f32 *S, const
                                float32_t *pSrc, float32_t *pDst, uint32_t blockSize)
```

Processing function for the floating-point Biquad cascade filter.

**Return** none

**Parameters**

- [in] *S*: points to an instance of the floating-point Biquad cascade structure
- [in] *pSrc*: points to the block of input data
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process

```
void riscv_biquad_cascade_df1_fast_q15 (const riscv_biquad_casd_df1_inst_q15 *S,
                                         const q15_t *pSrc, q15_t *pDst, uint32_t
                                         blockSize)
```

Processing function for the Q15 Biquad cascade filter (fast variant).

Fast but less precise processing function for the Q15 Biquad cascade filter for RISC-V Core with DSP enabled.

**Return** none

**Scaling and Overflow Behavior** This fast version uses a 32-bit accumulator with 2.30 format. The accumulator maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around and distorts the result. In order to avoid overflows completely the input signal must be scaled down by two bits and lie in the range [-0.25 +0.25). The 2.30 accumulator is then shifted by `postShift` bits and the result truncated to 1.15 format by discarding the low 16 bits.

**Remark** Refer to `riscv_biquad_cascade_df1_q15()` for a slower implementation of this function which uses 64-bit accumulation to avoid wrap around distortion. Both the slow and the fast versions use the same instance structure. Use the function `riscv_biquad_cascade_df1_init_q15()` to initialize the filter structure.

**Parameters**

- [in] *S*: points to an instance of the Q15 Biquad cascade structure
- [in] *pSrc*: points to the block of input data
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process per call

```
void riscv_biquad_cascade_df1_fast_q31 (const riscv_biquad_casd_df1_inst_q31 *S,
                                         const q31_t *pSrc, q31_t *pDst, uint32_t
                                         blockSize)
```

Processing function for the Q31 Biquad cascade filter (fast variant).

Fast but less precise processing function for the Q31 Biquad cascade filter for RISC-V Core with DSP enabled.

**Return** none

**Scaling and Overflow Behavior** This function is optimized for speed at the expense of fixed-point precision and overflow protection. The result of each 1.31 x 1.31 multiplication is truncated to 2.30 format. These intermediate results are added to a 2.30 accumulator. Finally, the accumulator is saturated and converted to a 1.31 result. The fast version has the same overflow behavior as the standard version and provides less precision since it discards the low 32 bits of each multiplication result. In order to avoid overflows completely the input signal must be scaled down by two bits and lie in the range  $[-0.25, +0.25]$ . Use the initialization function `riscv_biquad_cascade_df1_init_q31()` to initialize filter structure.

**Remark** Refer to `riscv_biquad_cascade_df1_q31()` for a slower implementation of this function which uses 64-bit accumulation to provide higher precision. Both the slow and the fast versions use the same instance structure. Use the function `riscv_biquad_cascade_df1_init_q31()` to initialize the filter structure.

#### Parameters

- [in] *S*: points to an instance of the Q31 Biquad cascade structure
- [in] *pSrc*: points to the block of input data
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process per call

```
void riscv_biquad_cascade_df1_init_f16 (riscv_biquad_casd_df1_inst_f16 *S, uint8_t num-
                                         Stages, const float16_t *pCoeffs, float16_t
                                         *pState)
```

Initialization function for the floating-point Biquad cascade filter.

The initialization function which must be used is `riscv_biquad_cascade_df1_mve_init_f16`.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array *pCoeffs* in the following order:

where *b1x* and *a1x* are the coefficients for the first stage, *b2x* and *a2x* are the coefficients for the second stage, and so on. The *pCoeffs* array contains a total of  $5 \times \text{numStages}$  values.

The *pState* is a pointer to state array. Each Biquad stage has 4 state variables  $x[n-1]$ ,  $x[n-2]$ ,  $y[n-1]$ , and  $y[n-2]$ . The state variables are arranged in the *pState* array as: The 4 state variables for stage 1 are first, then the 4 state variables for stage 2, and so on. The state array has a total length of  $4 \times \text{numStages}$  values. The state variables are updated after each block of data is processed; the coefficients are untouched.

**For MVE code, an additional buffer of modified coefficients is required.** Its size is *numStages* and each element of this buffer has type `riscv_biquad_mod_coef_f16`. So, its total size is  $96 \times \text{numStages}$  `float16_t` elements.

#### Parameters

- [inout] *S*: points to an instance of the floating-point Biquad cascade structure.
- [in] *numStages*: number of 2nd order stages in the filter.
- [in] *pCoeffs*: points to the filter coefficients.
- [in] *pState*: points to the state buffer.

```
void riscv_biquad_cascade_df1_init_f32 (riscv_biquad_casd_df1_inst_f32 *S, uint8_t num-
                                       Stages, const float32_t *pCoeffs, float32_t
                                       *pState)
```

Initialization function for the floating-point Biquad cascade filter.

The initialization function which must be used is `riscv_biquad_cascade_df1_mve_init_f32`.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array `pCoeffs` in the following order:

where `b1x` and `a1x` are the coefficients for the first stage, `b2x` and `a2x` are the coefficients for the second stage, and so on. The `pCoeffs` array contains a total of `5*numStages` values.

The `pState` is a pointer to state array. Each Biquad stage has 4 state variables `x[n-1]`, `x[n-2]`, `y[n-1]`, and `y[n-2]`. The state variables are arranged in the `pState` array as: The 4 state variables for stage 1 are first, then the 4 state variables for stage 2, and so on. The state array has a total length of `4*numStages` values. The state variables are updated after each block of data is processed; the coefficients are untouched.

**For MVE code, an additional buffer of modified coefficients is required.** Its size is `numStages` and each element of this buffer has type `riscv_biquad_mod_coef_f32`. So, its total size is `32*numStages` `float32_t` elements.

#### Parameters

- [inout] `S`: points to an instance of the floating-point Biquad cascade structure.
- [in] `numStages`: number of 2nd order stages in the filter.
- [in] `pCoeffs`: points to the filter coefficients.
- [in] `pState`: points to the state buffer.

```
void riscv_biquad_cascade_df1_init_q15 (riscv_biquad_casd_df1_inst_q15 *S, uint8_t
                                       numStages, const q15_t *pCoeffs, q15_t
                                       *pState, int8_t postShift)
```

Initialization function for the Q15 Biquad cascade filter.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array `pCoeffs` in the following order:

where `b1x` and `a1x` are the coefficients for the first stage, `b2x` and `a2x` are the coefficients for the second stage, and so on. The `pCoeffs` array contains a total of `6*numStages` values. The zero coefficient between `b1` and `b2` facilities use of 16-bit SIMD instructions on the RISC-V Core with DSP.

The state variables are stored in the array `pState`. Each Biquad stage has 4 state variables `x[n-1]`, `x[n-2]`, `y[n-1]`, and `y[n-2]`. The state variables are arranged in the `pState` array as: The 4 state variables for stage 1 are first, then the 4 state variables for stage 2, and so on. The state array has a total length of `4*numStages` values. The state variables are updated after each block of data is processed; the coefficients are untouched.

#### Parameters

- [inout] `S`: points to an instance of the Q15 Biquad cascade structure.
- [in] `numStages`: number of 2nd order stages in the filter.
- [in] `pCoeffs`: points to the filter coefficients.
- [in] `pState`: points to the state buffer.

- [in] `postShift`: Shift to be applied to the accumulator result. Varies according to the coefficients format

```
void riscv_biquad_cascade_df1_init_q31 (riscv_biquad_casd_df1_inst_q31 *S, uint8_t
                                     numStages, const q31_t *pCoeffs, q31_t
                                     *pState, int8_t postShift)
```

Initialization function for the Q31 Biquad cascade filter.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array `pCoeffs` in the following order:

where `b1x` and `a1x` are the coefficients for the first stage, `b2x` and `a2x` are the coefficients for the second stage, and so on. The `pCoeffs` array contains a total of  $5 \times \text{numStages}$  values.

The `pState` points to state variables array. Each Biquad stage has 4 state variables `x[n-1]`, `x[n-2]`, `y[n-1]`, and `y[n-2]`. The state variables are arranged in the `pState` array as: The 4 state variables for stage 1 are first, then the 4 state variables for stage 2, and so on. The state array has a total length of  $4 \times \text{numStages}$  values. The state variables are updated after each block of data is processed; the coefficients are untouched.

#### Parameters

- [inout] `S`: points to an instance of the Q31 Biquad cascade structure.
- [in] `numStages`: number of 2nd order stages in the filter.
- [in] `pCoeffs`: points to the filter coefficients.
- [in] `pState`: points to the state buffer.
- [in] `postShift`: Shift to be applied after the accumulator. Varies according to the coefficients format

```
void riscv_biquad_cascade_df1_q15 (const riscv_biquad_casd_df1_inst_q15 *S, const
                                   q15_t *pSrc, q15_t *pDst, uint32_t blockSize)
```

Processing function for the Q15 Biquad cascade filter.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both coefficients and state variables are represented in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. The accumulator is then shifted by `postShift` bits to truncate the result to 1.15 format by discarding the low 16 bits. Finally, the result is saturated to 1.15 format.

**Remark** Refer to `riscv_biquad_cascade_df1_fast_q15()` for a faster but less precise implementation of this filter.

#### Parameters

- [in] `S`: points to an instance of the Q15 Biquad cascade structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the location where the output result is written
- [in] `blockSize`: number of samples to process

```
void riscv_biquad_cascade_df1_q31 (const riscv_biquad_casd_df1_inst_q31 *S, const
                                   q31_t *pSrc, q31_t *pDst, uint32_t blockSize)
```

Processing function for the Q31 Biquad cascade filter.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clip. In order to avoid overflows completely the input signal must be scaled down by 2 bits and lie in the range  $[-0.25 +0.25)$ . After all 5 multiply-accumulates are performed, the 2.62 accumulator is shifted by `postShift` bits and the result truncated to 1.31 format by discarding the low 32 bits.

**Remark** Refer to `riscv_biquad_cascade_df1_fast_q31()` for a faster but less precise implementation of this filter.

**Parameters**

- [in] `S`: points to an instance of the Q31 Biquad cascade structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

### Biquad Cascade IIR Filters Using a Direct Form II Transposed Structure

```
LOW_OPTIMIZATION_ENTER void riscv_biquad_cascade_df2T_f16(const riscv_biquad_cascade_df2T
```

```
LOW_OPTIMIZATION_ENTER void riscv_biquad_cascade_df2T_f32(const riscv_biquad_cascade_df2T
```

```
LOW_OPTIMIZATION_ENTER void riscv_biquad_cascade_df2T_f64(const riscv_biquad_cascade_df2T
```

```
void riscv_biquad_cascade_df2T_init_f16 (riscv_biquad_cascade_df2T_instance_f16 *S, uint8_t
                                         numStages, const float16_t *pCoeffs, float16_t
                                         *pState)
```

```
void riscv_biquad_cascade_df2T_init_f32 (riscv_biquad_cascade_df2T_instance_f32 *S, uint8_t
                                         numStages, const float32_t *pCoeffs, float32_t
                                         *pState)
```

```
void riscv_biquad_cascade_df2T_init_f64 (riscv_biquad_cascade_df2T_instance_f64 *S, uint8_t
                                         numStages, const float64_t *pCoeffs, float64_t
                                         *pState)
```

```
LOW_OPTIMIZATION_ENTER void riscv_biquad_cascade_stereo_df2T_f16(const riscv_biquad_cascad
```

```
LOW_OPTIMIZATION_ENTER void riscv_biquad_cascade_stereo_df2T_f32(const riscv_biquad_cascad
```

```
void riscv_biquad_cascade_stereo_df2T_init_f16 (riscv_biquad_cascade_stereo_df2T_instance_f16
                                                *S, uint8_t numStages, const float16_t
                                                *pCoeffs, float16_t *pState)
```

```
void riscv_biquad_cascade_stereo_df2T_init_f32 (riscv_biquad_cascade_stereo_df2T_instance_f32
                                                *S, uint8_t numStages, const float32_t
                                                *pCoeffs, float32_t *pState)
```

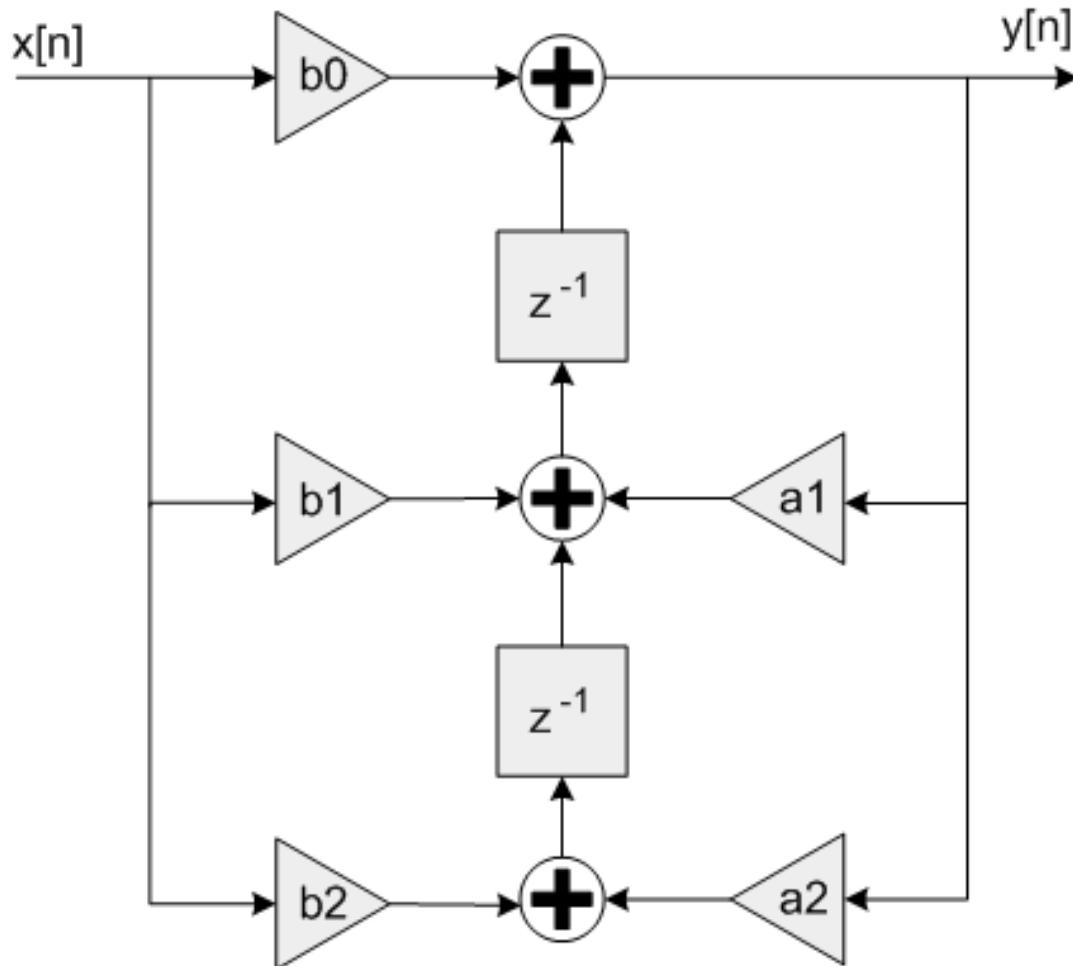
*group* **BiquadCascadeDF2T**

This set of functions implements arbitrary order recursive (IIR) filters using a transposed direct form II structure. The filters are implemented as a cascade of second order Biquad sections. These functions provide a slight memory savings as compared to the direct form I Biquad filter functions. Only floating-point data is supported.

This function operates on blocks of input and output data and each call to the function processes `blockSize` samples through the filter. `pSrc` points to the array of input data and `pDst` points to the array of output data. Both arrays contain `blockSize` values.

**Algorithm** Each Biquad stage implements a second order filter using the difference equation: where `d1` and `d2` represent the two state values.

A Biquad filter using a transposed Direct Form II structure is shown below. Coefficients `b0`, `b1`, and `b2` multiply the input signal  $x[n]$  and are referred to as the feedforward coefficients. Coefficients `a1` and `a2` multiply the output signal  $y[n]$  and are referred to as the feedback coefficients. Pay careful attention to the sign of the feedback coefficients. Some design tools flip the sign of the feedback coefficients: In this case the feedback coefficients `a1` and `a2` must be negated when used with the NMSIS DSP Library.



Higher order filters are realized as a cascade of second order sections. `numStages` refers to the number of second order stages used. For example, an 8th order filter would be realized with `numStages=4` second order stages. A 9th order filter would be realized with `numStages=5` second order stages with the coefficients for one of the stages configured as a first order filter (`b2=0` and `a2=0`).

`pState` points to the state variable array. Each Biquad stage has 2 state variables `d1` and `d2`. The state variables are arranged in the `pState` array as: where `d1x` refers to the state variables for the first Biquad and `d2x` refers to the state variables for the second Biquad. The state array has a total length of  $2 \times \text{numStages}$  values. The state variables are updated after each block of data is processed; the coefficients are untouched.

The NMSIS library contains Biquad filters in both Direct Form I and transposed Direct Form II. The advantage of the Direct Form I structure is that it is numerically more robust for fixed-point data types. That is why the Direct Form I structure supports Q15 and Q31 data types. The transposed Direct Form II structure, on the other hand, requires a wide dynamic range for the state variables `d1` and `d2`. Because of this, the NMSIS library only has a floating-point version of the Direct Form II Biquad. The advantage of the Direct Form II Biquad is that it requires half the number of state variables, 2 rather than 4, per Biquad stage.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient arrays may be shared among several instances while state variable arrays cannot be shared.

**Init Functions** There is also an associated initialization function. The initialization function performs following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numStages`, `pCoeffs`, `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. Set the values in the state buffer to zeros before static initialization. For example, to statically initialize the instance structure use where `numStages` is the number of Biquad stages in the filter; `pState` is the address of the state buffer. `pCoeffs` is the address of the coefficient buffer;

## Functions

**LOW\_OPTIMIZATION\_ENTER void riscv\_biquad\_cascade\_df2T\_f16(const riscv\_biquad\_cascade\_**  
Processing function for the floating-point transposed direct form II Biquad cascade filter.

**Return** none

### Parameters

- [in] `S`: points to an instance of the filter data structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

**LOW\_OPTIMIZATION\_ENTER void riscv\_biquad\_cascade\_df2T\_f32(const riscv\_biquad\_cascade\_**  
Processing function for the floating-point transposed direct form II Biquad cascade filter.

**Return** none

### Parameters

- [in] `S`: points to an instance of the filter data structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

**LOW\_OPTIMIZATION\_ENTER void riscv\_biquad\_cascade\_df2T\_f64(const riscv\_biquad\_cascade\_**  
Processing function for the floating-point transposed direct form II Biquad cascade filter.



**Return** none

**Parameters**

- [in] *S*: points to an instance of the filter data structure
- [in] *pSrc*: points to the block of input data
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process

```
void riscv_biquad_cascade_df2T_init_f16 (riscv_biquad_cascade_df2T_instance_f16 *S,
                                         uint8_t numStages, const float16_t *pCoeffs,
                                         float16_t *pState)
```

Initialization function for the floating-point transposed direct form II Biquad cascade filter.

For Neon version, this array is bigger. If  $\text{numStages} = 4x + y$ , then the array has size:  $32*x + 5*y$  and it must be initialized using the function `riscv_biquad_cascade_df2T_compute_coefs_f16` which is taking the standard array coefficient as parameters.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array *pCoeffs* in the following order in the not Neon version.

where *b1x* and *a1x* are the coefficients for the first stage, *b2x* and *a2x* are the coefficients for the second stage, and so on. The *pCoeffs* array contains a total of  $5 * \text{numStages}$  values.

**Parameters**

- [inout] *S*: points to an instance of the filter data structure.
- [in] *numStages*: number of 2nd order stages in the filter.
- [in] *pCoeffs*: points to the filter coefficients.
- [in] *pState*: points to the state buffer.

But, an array of  $8 * \text{numStages}$  is a good approximation.

Then, the initialization can be done with:

**In this example, neonCoefs is a bigger array of size  $8 * \text{numStages}$ .** *coefs* is the standard array:

The *pState* is a pointer to state array. Each Biquad stage has 2 state variables *d1*, and *d2*. The 2 state variables for stage 1 are first, then the 2 state variables for stage 2, and so on. The state array has a total length of  $2 * \text{numStages}$  values. The state variables are updated after each block of data is processed; the coefficients are untouched.

```
void riscv_biquad_cascade_df2T_init_f32 (riscv_biquad_cascade_df2T_instance_f32 *S,
                                         uint8_t numStages, const float32_t *pCoeffs,
                                         float32_t *pState)
```

Initialization function for the floating-point transposed direct form II Biquad cascade filter.

For Neon version, this array is bigger. If  $\text{numStages} = 4x + y$ , then the array has size:  $32*x + 5*y$  and it must be initialized using the function `riscv_biquad_cascade_df2T_compute_coefs_f32` which is taking the standard array coefficient as parameters.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array *pCoeffs* in the following order in the not Neon version.

where `b1x` and `a1x` are the coefficients for the first stage, `b2x` and `a2x` are the coefficients for the second stage, and so on. The `pCoeffs` array contains a total of `5*numStages` values.

#### Parameters

- [inout] `S`: points to an instance of the filter data structure.
- [in] `numStages`: number of 2nd order stages in the filter.
- [in] `pCoeffs`: points to the filter coefficients.
- [in] `pState`: points to the state buffer.

But, an array of `8*numstages` is a good approximation.

Then, the initialization can be done with:

**In this example, `neonCoefs` is a bigger array of size `8 * numStages`. `coefs` is the standard array:**

The `pState` is a pointer to state array. Each Biquad stage has 2 state variables `d1`, and `d2`. The 2 state variables for stage 1 are first, then the 2 state variables for stage 2, and so on. The state array has a total length of `2*numStages` values. The state variables are updated after each block of data is processed; the coefficients are untouched.

```
void riscv_biquad_cascade_df2T_init_f64 (riscv_biquad_cascade_df2T_instance_f64 *S,  
                                         uint8_t numStages, const float64_t *pCoeffs,  
                                         float64_t *pState)
```

Initialization function for the floating-point transposed direct form II Biquad cascade filter.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array `pCoeffs` in the following order:

where `b1x` and `a1x` are the coefficients for the first stage, `b2x` and `a2x` are the coefficients for the second stage, and so on. The `pCoeffs` array contains a total of `5*numStages` values.

The `pState` is a pointer to state array. Each Biquad stage has 2 state variables `d1`, and `d2`. The 2 state variables for stage 1 are first, then the 2 state variables for stage 2, and so on. The state array has a total length of `2*numStages` values. The state variables are updated after each block of data is processed; the coefficients are untouched.

#### Parameters

- [inout] `S`: points to an instance of the filter data structure
- [in] `numStages`: number of 2nd order stages in the filter
- [in] `pCoeffs`: points to the filter coefficients
- [in] `pState`: points to the state buffer

```
LOW_OPTIMIZATION_ENTER void riscv_biquad_cascade_stereo_df2T_f16(const riscv_biquad_c
```

Processing function for the floating-point transposed direct form II Biquad cascade filter.

Processing function for the floating-point transposed direct form II Biquad cascade filter. 2 channels.

**Return** none

#### Parameters

- [in] `S`: points to an instance of the filter data structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data

- [in] blockSize: number of samples to process

**LOW\_OPTIMIZATION\_ENTER** void **riscv\_biquad\_cascade\_stereo\_df2T\_f32**(const riscv\_biquad\_cascade\_stereo\_df2T\_instance\_f32 \*S, uint8\_t numStages, const float32\_t \*pCoeffs, float32\_t \*pState)

Processing function for the floating-point transposed direct form II Biquad cascade filter.

Processing function for the floating-point transposed direct form II Biquad cascade filter. 2 channels.

**Return** none

#### Parameters

- [in] S: points to an instance of the filter data structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

void **riscv\_biquad\_cascade\_stereo\_df2T\_init\_f16**(riscv\_biquad\_cascade\_stereo\_df2T\_instance\_f16 \*S, uint8\_t numStages, const float16\_t \*pCoeffs, float16\_t \*pState)

Initialization function for the floating-point transposed direct form II Biquad cascade filter.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array pCoeffs in the following order:

where b1x and a1x are the coefficients for the first stage, b2x and a2x are the coefficients for the second stage, and so on. The pCoeffs array contains a total of 5\*numStages values.

The pState is a pointer to state array. Each Biquad stage has 2 state variables d1, and d2 for each channel. The 2 state variables for stage 1 are first, then the 2 state variables for stage 2, and so on.

The state array has a total length of 2\*numStages values. The state variables are updated after each block of data is processed; the coefficients are untouched.

#### Parameters

- [inout] S: points to an instance of the filter data structure.
- [in] numStages: number of 2nd order stages in the filter.
- [in] pCoeffs: points to the filter coefficients.
- [in] pState: points to the state buffer.

void **riscv\_biquad\_cascade\_stereo\_df2T\_init\_f32**(riscv\_biquad\_cascade\_stereo\_df2T\_instance\_f32 \*S, uint8\_t numStages, const float32\_t \*pCoeffs, float32\_t \*pState)

Initialization function for the floating-point transposed direct form II Biquad cascade filter.

**Return** none

**Coefficient and State Ordering** The coefficients are stored in the array pCoeffs in the following order:

where b1x and a1x are the coefficients for the first stage, b2x and a2x are the coefficients for the second stage, and so on. The pCoeffs array contains a total of 5\*numStages values.

The pState is a pointer to state array. Each Biquad stage has 2 state variables d1, and d2 for each channel. The 2 state variables for stage 1 are first, then the 2 state variables for stage 2, and so on.

The state array has a total length of  $2 * \text{numStages}$  values. The state variables are updated after each block of data is processed; the coefficients are untouched.

#### Parameters

- [inout] *S*: points to an instance of the filter data structure.
- [in] *numStages*: number of 2nd order stages in the filter.
- [in] *pCoeffs*: points to the filter coefficients.
- [in] *pState*: points to the state buffer.

#### Convolution

```
void riscv_conv_f32 (const float32_t *pSrcA, uint32_t srcALen, const float32_t *pSrcB, uint32_t srcBLen, float32_t *pDst)
void riscv_conv_fast_opt_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB, uint32_t srcBLen, q15_t *pDst, q15_t *pScratch1, q15_t *pScratch2)
void riscv_conv_fast_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB, uint32_t srcBLen, q15_t *pDst)
void riscv_conv_fast_q31 (const q31_t *pSrcA, uint32_t srcALen, const q31_t *pSrcB, uint32_t srcBLen, q31_t *pDst)
void riscv_conv_opt_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB, uint32_t srcBLen, q15_t *pDst, q15_t *pScratch1, q15_t *pScratch2)
void riscv_conv_opt_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t *pSrcB, uint32_t srcBLen, q7_t *pDst, q15_t *pScratch1, q15_t *pScratch2)
void riscv_conv_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB, uint32_t srcBLen, q15_t *pDst)
void riscv_conv_q31 (const q31_t *pSrcA, uint32_t srcALen, const q31_t *pSrcB, uint32_t srcBLen, q31_t *pDst)
void riscv_conv_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t *pSrcB, uint32_t srcBLen, q7_t *pDst)
```

#### group Conv

Convolution is a mathematical operation that operates on two finite length vectors to generate a finite length output vector. Convolution is similar to correlation and is frequently used in filtering and data analysis. The NMSIS DSP library contains functions for convolving Q7, Q15, Q31, and floating-point data types. The library also provides fast versions of the Q15 and Q31 functions.

**Algorithm** Let  $a[n]$  and  $b[n]$  be sequences of length *srcALen* and *srcBLen* samples respectively. Then the convolution

$$c[n] = \sum_{k=0}^{\text{srcALen}} a[k]b[n-k]$$

is defined as

Note that  $c[n]$  is of length  $\text{srcALen} + \text{srcBLen} - 1$  and is defined over the interval  $n=0, 1, 2, \dots, \text{srcALen} + \text{srcBLen} - 2$ . *pSrcA* points to the first input vector of length *srcALen* and *pSrcB* points to the second input vector of length *srcBLen*. The output result is written to *pDst* and the calling function must allocate  $\text{srcALen} + \text{srcBLen} - 1$  words for the result.

Conceptually, when two signals  $a[n]$  and  $b[n]$  are convolved, the signal  $b[n]$  slides over  $a[n]$ . For each offset  $n$ , the overlapping portions of  $a[n]$  and  $b[n]$  are multiplied and summed together.

Note that convolution is a commutative operation:

This means that switching the A and B arguments to the convolution functions has no effect.

**Fixed-Point Behavior** Convolution requires summing up a large number of intermediate products. As such, the Q7, Q15, and Q31 functions run a risk of overflow and saturation. Refer to the function specific documentation below for further details of the particular algorithm used.

**Fast Versions** Fast versions are supported for Q31 and Q15. Cycles for Fast versions are less compared to Q31 and Q15 of conv and the design requires the input signals should be scaled down to avoid intermediate overflows.

**Opt Versions** Opt versions are supported for Q15 and Q7. Design uses internal scratch buffer for getting good optimisation. These versions are optimised in cycles and consumes more memory (Scratch memory) compared to Q15 and Q7 versions

## Functions

void **riscv\_conv\_f32** (const float32\_t \*pSrcA, uint32\_t srcALen, const float32\_t \*pSrcB, uint32\_t srcBLen, float32\_t \*pDst)  
Convolution of floating-point sequences.

**Return** none

**Parameters**

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written. Length srcALen+srcBLen-1.

void **riscv\_conv\_fast\_opt\_q15** (const q15\_t \*pSrcA, uint32\_t srcALen, const q15\_t \*pSrcB, uint32\_t srcBLen, q15\_t \*pDst, q15\_t \*pScratch1, q15\_t \*pScratch2)  
Convolution of Q15 sequences (fast version).

Convolution of Q15 sequences (fast version) for RISC-V Core with DSP enabled.

**Return** none

**Scaling and Overflow Behavior** This fast version uses a 32-bit accumulator with 2.30 format. The accumulator maintains full precision of the intermediate multiplication results but provides only a single guard bit. There is no saturation on intermediate additions. Thus, if the accumulator overflows it wraps around and distorts the result. The input signals should be scaled down to avoid intermediate overflows. Scale down the inputs by  $\log_2(\min(\text{srcALen}, \text{srcBLen}))$  ( $\log_2$  is read as log to the base 2) times to avoid overflows, as maximum of  $\min(\text{srcALen}, \text{srcBLen})$  number of additions are carried internally. The 2.30 accumulator is right shifted by 15 bits and then saturated to 1.15 format to yield the final result.

**Remark** Refer to riscv\_conv\_q15() for a slower implementation of this function which uses 64-bit accumulation to avoid wrap around distortion.

**Parameters**

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written. Length srcALen+srcBLen-1
- [in] pScratch1: points to scratch buffer of size  $\max(\text{srcALen}, \text{srcBLen}) + 2 * \min(\text{srcALen}, \text{srcBLen}) - 2$
- [in] pScratch2: points to scratch buffer of size  $\min(\text{srcALen}, \text{srcBLen})$

```
void riscv_conv_fast_q15(const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB,  
                        uint32_t srcBLen, q15_t *pDst)
```

Convolution of Q15 sequences (fast version).

Convolution of Q15 sequences (fast version) for RISC-V Core with DSP enabled.

**Return** none

**Scaling and Overflow Behavior** This fast version uses a 32-bit accumulator with 2.30 format. The accumulator maintains full precision of the intermediate multiplication results but provides only a single guard bit. There is no saturation on intermediate additions. Thus, if the accumulator overflows it wraps around and distorts the result. The input signals should be scaled down to avoid intermediate overflows. Scale down the inputs by  $\log_2(\min(\text{srcALen}, \text{srcBLen}))$  ( $\log_2$  is read as log to the base 2) times to avoid overflows, as maximum of  $\min(\text{srcALen}, \text{srcBLen})$  number of additions are carried internally. The 2.30 accumulator is right shifted by 15 bits and then saturated to 1.15 format to yield the final result.

**Remark** Refer to `riscv_conv_q15()` for a slower implementation of this function which uses 64-bit accumulation to avoid wrap around distortion.

**Parameters**

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written. Length srcALen+srcBLen-1

```
void riscv_conv_fast_q31(const q31_t *pSrcA, uint32_t srcALen, const q31_t *pSrcB,  
                        uint32_t srcBLen, q31_t *pDst)
```

Convolution of Q31 sequences (fast version).

Convolution of Q31 sequences (fast version) for RISC-V Core with DSP enabled.

**Return** none

**Scaling and Overflow Behavior** This function is optimized for speed at the expense of fixed-point precision and overflow protection. The result of each 1.31 x 1.31 multiplication is truncated to 2.30 format. These intermediate results are accumulated in a 32-bit register in 2.30 format. Finally, the accumulator is saturated and converted to a 1.31 result.

The fast version has the same overflow behavior as the standard version but provides less precision since it discards the low 32 bits of each multiplication result. In order to avoid overflows completely the input signals must be scaled down. Scale down the inputs by  $\log_2(\min(\text{srcALen}, \text{srcBLen}))$  ( $\log_2$  is read as log to the base 2) times to avoid overflows, as maximum of  $\min(\text{srcALen}, \text{srcBLen})$  number of additions are carried internally.

**Remark** Refer to `riscv_conv_q31()` for a slower implementation of this function which uses 64-bit accumulation to provide higher precision.

#### Parameters

- [in] `pSrcA`: points to the first input sequence.
- [in] `srcALen`: length of the first input sequence.
- [in] `pSrcB`: points to the second input sequence.
- [in] `srcBLen`: length of the second input sequence.
- [out] `pDst`: points to the location where the output result is written. Length `srcALen+srcBLen-1`.

void **riscv\_conv\_opt\_q15** (**const** q15\_t \**pSrcA*, uint32\_t *srcALen*, **const** q15\_t \**pSrcB*, uint32\_t *srcBLen*, q15\_t \**pDst*, q15\_t \**pScratch1*, q15\_t \**pScratch2*)  
Convolution of Q15 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both inputs are in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. This approach provides 33 guard bits and there is no risk of overflow. The 34.30 result is then truncated to 34.15 format by discarding the low 15 bits and then saturated to 1.15 format.

**Remark** Refer to `riscv_conv_fast_q15()` for a faster but less precise version of this function.

#### Parameters

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written. Length `srcALen+srcBLen-1`.
- [in] `pScratch1`: points to scratch buffer of size  $\max(\text{srcALen}, \text{srcBLen}) + 2 * \min(\text{srcALen}, \text{srcBLen}) - 2$ .
- [in] `pScratch2`: points to scratch buffer of size  $\min(\text{srcALen}, \text{srcBLen})$ .

void **riscv\_conv\_opt\_q7** (**const** q7\_t \**pSrcA*, uint32\_t *srcALen*, **const** q7\_t \**pSrcB*, uint32\_t *srcBLen*, q7\_t \**pDst*, q15\_t \**pScratch1*, q15\_t \**pScratch2*)  
Convolution of Q7 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. Both the inputs are represented in 1.7 format and multiplications yield a 2.14 result. The 2.14 intermediate results are accumulated in a 32-bit accumulator in 18.14 format. This approach provides 17 guard

bits and there is no risk of overflow as long as  $\max(\text{srcALen}, \text{srcBLen}) < 131072$ . The 18.14 result is then truncated to 18.7 format by discarding the low 7 bits and then saturated to 1.7 format.

**Parameters**

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written. Length  $\text{srcALen} + \text{srcBLen} - 1$ .
- [in] `pScratch1`: points to scratch buffer (of type `q15_t`) of size  $\max(\text{srcALen}, \text{srcBLen}) + 2 * \min(\text{srcALen}, \text{srcBLen}) - 2$ .
- [in] `pScratch2`: points to scratch buffer (of type `q15_t`) of size  $\min(\text{srcALen}, \text{srcBLen})$ .

void **riscv\_conv\_q15** (**const** `q15_t *pSrcA`, `uint32_t srcALen`, **const** `q15_t *pSrcB`, `uint32_t srcBLen`, `q15_t *pDst`)  
Convolution of Q15 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both inputs are in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. This approach provides 33 guard bits and there is no risk of overflow. The 34.30 result is then truncated to 34.15 format by discarding the low 15 bits and then saturated to 1.15 format.

**Remark** Refer to `riscv_conv_fast_q15()` for a faster but less precise version of this function.

**Remark** Refer to `riscv_conv_opt_q15()` for a faster implementation of this function using scratch buffers.

**Parameters**

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written. Length  $\text{srcALen} + \text{srcBLen} - 1$ .

void **riscv\_conv\_q31** (**const** `q31_t *pSrcA`, `uint32_t srcALen`, **const** `q31_t *pSrcB`, `uint32_t srcBLen`, `q31_t *pDst`)  
Convolution of Q31 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. There is no saturation on intermediate additions. Thus, if the accumulator overflows it wraps around and distorts the result. The input signals should be scaled down to avoid intermediate overflows. Scale down the inputs by  $\log_2(\min(\text{srcALen}, \text{srcBLen}))$  ( $\log_2$  is read as log to the base 2) times to avoid overflows, as maximum of  $\min(\text{srcALen}, \text{srcBLen})$  number



of additions are carried internally. The 2.62 accumulator is right shifted by 31 bits and saturated to 1.31 format to yield the final result.

**Remark** Refer to `riscv_conv_fast_q31()` for a faster but less precise implementation of this function.

#### Parameters

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written. Length `srcALen+srcBLen-1`.

void **riscv\_conv\_q7** (**const** q7\_t \**pSrcA*, uint32\_t *srcALen*, **const** q7\_t \**pSrcB*, uint32\_t *srcBLen*,  
q7\_t \**pDst*)  
Convolution of Q7 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. Both the inputs are represented in 1.7 format and multiplications yield a 2.14 result. The 2.14 intermediate results are accumulated in a 32-bit accumulator in 18.14 format. This approach provides 17 guard bits and there is no risk of overflow as long as `max(srcALen, srcBLen) < 131072`. The 18.14 result is then truncated to 18.7 format by discarding the low 7 bits and then saturated to 1.7 format.

**Remark** Refer to `riscv_conv_opt_q7()` for a faster implementation of this function.

#### Parameters

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written. Length `srcALen+srcBLen-1`.

### Partial Convolution

riscv\_status **riscv\_conv\_partial\_f32** (**const** float32\_t \**pSrcA*, uint32\_t *srcALen*, **const** float32\_t \**pSrcB*, uint32\_t *srcBLen*, float32\_t \**pDst*, uint32\_t *firstIndex*,  
uint32\_t *numPoints*)

riscv\_status **riscv\_conv\_partial\_fast\_opt\_q15** (**const** q15\_t \**pSrcA*, uint32\_t *srcALen*, **const** q15\_t \**pSrcB*, uint32\_t *srcBLen*, q15\_t \**pDst*,  
uint32\_t *firstIndex*, uint32\_t *numPoints*, q15\_t \**pScratch1*, q15\_t \**pScratch2*)

riscv\_status **riscv\_conv\_partial\_fast\_q15** (**const** q15\_t \**pSrcA*, uint32\_t *srcALen*, **const** q15\_t \**pSrcB*, uint32\_t *srcBLen*, q15\_t \**pDst*, uint32\_t *firstIndex*,  
uint32\_t *numPoints*)

riscv\_status **riscv\_conv\_partial\_fast\_q31** (**const** q31\_t \**pSrcA*, uint32\_t *srcALen*, **const** q31\_t \**pSrcB*, uint32\_t *srcBLen*, q31\_t \**pDst*, uint32\_t *firstIndex*,  
uint32\_t *numPoints*)

```

riscv_status riscv_conv_partial_opt_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t
                                         *pSrcB, uint32_t srcBLen, q15_t *pDst, uint32_t firstIndex,
                                         uint32_t numPoints, q15_t *pScratch1, q15_t
                                         *pScratch2)

riscv_status riscv_conv_partial_opt_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t
                                         *pSrcB, uint32_t srcBLen, q7_t *pDst, uint32_t firstIndex,
                                         uint32_t numPoints, q15_t *pScratch1, q15_t *pScratch2)

riscv_status riscv_conv_partial_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB,
                                       uint32_t srcBLen, q15_t *pDst, uint32_t firstIndex, uint32_t
                                       numPoints)

riscv_status riscv_conv_partial_q31 (const q31_t *pSrcA, uint32_t srcALen, const q31_t *pSrcB,
                                       uint32_t srcBLen, q31_t *pDst, uint32_t firstIndex, uint32_t
                                       numPoints)

riscv_status riscv_conv_partial_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t *pSrcB,
                                       uint32_t srcBLen, q7_t *pDst, uint32_t firstIndex, uint32_t num-
                                       Points)

```

#### group **PartialConv**

Partial Convolution is equivalent to Convolution except that a subset of the output samples is generated. Each function has two additional arguments. *firstIndex* specifies the starting index of the subset of output samples. *numPoints* is the number of output samples to compute. The function computes the output in the range [*firstIndex*, ..., *firstIndex*+*numPoints*-1]. The output array *pDst* contains *numPoints* values.

The allowable range of output indices is [0 *srcALen*+*srcBLen*-2]. If the requested subset does not fall in this range then the functions return **RISCV\_MATH\_ARGUMENT\_ERROR**. Otherwise the functions return **RISCV\_MATH\_SUCCESS**.

**Note** Refer to `riscv_conv_f32()` for details on fixed point behavior.

**Fast Versions** Fast versions are supported for Q31 and Q15 of partial convolution. Cycles for Fast versions are less compared to Q31 and Q15 of partial conv and the design requires the input signals should be scaled down to avoid intermediate overflows.

**Opt Versions** Opt versions are supported for Q15 and Q7. Design uses internal scratch buffer for getting good optimisation. These versions are optimised in cycles and consumes more memory (Scratch memory) compared to Q15 and Q7 versions of partial convolution

## Functions

```

riscv_status riscv_conv_partial_f32 (const float32_t *pSrcA, uint32_t srcALen, const
                                       float32_t *pSrcB, uint32_t srcBLen, float32_t *pDst,
                                       uint32_t firstIndex, uint32_t numPoints)

```

Partial convolution of floating-point sequences.

**Return** execution status

- **RISCV\_MATH\_SUCCESS** : Operation successful
- **RISCV\_MATH\_ARGUMENT\_ERROR** : requested subset is not in the range [0 *srcALen*+*srcBLen*-2]

**Parameters**

- [in] *pSrcA*: points to the first input sequence
- [in] *srcALen*: length of the first input sequence

- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written
- [in] firstIndex: is the first output sample to start with
- [in] numPoints: is the number of output points to be computed

```
riscv_status riscv_conv_partial_fast_opt_q15(const q15_t *pSrcA, uint32_t srcALen,
                                             const q15_t *pSrcB, uint32_t srcBLen,
                                             q15_t *pDst, uint32_t firstIndex, uint32_t
                                             numPoints, q15_t *pScratch1, q15_t
                                             *pScratch2)
```

Partial convolution of Q15 sequences (fast version).

Partial convolution of Q15 sequences (fast version) for RISC-V Core with DSP enabled.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : requested subset is not in the range [0 srcALen+srcBLen-2]

**Remark** Refer to riscv\_conv\_partial\_q15() for a slower implementation of this function which uses a 64-bit accumulator to avoid wrap around distortion.

**Parameters**

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written
- [in] firstIndex: is the first output sample to start with
- [in] numPoints: is the number of output points to be computed
- [in] pScratch1: points to scratch buffer of size max(srcALen, srcBLen) + 2\*min(srcALen, srcBLen) - 2
- [in] pScratch2: points to scratch buffer of size min(srcALen, srcBLen)

```
riscv_status riscv_conv_partial_fast_q15(const q15_t *pSrcA, uint32_t srcALen, const
                                         q15_t *pSrcB, uint32_t srcBLen, q15_t *pDst,
                                         uint32_t firstIndex, uint32_t numPoints)
```

Partial convolution of Q15 sequences (fast version).

Partial convolution of Q15 sequences (fast version) for RISC-V Core with DSP enabled.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : requested subset is not in the range [0 srcALen+srcBLen-2]

**Remark** Refer to `riscv_conv_partial_q15()` for a slower implementation of this function which uses a 64-bit accumulator to avoid wrap around distortion.

**Parameters**

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written
- [in] `firstIndex`: is the first output sample to start with
- [in] `numPoints`: is the number of output points to be computed

```
riscv_status riscv_conv_partial_fast_q31(const q31_t *pSrcA, uint32_t srcALen, const  
                                         q31_t *pSrcB, uint32_t srcBLen, q31_t *pDst,  
                                         uint32_t firstIndex, uint32_t numPoints)
```

Partial convolution of Q31 sequences (fast version).

Partial convolution of Q31 sequences (fast version) for RISC-V Core with DSP enabled.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_ARGUMENT_ERROR` : requested subset is not in the range [0 srcALen+srcBLen-2]

**Remark** Refer to `riscv_conv_partial_q31()` for a slower implementation of this function which uses a 64-bit accumulator to provide higher precision.

**Parameters**

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written
- [in] `firstIndex`: is the first output sample to start with
- [in] `numPoints`: is the number of output points to be computed

```
riscv_status riscv_conv_partial_opt_q15(const q15_t *pSrcA, uint32_t srcALen, const  
                                         q15_t *pSrcB, uint32_t srcBLen, q15_t *pDst,  
                                         uint32_t firstIndex, uint32_t numPoints, q15_t  
                                         *pScratch1, q15_t *pScratch2)
```

Partial convolution of Q15 sequences.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_ARGUMENT_ERROR` : requested subset is not in the range [0 srcALen+srcBLen-2]

**Remark** Refer to `riscv_conv_partial_fast_q15()` for a faster but less precise version of this function.

**Parameters**

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written
- [in] firstIndex: is the first output sample to start with
- [in] numPoints: is the number of output points to be computed
- [in] pScratch1: points to scratch buffer of size  $\max(\text{srcALen}, \text{srcBLen}) + 2 \cdot \min(\text{srcALen}, \text{srcBLen}) - 2$ .
- [in] pScratch2: points to scratch buffer of size  $\min(\text{srcALen}, \text{srcBLen})$ .

```
riscv_status riscv_conv_partial_opt_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t
                                         *pSrcB, uint32_t srcBLen, q7_t *pDst, uint32_t firstIndex,
                                         uint32_t numPoints, q15_t *pScratch1, q15_t
                                         *pScratch2)
```

Partial convolution of Q7 sequences.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : requested subset is not in the range  $[0, \text{srcALen} + \text{srcBLen} - 2]$

**Parameters**

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written
- [in] firstIndex: is the first output sample to start with
- [in] numPoints: is the number of output points to be computed
- [in] pScratch1: points to scratch buffer (of type q15\_t) of size  $\max(\text{srcALen}, \text{srcBLen}) + 2 \cdot \min(\text{srcALen}, \text{srcBLen}) - 2$ .
- [in] pScratch2: points to scratch buffer (of type q15\_t) of size  $\min(\text{srcALen}, \text{srcBLen})$ .

```
riscv_status riscv_conv_partial_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t
                                       *pSrcB, uint32_t srcBLen, q15_t *pDst, uint32_t firstIndex,
                                       uint32_t numPoints)
```

Partial convolution of Q15 sequences.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : requested subset is not in the range  $[0, \text{srcALen} + \text{srcBLen} - 2]$

**Remark** Refer to `riscv_conv_partial_fast_q15()` for a faster but less precise version of this function.

**Remark** Refer to `riscv_conv_partial_opt_q15()` for a faster implementation of this function using scratch buffers.

**Parameters**

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written
- [in] `firstIndex`: is the first output sample to start with
- [in] `numPoints`: is the number of output points to be computed

`riscv_status riscv_conv_partial_q31 (const q31_t *pSrcA, uint32_t srcALen, const q31_t *pSrcB, uint32_t srcBLen, q31_t *pDst, uint32_t firstIndex, uint32_t numPoints)`

Partial convolution of Q31 sequences.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_ARGUMENT_ERROR` : requested subset is not in the range `[0 srcALen+srcBLen-2]`

**Remark** Refer to `riscv_conv_partial_fast_q31()` for a faster but less precise implementation of this function.

**Parameters**

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written
- [in] `firstIndex`: is the first output sample to start with
- [in] `numPoints`: is the number of output points to be computed

`riscv_status riscv_conv_partial_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t *pSrcB, uint32_t srcBLen, q7_t *pDst, uint32_t firstIndex, uint32_t numPoints)`

Partial convolution of Q7 sequences.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_ARGUMENT_ERROR` : requested subset is not in the range `[0 srcALen+srcBLen-2]`

**Remark** Refer to `riscv_conv_partial_opt_q7()` for a faster implementation of this function.

### Parameters

- [in] `pSrcA`: points to the first input sequence
- [in] `srcALen`: length of the first input sequence
- [in] `pSrcB`: points to the second input sequence
- [in] `srcBLen`: length of the second input sequence
- [out] `pDst`: points to the location where the output result is written
- [in] `firstIndex`: is the first output sample to start with
- [in] `numPoints`: is the number of output points to be computed

### Correlation

```
void riscv_correlate_f16 (const float16_t *pSrcA, uint32_t srcALen, const float16_t *pSrcB,
                          uint32_t srcBLen, float16_t *pDst)
void riscv_correlate_f32 (const float32_t *pSrcA, uint32_t srcALen, const float32_t *pSrcB,
                          uint32_t srcBLen, float32_t *pDst)
void riscv_correlate_fast_opt_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB,
                                   uint32_t srcBLen, q15_t *pDst, q15_t *pScratch)
void riscv_correlate_fast_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB,
                               uint32_t srcBLen, q15_t *pDst)
void riscv_correlate_fast_q31 (const q31_t *pSrcA, uint32_t srcALen, const q31_t *pSrcB,
                               uint32_t srcBLen, q31_t *pDst)
void riscv_correlate_opt_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB,
                              uint32_t srcBLen, q15_t *pDst, q15_t *pScratch)
void riscv_correlate_opt_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t *pSrcB, uint32_t
                             srcBLen, q7_t *pDst, q15_t *pScratch1, q15_t *pScratch2)
void riscv_correlate_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB, uint32_t sr-
                          cBLen, q15_t *pDst)
void riscv_correlate_q31 (const q31_t *pSrcA, uint32_t srcALen, const q31_t *pSrcB, uint32_t sr-
                          cBLen, q31_t *pDst)
void riscv_correlate_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t *pSrcB, uint32_t sr-
                        cBLen, q7_t *pDst)
```

#### group Corr

Correlation is a mathematical operation that is similar to convolution. As with convolution, correlation uses two signals to produce a third signal. The underlying algorithms in correlation and convolution are identical except that one of the inputs is flipped in convolution. Correlation is commonly used to measure the similarity between two signals. It has applications in pattern recognition, cryptanalysis, and searching. The NMSIS library provides correlation functions for Q7, Q15, Q31 and floating-point data types. Fast versions of the Q15 and Q31 functions are also provided.

**Algorithm** Let  $a[n]$  and  $b[n]$  be sequences of length `srcALen` and `srcBLen` samples respectively. The convolution of the two signals is denoted by  $\text{In}$  correlation, one of the signals is flipped in time

$$c[n] = \sum_{k=0}^{srcALen} a[k] b[k - n]$$

and this is mathematically defined as

The `pSrcA` points to the first input vector of length `srcALen` and `pSrcB` points to the second input vector of length `srcBLen`. The result `c[n]` is of length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$  and is defined over the interval  $n=0, 1, 2, \dots, (2 * \max(\text{srcALen}, \text{srcBLen}) - 2)$ . The output result is written to `pDst` and the calling function must allocate  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$  words for the result.

**Note** The `pDst` should be initialized to all zeros before being used.

**Fixed-Point Behavior** Correlation requires summing up a large number of intermediate products. As such, the Q7, Q15, and Q31 functions run a risk of overflow and saturation. Refer to the function specific documentation below for further details of the particular algorithm used.

**Fast Versions** Fast versions are supported for Q31 and Q15. Cycles for Fast versions are less compared to Q31 and Q15 of correlate and the design requires the input signals should be scaled down to avoid intermediate overflows.

**Opt Versions** Opt versions are supported for Q15 and Q7. Design uses internal scratch buffer for getting good optimisation. These versions are optimised in cycles and consumes more memory (Scratch memory) compared to Q15 and Q7 versions of correlate

## Functions

void **riscv\_correlate\_f16** (**const** float16\_t \*pSrcA, uint32\_t srcALen, **const** float16\_t \*pSrcB, uint32\_t srcBLen, float16\_t \*pDst)  
Correlation of floating-point sequences.

**Return** none

### Parameters

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .

void **riscv\_correlate\_f32** (**const** float32\_t \*pSrcA, uint32\_t srcALen, **const** float32\_t \*pSrcB, uint32\_t srcBLen, float32\_t \*pDst)  
Correlation of floating-point sequences.

**Return** none

### Parameters

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence



- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .

```
void riscv_correlate_fast_opt_q15 (const q15_t *pSrcA, uint32_t srcALen, const
                                   q15_t *pSrcB, uint32_t srcBLen, q15_t *pDst, q15_t
                                   *pScratch)
```

Correlation of Q15 sequences (fast version).

**Return** none

**Scaling and Overflow Behavior** This fast version uses a 32-bit accumulator with 2.30 format. The accumulator maintains full precision of the intermediate multiplication results but provides only a single guard bit. There is no saturation on intermediate additions. Thus, if the accumulator overflows it wraps around and distorts the result. The input signals should be scaled down to avoid intermediate overflows. Scale down one of the inputs by  $1/\min(\text{srcALen}, \text{srcBLen})$  to avoid overflow since a maximum of  $\min(\text{srcALen}, \text{srcBLen})$  number of additions is carried internally. The 2.30 accumulator is right shifted by 15 bits and then saturated to 1.15 format to yield the final result.

**Remark** Refer to `riscv_correlate_q15()` for a slower implementation of this function which uses a 64-bit accumulator to avoid wrap around distortion.

#### Parameters

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence.
- [out] pDst: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .
- [in] pScratch: points to scratch buffer of size  $\max(\text{srcALen}, \text{srcBLen}) + 2 * \min(\text{srcALen}, \text{srcBLen}) - 2$ .

```
void riscv_correlate_fast_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB,
                               uint32_t srcBLen, q15_t *pDst)
```

Correlation of Q15 sequences (fast version).

**Return** none

**Scaling and Overflow Behavior** This fast version uses a 32-bit accumulator with 2.30 format. The accumulator maintains full precision of the intermediate multiplication results but provides only a single guard bit. There is no saturation on intermediate additions. Thus, if the accumulator overflows it wraps around and distorts the result. The input signals should be scaled down to avoid intermediate overflows. Scale down one of the inputs by  $1/\min(\text{srcALen}, \text{srcBLen})$  to avoid overflow since a maximum of  $\min(\text{srcALen}, \text{srcBLen})$  number of additions is carried internally. The 2.30 accumulator is right shifted by 15 bits and then saturated to 1.15 format to yield the final result.

**Remark** Refer to `riscv_correlate_q15()` for a slower implementation of this function which uses a 64-bit accumulator to avoid wrap around distortion.

#### Parameters

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence

- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .

void **riscv\_correlate\_fast\_q31** (const q31\_t \*pSrcA, uint32\_t srcALen, const q31\_t \*pSrcB, uint32\_t srcBLen, q31\_t \*pDst)

Correlation of Q31 sequences (fast version).

**Return** none

**Scaling and Overflow Behavior** This function is optimized for speed at the expense of fixed-point precision and overflow protection. The result of each 1.31 x 1.31 multiplication is truncated to 2.30 format. These intermediate results are accumulated in a 32-bit register in 2.30 format. Finally, the accumulator is saturated and converted to a 1.31 result.

The fast version has the same overflow behavior as the standard version but provides less precision since it discards the low 32 bits of each multiplication result. In order to avoid overflows completely the input signals must be scaled down. The input signals should be scaled down to avoid intermediate overflows. Scale down one of the inputs by  $1/\min(\text{srcALen}, \text{srcBLen})$  to avoid overflows since a maximum of  $\min(\text{srcALen}, \text{srcBLen})$  number of additions is carried internally.

**Remark** Refer to `riscv_correlate_q31()` for a slower implementation of this function which uses 64-bit accumulation to provide higher precision.

#### Parameters

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .

void **riscv\_correlate\_opt\_q15** (const q15\_t \*pSrcA, uint32\_t srcALen, const q15\_t \*pSrcB, uint32\_t srcBLen, q15\_t \*pDst, q15\_t \*pScratch)

Correlation of Q15 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both inputs are in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. This approach provides 33 guard bits and there is no risk of overflow. The 34.30 result is then truncated to 34.15 format by discarding the low 15 bits and then saturated to 1.15 format.

**Remark** Refer to `riscv_correlate_fast_q15()` for a faster but less precise version of this function.

#### Parameters

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence

- [out] pDst: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .
- [in] pScratch: points to scratch buffer of size  $\max(\text{srcALen}, \text{srcBLen}) + 2 * \min(\text{srcALen}, \text{srcBLen}) - 2$ .

```
void riscv_correlate_opt_q7 (const q7_t *pSrcA, uint32_t srcALen, const q7_t *pSrcB,
                             uint32_t srcBLen, q7_t *pDst, q15_t *pScratch1, q15_t
                             *pScratch2)
```

Correlation of Q7 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. Both the inputs are represented in 1.7 format and multiplications yield a 2.14 result. The 2.14 intermediate results are accumulated in a 32-bit accumulator in 18.14 format. This approach provides 17 guard bits and there is no risk of overflow as long as  $\max(\text{srcALen}, \text{srcBLen}) < 131072$ . The 18.14 result is then truncated to 18.7 format by discarding the low 7 bits and then saturated to 1.7 format.

#### Parameters

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence
- [out] pDst: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .
- [in] pScratch1: points to scratch buffer (of type q15\_t) of size  $\max(\text{srcALen}, \text{srcBLen}) + 2 * \min(\text{srcALen}, \text{srcBLen}) - 2$ .
- [in] pScratch2: points to scratch buffer (of type q15\_t) of size  $\min(\text{srcALen}, \text{srcBLen})$ .

```
void riscv_correlate_q15 (const q15_t *pSrcA, uint32_t srcALen, const q15_t *pSrcB,
                          uint32_t srcBLen, q15_t *pDst)
```

Correlation of Q15 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both inputs are in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. This approach provides 33 guard bits and there is no risk of overflow. The 34.30 result is then truncated to 34.15 format by discarding the low 15 bits and then saturated to 1.15 format.

**Remark** Refer to `riscv_correlate_fast_q15()` for a faster but less precise version of this function.

**Remark** Refer to `riscv_correlate_opt_q15()` for a faster implementation of this function using scratch buffers.

#### Parameters

- [in] pSrcA: points to the first input sequence
- [in] srcALen: length of the first input sequence
- [in] pSrcB: points to the second input sequence
- [in] srcBLen: length of the second input sequence

- [out] *pDst*: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .

void **riscv\_correlate\_q31**(const q31\_t \**pSrcA*, uint32\_t *srcALen*, const q31\_t \**pSrcB*,  
uint32\_t *srcBLen*, q31\_t \**pDst*)

Correlation of Q31 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. There is no saturation on intermediate additions. Thus, if the accumulator overflows it wraps around and distorts the result. The input signals should be scaled down to avoid intermediate overflows. Scale down one of the inputs by  $1/\min(\text{srcALen}, \text{srcBLen})$  to avoid overflows since a maximum of  $\min(\text{srcALen}, \text{srcBLen})$  number of additions is carried internally. The 2.62 accumulator is right shifted by 31 bits and saturated to 1.31 format to yield the final result.

**Remark** Refer to `riscv_correlate_fast_q31()` for a faster but less precise implementation of this function.

**Parameters**

- [in] *pSrcA*: points to the first input sequence
- [in] *srcALen*: length of the first input sequence
- [in] *pSrcB*: points to the second input sequence
- [in] *srcBLen*: length of the second input sequence
- [out] *pDst*: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .

void **riscv\_correlate\_q7**(const q7\_t \**pSrcA*, uint32\_t *srcALen*, const q7\_t \**pSrcB*, uint32\_t  
*srcBLen*, q7\_t \**pDst*)

Correlation of Q7 sequences.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. Both the inputs are represented in 1.7 format and multiplications yield a 2.14 result. The 2.14 intermediate results are accumulated in a 32-bit accumulator in 18.14 format. This approach provides 17 guard bits and there is no risk of overflow as long as  $\max(\text{srcALen}, \text{srcBLen}) < 131072$ . The 18.14 result is then truncated to 18.7 format by discarding the low 7 bits and saturated to 1.7 format.

**Remark** Refer to `riscv_correlate_opt_q7()` for a faster implementation of this function.

**Parameters**

- [in] *pSrcA*: points to the first input sequence
- [in] *srcALen*: length of the first input sequence
- [in] *pSrcB*: points to the second input sequence
- [in] *srcBLen*: length of the second input sequence
- [out] *pDst*: points to the location where the output result is written. Length  $2 * \max(\text{srcALen}, \text{srcBLen}) - 1$ .

## Finite Impulse Response (FIR) Decimator

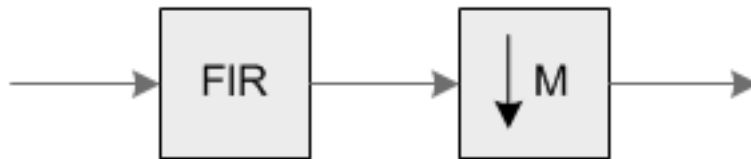
```

void riscv_fir_decimate_f32 (const riscv_fir_decimate_instance_f32 *S, const float32_t *pSrc,
                             float32_t *pDst, uint32_t blockSize)
void riscv_fir_decimate_fast_q15 (const riscv_fir_decimate_instance_q15 *S, const q15_t
                                  *pSrc, q15_t *pDst, uint32_t blockSize)
void riscv_fir_decimate_fast_q31 (const riscv_fir_decimate_instance_q31 *S, const q31_t
                                  *pSrc, q31_t *pDst, uint32_t blockSize)
riscv_status riscv_fir_decimate_init_f32 (riscv_fir_decimate_instance_f32 *S, uint16_t numTaps,
                                           uint8_t M, const float32_t *pCoeffs, float32_t *pState,
                                           uint32_t blockSize)
riscv_status riscv_fir_decimate_init_q15 (riscv_fir_decimate_instance_q15 *S, uint16_t numTaps,
                                           uint8_t M, const q15_t *pCoeffs, q15_t *pState,
                                           uint32_t blockSize)
riscv_status riscv_fir_decimate_init_q31 (riscv_fir_decimate_instance_q31 *S, uint16_t numTaps,
                                           uint8_t M, const q31_t *pCoeffs, q31_t *pState,
                                           uint32_t blockSize)
void riscv_fir_decimate_q15 (const riscv_fir_decimate_instance_q15 *S, const q15_t *pSrc,
                             q15_t *pDst, uint32_t blockSize)
void riscv_fir_decimate_q31 (const riscv_fir_decimate_instance_q31 *S, const q31_t *pSrc,
                             q31_t *pDst, uint32_t blockSize)

```

### group **FIR\_decimate**

These functions combine an FIR filter together with a decimator. They are used in multirate systems for reducing the sample rate of a signal without introducing aliasing distortion. Conceptually, the functions are equivalent to the block diagram below: When decimating by a factor of  $M$ , the signal should be prefiltered by a lowpass filter with a normalized cutoff frequency of  $1/M$  in order to prevent aliasing distortion. The user of the function is responsible for providing the filter



coefficients.

The FIR decimator functions provided in the NMSIS DSP Library combine the FIR filter and the decimator in an efficient manner. Instead of calculating all of the FIR filter outputs and discarding  $M-1$  out of every  $M$ , only the samples output by the decimator are computed. The functions operate on blocks of input and output data. `pSrc` points to an array of `blockSize` input values and `pDst` points to an array of `blockSize/M` output values. In order to have an integer number of output samples `blockSize` must always be a multiple of the decimation factor  $M$ .

The library provides separate functions for Q15, Q31 and floating-point data types.

**Algorithm:** The FIR portion of the algorithm uses the standard form filter: where,  $b[n]$  are the filter coefficients.

The `pCoeffs` points to a coefficient array of size `numTaps`. Coefficients are stored in time reversed order.

`pState` points to a state array of size `numTaps + blockSize - 1`. Samples in the state buffer are stored in the order:

The state variables are updated after each block of data is processed, the coefficients are untouched.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient arrays may be shared among several instances while state variable array should be allocated separately. There are separate instance structure declarations for each of the 3 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer.
- Checks to make sure that the size of the input is a multiple of the decimation factor. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numTaps`, `pCoeffs`, `M` (decimation factor), `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. The code below statically initializes each of the 3 different data type filter instance structures where `M` is the decimation factor; `numTaps` is the number of filter coefficients in the filter; `pCoeffs` is the address of the coefficient buffer; `pState` is the address of the state buffer. Be sure to set the values in the state buffer to zeros when doing static initialization.

**Fixed-Point Behavior** Care must be taken when using the fixed-point versions of the FIR decimate filter functions. In particular, the overflow and saturation behavior of the accumulator used in each function must be considered. Refer to the function specific documentation below for usage guidelines.

## Functions

```
void riscv_fir_decimate_f32 (const riscv_fir_decimate_instance_f32 *S, const float32_t
                             *pSrc, float32_t *pDst, uint32_t blockSize)
```

Processing function for floating-point FIR decimator.

**Return** none

**Parameters**

- [in] `S`: points to an instance of the floating-point FIR decimator structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

```
void riscv_fir_decimate_fast_q15 (const riscv_fir_decimate_instance_q15 *S, const
                                  q15_t *pSrc, q15_t *pDst, uint32_t blockSize)
```

Processing function for the Q15 FIR decimator (fast variant).

Processing function for the Q15 FIR decimator (fast variant) for RISC-V Core with DSP enabled.

**Return** none

**Scaling and Overflow Behavior** This fast version uses a 32-bit accumulator with 2.30 format. The accumulator maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around and distorts the result. In order to

avoid overflows completely the input signal must be scaled down by  $\log_2(\text{numTaps})$  bits ( $\log_2$  is read as log to the base 2). The 2.30 accumulator is then truncated to 2.15 format and saturated to yield the 1.15 result.

**Remark** Refer to `riscv_fir_decimate_q15()` for a slower implementation of this function which uses 64-bit accumulation to avoid wrap around distortion. Both the slow and the fast versions use the same instance structure. Use function `riscv_fir_decimate_init_q15()` to initialize the filter structure.

#### Parameters

- [in] `S`: points to an instance of the Q15 FIR decimator structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of input samples to process per call

```
void riscv_fir_decimate_fast_q31 (const riscv_fir_decimate_instance_q31 *S, const
                                q31_t *pSrc, q31_t *pDst, uint32_t blockSize)
```

Processing function for the Q31 FIR decimator (fast variant).

Processing function for the Q31 FIR decimator (fast variant) for RISC-V Core with DSP enabled.

**Return** none

**Scaling and Overflow Behavior** This function is optimized for speed at the expense of fixed-point precision and overflow protection. The result of each  $1.31 \times 1.31$  multiplication is truncated to 2.30 format. These intermediate results are added to a 2.30 accumulator. Finally, the accumulator is saturated and converted to a 1.31 result. The fast version has the same overflow behavior as the standard version and provides less precision since it discards the low 32 bits of each multiplication result. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{numTaps})$  bits (where  $\log_2$  is read as log to the base 2).

**Remark** Refer to `riscv_fir_decimate_q31()` for a slower implementation of this function which uses a 64-bit accumulator to provide higher precision. Both the slow and the fast versions use the same instance structure. Use function `riscv_fir_decimate_init_q31()` to initialize the filter structure.

#### Parameters

- [in] `S`: points to an instance of the Q31 FIR decimator structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

```
riscv_status riscv_fir_decimate_init_f32 (riscv_fir_decimate_instance_f32 *S, uint16_t num-
                                           Taps, uint8_t M, const float32_t *pCoeffs,
                                           float32_t *pState, uint32_t blockSize)
```

Initialization function for the floating-point FIR decimator.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_LENGTH_ERROR` : `blockSize` is not a multiple of `M`

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order:

pState points to the array of state variables. pState is of length numTaps+blockSize-1 words where blockSize is the number of input samples passed to riscv\_fir\_decimate\_f32(). M is the decimation factor.

**Parameters**

- [inout] S: points to an instance of the floating-point FIR decimator structure
- [in] numTaps: number of coefficients in the filter
- [in] M: decimation factor
- [in] pCoeffs: points to the filter coefficients
- [in] pState: points to the state buffer
- [in] blockSize: number of input samples to process per call

riscv\_status **riscv\_fir\_decimate\_init\_q15** (riscv\_fir\_decimate\_instance\_q15 \*S, uint16\_t numTaps, uint8\_t M, **const** q15\_t \*pCoeffs, q15\_t \*pState, uint32\_t blockSize)

Initialization function for the Q15 FIR decimator.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_LENGTH\_ERROR : blockSize is not a multiple of M

**Details** pCoeffs points to the array of filter coefficients stored in time reversed order:

pState points to the array of state variables. pState is of length numTaps+blockSize-1 words where blockSize is the number of input samples to the call riscv\_fir\_decimate\_q15(). M is the decimation factor.

**Parameters**

- [inout] S: points to an instance of the Q15 FIR decimator structure
- [in] numTaps: number of coefficients in the filter
- [in] M: decimation factor
- [in] pCoeffs: points to the filter coefficients
- [in] pState: points to the state buffer
- [in] blockSize: number of input samples to process

riscv\_status **riscv\_fir\_decimate\_init\_q31** (riscv\_fir\_decimate\_instance\_q31 \*S, uint16\_t numTaps, uint8\_t M, **const** q31\_t \*pCoeffs, q31\_t \*pState, uint32\_t blockSize)

Initialization function for the Q31 FIR decimator.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_LENGTH\_ERROR : blockSize is not a multiple of M

**Details** pCoeffs points to the array of filter coefficients stored in time reversed order:

pState points to the array of state variables. pState is of length numTaps+blockSize-1 words where blockSize is the number of input samples passed to riscv\_fir\_decimate\_q31(). M is the decimation factor.



**Parameters**

- [inout] *S*: points to an instance of the Q31 FIR decimator structure
- [in] *numTaps*: number of coefficients in the filter
- [in] *M*: decimation factor
- [in] *pCoeffs*: points to the filter coefficients
- [in] *pState*: points to the state buffer
- [in] *blockSize*: number of input samples to process

void **riscv\_fir\_decimate\_q15** (**const** riscv\_fir\_decimate\_instance\_q15 \**S*, **const** q15\_t \**pSrc*,  
q15\_t \**pDst*, uint32\_t *blockSize*)

Processing function for the Q15 FIR decimator.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both coefficients and state variables are represented in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. After all additions have been performed, the accumulator is truncated to 34.15 format by discarding low 15 bits. Lastly, the accumulator is saturated to yield a result in 1.15 format.

**Remark** Refer to `riscv_fir_decimate_fast_q15()` for a faster but less precise implementation of this function.

**Parameters**

- [in] *S*: points to an instance of the Q15 FIR decimator structure
- [in] *pSrc*: points to the block of input data
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of input samples to process per call

void **riscv\_fir\_decimate\_q31** (**const** riscv\_fir\_decimate\_instance\_q31 \**S*, **const** q31\_t \**pSrc*,  
q31\_t \**pDst*, uint32\_t *blockSize*)

Processing function for the Q31 FIR decimator.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clip. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{numTaps})$  bits (where  $\log_2$  is read as log to the base 2). After all multiply-accumulates are performed, the 2.62 accumulator is truncated to 1.32 format and then saturated to 1.31 format.

**Remark** Refer to `riscv_fir_decimate_fast_q31()` for a faster but less precise implementation of this function.

**Parameters**

- [in] *S*: points to an instance of the Q31 FIR decimator structure
- [in] *pSrc*: points to the block of input data
- [out] *pDst*: points to the block of output data

- [in] blockSize: number of samples to process

### Finite Impulse Response (FIR) Filters

```
void riscv_fir_f16 (const riscv_fir_instance_f16 *S, const float16_t *pSrc, float16_t *pDst, uint32_t  
    blockSize)
```

```
void riscv_fir_f32 (const riscv_fir_instance_f32 *S, const float32_t *pSrc, float32_t *pDst, uint32_t  
    blockSize)
```

```
void riscv_fir_fast_q15 (const riscv_fir_instance_q15 *S, const q15_t *pSrc, q15_t *pDst, uint32_t  
    blockSize)
```

```
IAR_ONLY_LOW_OPTIMIZATION_ENTER void riscv_fir_fast_q31 (const riscv_fir_instance_q31 *S,
```

```
void riscv_fir_init_f16 (riscv_fir_instance_f16 *S, uint16_t numTaps, const float16_t *pCoeffs,  
    float16_t *pState, uint32_t blockSize)
```

```
void riscv_fir_init_f32 (riscv_fir_instance_f32 *S, uint16_t numTaps, const float32_t *pCoeffs,  
    float32_t *pState, uint32_t blockSize)
```

```
riscv_status riscv_fir_init_q15 (riscv_fir_instance_q15 *S, uint16_t numTaps, const q15_t *pCoeffs,  
    q15_t *pState, uint32_t blockSize)
```

```
void riscv_fir_init_q31 (riscv_fir_instance_q31 *S, uint16_t numTaps, const q31_t *pCoeffs, q31_t  
    *pState, uint32_t blockSize)
```

```
void riscv_fir_init_q7 (riscv_fir_instance_q7 *S, uint16_t numTaps, const q7_t *pCoeffs, q7_t  
    *pState, uint32_t blockSize)
```

```
void riscv_fir_q15 (const riscv_fir_instance_q15 *S, const q15_t *pSrc, q15_t *pDst, uint32_t block-  
    Size)
```

```
void riscv_fir_q31 (const riscv_fir_instance_q31 *S, const q31_t *pSrc, q31_t *pDst, uint32_t block-  
    Size)
```

```
void riscv_fir_q7 (const riscv_fir_instance_q7 *S, const q7_t *pSrc, q7_t *pDst, uint32_t blockSize)
```

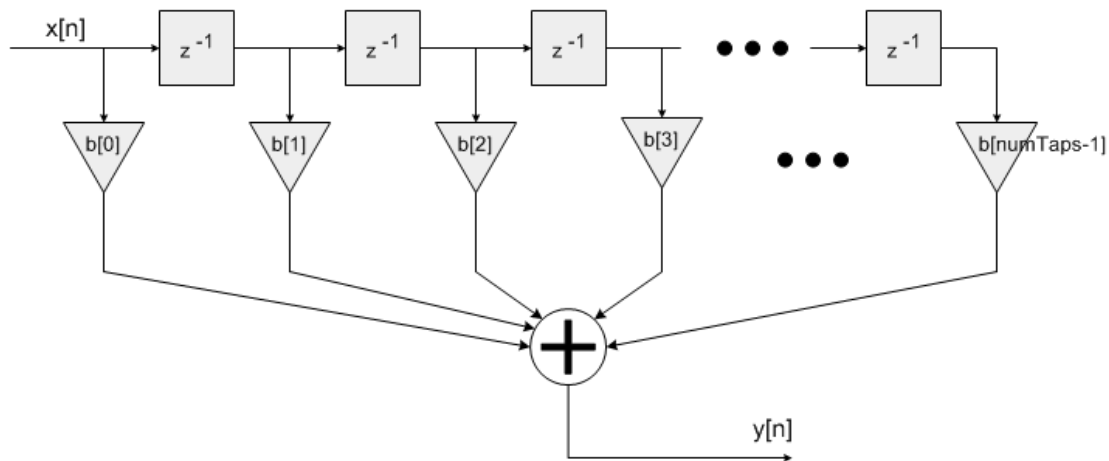
#### group **FIR**

This set of functions implements Finite Impulse Response (FIR) filters for Q7, Q15, Q31, and floating-point data types. Fast versions of Q15 and Q31 are also provided. The functions operate on blocks of input and output data and each call to the function processes `blockSize` samples through the filter. `pSrc` and `pDst` points to input and output arrays containing `blockSize` values.

The array length `L` must be a multiple of `x`.  $L = x * a$  :

- `x` is 4 for `f32`
- `x` is 4 for `q31`
- `x` is 4 for `f16` (so managed like the `f32` version and not like the `q15` one)
- `x` is 8 for `q15`
- `x` is 16 for `q7`

**Algorithm** The FIR filter algorithm is based upon a sequence of multiply-accumulate (MAC) operations. Each filter coefficient `b[n]` is multiplied by a state variable which equals a previous input sample `x[n]`.



`pCoeffs` points to a coefficient array of size `numTaps`. Coefficients are stored in time reversed order.

`pState` points to a state array of size `numTaps + blockSize - 1`. Samples in the state buffer are stored in the following order.

Note that the length of the state buffer exceeds the length of the coefficient array by `blockSize-1`. The increased state buffer length allows circular addressing, which is traditionally used in the FIR filters, to be avoided and yields a significant speed improvement. The state variables are updated after each block of data is processed; the coefficients are untouched.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient arrays may be shared among several instances while state variable arrays cannot be shared. There are separate instance structure declarations for each of the 4 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numTaps`, `pCoeffs`, `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. Set the values in the state buffer to zeros before static initialization. The code below statically initializes each of the 4 different data type filter instance structures where `numTaps` is the number of filter coefficients in the filter; `pState` is the address of the state buffer; `pCoeffs` is the address of the coefficient buffer.

**Initialization of Helium version** For Helium version the array of coefficients must be padded with zero to contain a full number of lanes.

The additional coefficients (`x * a - numTaps`) must be set to 0. `numTaps` is still set to its right value in the init function. It means that the implementation may require to read more coefficients due to the vectorization and to avoid having to manage too many different cases in the code.

**Helium state buffer** The state buffer must contain some additional temporary data used during the computation but which is not the state of the FIR. The first  $A$  samples are temporary data. The remaining samples are the state of the FIR filter.

So the state buffer has size  $\text{numTaps} + A + \text{blockSize} - 1$  :

- $A$  is  $\text{blockSize}$  for f32
- $A$  is  $8 * \text{ceil}(\text{blockSize}/8)$  for f16
- $A$  is  $8 * \text{ceil}(\text{blockSize}/4)$  for q31
- $A$  is 0 for other datatypes (q15 and q7)

**Fixed-Point Behavior** Care must be taken when using the fixed-point versions of the FIR filter functions. In particular, the overflow and saturation behavior of the accumulator used in each function must be considered. Refer to the function specific documentation below for usage guidelines.

## Functions

void **riscv\_fir\_f16**(**const** riscv\_fir\_instance\_f16 \*S, **const** float16\_t \*pSrc, float16\_t \*pDst,  
uint32\_t blockSize)

Processing function for floating-point FIR filter.

Processing function for the floating-point FIR filter.

**Return** none

### Parameters

- [in] S: points to an instance of the floating-point FIR filter structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

void **riscv\_fir\_f32**(**const** riscv\_fir\_instance\_f32 \*S, **const** float32\_t \*pSrc, float32\_t \*pDst,  
uint32\_t blockSize)

Processing function for floating-point FIR filter.

Processing function for the floating-point FIR filter.

**Return** none

### Parameters

- [in] S: points to an instance of the floating-point FIR filter structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

void **riscv\_fir\_fast\_q15**(**const** riscv\_fir\_instance\_q15 \*S, **const** q15\_t \*pSrc, q15\_t \*pDst,  
uint32\_t blockSize)

Processing function for the Q15 FIR filter (fast version).

Processing function for the fast Q15 FIR filter (fast version).

**Return** none

**Scaling and Overflow Behavior** This fast version uses a 32-bit accumulator with 2.30 format. The accumulator maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around and distorts the result. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{numTaps})$  bits. The 2.30 accumulator is then truncated to 2.15 format and saturated to yield the 1.15 result.

**Remark** Refer to `riscv_fir_q15()` for a slower implementation of this function which uses 64-bit accumulation to avoid wrap around distortion. Both the slow and the fast versions use the same instance structure. Use function `riscv_fir_init_q15()` to initialize the filter structure.

#### Parameters

- [in] `S`: points to an instance of the Q15 FIR filter structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

**IAR\_ONLY\_LOW\_OPTIMIZATION\_ENTER** `void riscv_fir_fast_q31(const riscv_fir_instance_q31`  
Processing function for the Q31 FIR filter (fast version).

Processing function for the fast Q31 FIR filter (fast version).

**Return** none

**Scaling and Overflow Behavior** This function is optimized for speed at the expense of fixed-point precision and overflow protection. The result of each  $1.31 \times 1.31$  multiplication is truncated to 2.30 format. These intermediate results are added to a 2.30 accumulator. Finally, the accumulator is saturated and converted to a 1.31 result. The fast version has the same overflow behavior as the standard version and provides less precision since it discards the low 32 bits of each multiplication result. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{numTaps})$  bits.

**Remark** Refer to `riscv_fir_q31()` for a slower implementation of this function which uses a 64-bit accumulator to provide higher precision. Both the slow and the fast versions use the same instance structure. Use function `riscv_fir_init_q31()` to initialize the filter structure.

#### Parameters

- [in] `S`: points to an instance of the Q31 structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

`void riscv_fir_init_f16(riscv_fir_instance_f16 *S, uint16_t numTaps, const float16_t *pCoeffs,`  
`float16_t *pState, uint32_t blockSize)`  
Initialization function for the floating-point FIR filter.

**Return** none

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order:

`pState` points to the array of state variables. `pState` is of length `numTaps+blockSize-1` samples (except for Helium - see below), where `blockSize` is the number of input samples processed by each call to `riscv_fir_f16()`.

**Initialization of Helium version** For Helium version the array of coefficients must be a multiple of 4 (4a) even if less than 4a coefficients are defined in the FIR. The additional coefficients (4a - numTaps) must be set to 0. numTaps is still set to its right value in the init function. It means that the implementation

may require to read more coefficients due to the vectorization and to avoid having to manage too many different cases in the code.

**Helium state buffer** The state buffer must contain some additional temporary data used during the computation but which is not the state of the FIR. The first  $8 \times \text{ceil}(\text{blockSize}/8)$  samples are temporary data. The remaining samples are the state of the FIR filter. So the state buffer has size  $\text{numTaps} + 8 \times \text{ceil}(\text{blockSize}/8) + \text{blockSize} - 1$

#### Parameters

- [inout] *S*: points to an instance of the floating-point FIR filter structure
- [in] *numTaps*: number of filter coefficients in the filter
- [in] *pCoeffs*: points to the filter coefficients buffer
- [in] *pState*: points to the state buffer
- [in] *blockSize*: number of samples processed per call

```
void riscv_fir_init_f32 (riscv_fir_instance_f32 *S, uint16_t numTaps, const float32_t *pCoeffs,  
                        float32_t *pState, uint32_t blockSize)
```

Initialization function for the floating-point FIR filter.

**Return** none

**Details** *pCoeffs* points to the array of filter coefficients stored in time reversed order:

*pState* points to the array of state variables and some working memory for the Helium version. *pState* is of length  $\text{numTaps} + \text{blockSize} - 1$  samples (except for Helium - see below), where *blockSize* is the number of input samples processed by each call to `riscv_fir_f32()`.

**Initialization of Helium version** For Helium version the array of coefficients must be a multiple of 4 (4a) even if less than 4a coefficients are defined in the FIR. The additional coefficients ( $4a - \text{numTaps}$ ) must be set to 0. *numTaps* is still set to its right value in the init function. It means that the implementation may require to read more coefficients due to the vectorization and to avoid having to manage too many different cases in the code.

**Helium state buffer** The state buffer must contain some additional temporary data used during the computation but which is not the state of the FIR. The first *blockSize* samples are temporary data. The remaining samples are the state of the FIR filter. So the state buffer has size  $\text{numTaps} + 2 * \text{blockSize} - 1$

#### Parameters

- [inout] *S*: points to an instance of the floating-point FIR filter structure
- [in] *numTaps*: number of filter coefficients in the filter
- [in] *pCoeffs*: points to the filter coefficients buffer
- [in] *pState*: points to the state buffer
- [in] *blockSize*: number of samples processed per call

```
riscv_status riscv_fir_init_q15 (riscv_fir_instance_q15 *S, uint16_t numTaps, const q15_t  
                                *pCoeffs, q15_t *pState, uint32_t blockSize)
```

Initialization function for the Q15 FIR filter.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_ARGUMENT_ERROR` : *numTaps* is not greater than or equal to 4 and even

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: Note that `numTaps` must be even and greater than or equal to 4. To implement an odd length filter simply increase `numTaps` by 1 and set the last coefficient to zero. For example, to implement a filter with `numTaps=3` and coefficients set `numTaps=4` and use the coefficients: Similarly, to implement a two point filter set `numTaps=4` and use the coefficients: `pState` points to the array of state variables. `pState` is of length `numTaps+blockSize`, when running on RISC-V Core with DSP enabled and is of length `numTaps+blockSize-1`, when running on RISC-V Core without DSP where `blockSize` is the number of input samples processed by each call to `riscv_fir_q15()`.

**Initialization of Helium version** For Helium version the array of coefficients must be a multiple of 8 (8a) even if less than 8a coefficients are defined in the FIR. The additional coefficients (8a - `numTaps`) must be set to 0. `numTaps` is still set to its right value in the init function. It means that the implementation may require to read more coefficients due to the vectorization and to avoid having to manage too many different cases in the code.

#### Parameters

- [inout] `S`: points to an instance of the Q15 FIR filter structure.
- [in] `numTaps`: number of filter coefficients in the filter. Must be even and greater than or equal to 4.
- [in] `pCoeffs`: points to the filter coefficients buffer.
- [in] `pState`: points to the state buffer.
- [in] `blockSize`: number of samples processed per call.

```
void riscv_fir_init_q31(riscv_fir_instance_q31 *S, uint16_t numTaps, const q31_t *pCoeffs,
                      q31_t *pState, uint32_t blockSize)
```

Initialization function for the Q31 FIR filter.

**Return** none

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: `pState` points to the array of state variables. `pState` is of length `numTaps+blockSize-1` samples (except for Helium - see below), where `blockSize` is the number of input samples processed by each call to `riscv_fir_q31()`.

**Initialization of Helium version** For Helium version the array of coefficients must be a multiple of 4 (4a) even if less than 4a coefficients are defined in the FIR. The additional coefficients (4a - `numTaps`) must be set to 0. `numTaps` is still set to its right value in the init function. It means that the implementation may require to read more coefficients due to the vectorization and to avoid having to manage too many different cases in the code.

**Helium state buffer** The state buffer must contain some additional temporary data used during the computation but which is not the state of the FIR. The first  $2*4*\text{ceil}(\text{blockSize}/4)$  samples are temporary data. The remaining samples are the state of the FIR filter. So the state buffer has size `numTaps + 8*ceil(blockSize/4) + blockSize - 1`

#### Parameters

- [inout] `S`: points to an instance of the Q31 FIR filter structure
- [in] `numTaps`: number of filter coefficients in the filter
- [in] `pCoeffs`: points to the filter coefficients buffer
- [in] `pState`: points to the state buffer
- [in] `blockSize`: number of samples processed

```
void riscv_fir_init_q7 (riscv_fir_instance_q7 *S, uint16_t numTaps, const q7_t *pCoeffs, q7_t  
                      *pState, uint32_t blockSize)
```

Initialization function for the Q7 FIR filter.

**Return** none

**Details** pCoeffs points to the array of filter coefficients stored in time reversed order:

pState points to the array of state variables. pState is of length numTaps+blockSize-1 samples, where blockSize is the number of input samples processed by each call to riscv\_fir\_q7().

**Initialization of Helium version** For Helium version the array of coefficients must be a multiple of 16 (16a) even if less than 16a coefficients are defined in the FIR. The additional coefficients (16a - numTaps) must be set to 0. numTaps is still set to its right value in the init function. It means that the implementation may require to read more coefficients due to the vectorization and to avoid having to manage too many different cases in the code.

**Parameters**

- [inout] S: points to an instance of the Q7 FIR filter structure
- [in] numTaps: number of filter coefficients in the filter
- [in] pCoeffs: points to the filter coefficients buffer
- [in] pState: points to the state buffer
- [in] blockSize: number of samples processed

```
void riscv_fir_q15 (const riscv_fir_instance_q15 *S, const q15_t *pSrc, q15_t *pDst, uint32_t  
                  blockSize)
```

Processing function for the Q15 FIR filter.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both coefficients and state variables are represented in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. After all additions have been performed, the accumulator is truncated to 34.15 format by discarding low 15 bits. Lastly, the accumulator is saturated to yield a result in 1.15 format.

**Remark** Refer to riscv\_fir\_fast\_q15() for a faster but less precise implementation of this function.

**Parameters**

- [in] S: points to an instance of the Q15 FIR filter structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

```
void riscv_fir_q31 (const riscv_fir_instance_q31 *S, const q31_t *pSrc, q31_t *pDst, uint32_t  
                  blockSize)
```

Processing function for Q31 FIR filter.

Processing function for the Q31 FIR filter.

**Return** none



**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clip. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{numTaps})$  bits. After all multiply-accumulates are performed, the 2.62 accumulator is right shifted by 31 bits and saturated to 1.31 format to yield the final result.

**Remark** Refer to `riscv_fir_fast_q31()` for a faster but less precise implementation of this filter.

#### Parameters

- [in] `S`: points to an instance of the Q31 FIR filter structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

```
void riscv_fir_q7 (const riscv_fir_instance_q7 *S, const q7_t *pSrc, q7_t *pDst, uint32_t blockSize)
```

Processing function for Q7 FIR filter.

Processing function for the Q7 FIR filter.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. Both coefficients and state variables are represented in 1.7 format and multiplications yield a 2.14 result. The 2.14 intermediate results are accumulated in a 32-bit accumulator in 18.14 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. The accumulator is converted to 18.7 format by discarding the low 7 bits. Finally, the result is truncated to 1.7 format.

#### Parameters

- [in] `S`: points to an instance of the Q7 FIR filter structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

### Finite Impulse Response (FIR) Lattice Filters

```
void riscv_fir_lattice_f32 (const riscv_fir_lattice_instance_f32 *S, const float32_t *pSrc, float32_t *pDst, uint32_t blockSize)
```

```
void riscv_fir_lattice_init_f32 (riscv_fir_lattice_instance_f32 *S, uint16_t numStages, const float32_t *pCoeffs, float32_t *pState)
```

```
void riscv_fir_lattice_init_q15 (riscv_fir_lattice_instance_q15 *S, uint16_t numStages, const q15_t *pCoeffs, q15_t *pState)
```

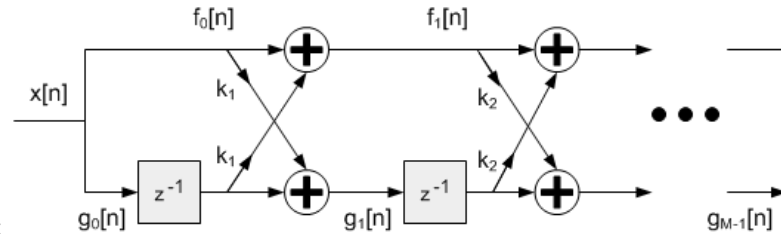
```
void riscv_fir_lattice_init_q31 (riscv_fir_lattice_instance_q31 *S, uint16_t numStages, const q31_t *pCoeffs, q31_t *pState)
```

```
void riscv_fir_lattice_q15 (const riscv_fir_lattice_instance_q15 *S, const q15_t *pSrc, q15_t *pDst, uint32_t blockSize)
```

```
void riscv_fir_lattice_q31 (const riscv_fir_lattice_instance_q31 *S, const q31_t *pSrc, q31_t *pDst, uint32_t blockSize)
```

**group FIR\_Lattice**

This set of functions implements Finite Impulse Response (FIR) lattice filters for Q15, Q31 and floating-point data types. Lattice filters are used in a variety of adaptive filter applications. The filter structure is feedforward and the net impulse response is finite length. The functions operate on blocks of input and output data and each call to the function processes `blockSize` samples through the filter. `pSrc` and `pDst` point to input and output arrays containing `blockSize` values.

**Algorithm**

The following difference equation is implemented:

`pCoeffs` points to the array of reflection coefficients of size `numStages`. Reflection Coefficients are stored in the following order.

where `M` is number of stages

`pState` points to a state array of size `numStages`. The state variables (`g` values) hold previous inputs and are stored in the following order. The state variables are updated after each block of data is processed; the coefficients are untouched.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient arrays may be shared among several instances while state variable arrays cannot be shared. There are separate instance structure declarations for each of the 3 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numStages`, `pCoeffs`, `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. Set the values in the state buffer to zeros and then manually initialize the instance structure as follows:

where `numStages` is the number of stages in the filter; `pState` is the address of the state buffer; `pCoeffs` is the address of the coefficient buffer.

**Fixed-Point Behavior** Care must be taken when using the fixed-point versions of the FIR Lattice filter functions. In particular, the overflow and saturation behavior of the accumulator used in each function must be considered. Refer to the function specific documentation below for usage guidelines.

## Functions

void **riscv\_fir\_lattice\_f32** (**const** riscv\_fir\_lattice\_instance\_f32 \*S, **const** float32\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)

Processing function for the floating-point FIR lattice filter.

**Return** none

### Parameters

- [in] S: points to an instance of the floating-point FIR lattice structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

void **riscv\_fir\_lattice\_init\_f32** (riscv\_fir\_lattice\_instance\_f32 \*S, uint16\_t numStages, **const** float32\_t \*pCoeffs, float32\_t \*pState)

Initialization function for the floating-point FIR lattice filter.

**Return** none

### Parameters

- [in] S: points to an instance of the floating-point FIR lattice structure
- [in] numStages: number of filter stages
- [in] pCoeffs: points to the coefficient buffer. The array is of length numStages
- [in] pState: points to the state buffer. The array is of length numStages

void **riscv\_fir\_lattice\_init\_q15** (riscv\_fir\_lattice\_instance\_q15 \*S, uint16\_t numStages, **const** q15\_t \*pCoeffs, q15\_t \*pState)

Initialization function for the Q15 FIR lattice filter.

**Return** none

### Parameters

- [in] S: points to an instance of the Q15 FIR lattice structure
- [in] numStages: number of filter stages
- [in] pCoeffs: points to the coefficient buffer. The array is of length numStages
- [in] pState: points to the state buffer. The array is of length numStages

void **riscv\_fir\_lattice\_init\_q31** (riscv\_fir\_lattice\_instance\_q31 \*S, uint16\_t numStages, **const** q31\_t \*pCoeffs, q31\_t \*pState)

Initialization function for the Q31 FIR lattice filter.

**Return** none

### Parameters

- [in] S: points to an instance of the Q31 FIR lattice structure
- [in] numStages: number of filter stages
- [in] pCoeffs: points to the coefficient buffer. The array is of length numStages
- [in] pState: points to the state buffer. The array is of length numStages

```
void riscv_fir_lattice_q15 (const riscv_fir_lattice_instance_q15 *S, const q15_t *pSrc,  
                        q15_t *pDst, uint32_t blockSize)
```

Processing function for Q15 FIR lattice filter.

Processing function for the Q15 FIR lattice filter.

**Return** none

**Parameters**

- [in] S: points to an instance of the Q15 FIR lattice structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

```
void riscv_fir_lattice_q31 (const riscv_fir_lattice_instance_q31 *S, const q31_t *pSrc,  
                        q31_t *pDst, uint32_t blockSize)
```

Processing function for the Q31 FIR lattice filter.

**Return** none

**Scaling and Overflow Behavior** In order to avoid overflows the input signal must be scaled down by  $2 \cdot \log_2(\text{numStages})$  bits.

**Parameters**

- [in] S: points to an instance of the Q31 FIR lattice structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

## Finite Impulse Response (FIR) Sparse Filters

```
void riscv_fir_sparse_f32 (riscv_fir_sparse_instance_f32 *S, const float32_t *pSrc, float32_t *pDst,  
                        float32_t *pScratchIn, uint32_t blockSize)
```

```
void riscv_fir_sparse_init_f32 (riscv_fir_sparse_instance_f32 *S, uint16_t numTaps, const  
                        float32_t *pCoeffs, float32_t *pState, int32_t *pTapDelay, uint16_t  
                        maxDelay, uint32_t blockSize)
```

```
void riscv_fir_sparse_init_q15 (riscv_fir_sparse_instance_q15 *S, uint16_t numTaps, const q15_t  
                        *pCoeffs, q15_t *pState, int32_t *pTapDelay, uint16_t maxDelay,  
                        uint32_t blockSize)
```

```
void riscv_fir_sparse_init_q31 (riscv_fir_sparse_instance_q31 *S, uint16_t numTaps, const q31_t  
                        *pCoeffs, q31_t *pState, int32_t *pTapDelay, uint16_t maxDelay,  
                        uint32_t blockSize)
```

```
void riscv_fir_sparse_init_q7 (riscv_fir_sparse_instance_q7 *S, uint16_t numTaps, const q7_t  
                        *pCoeffs, q7_t *pState, int32_t *pTapDelay, uint16_t maxDelay,  
                        uint32_t blockSize)
```

```
void riscv_fir_sparse_q15 (riscv_fir_sparse_instance_q15 *S, const q15_t *pSrc, q15_t *pDst, q15_t  
                        *pScratchIn, q31_t *pScratchOut, uint32_t blockSize)
```

```
void riscv_fir_sparse_q31 (riscv_fir_sparse_instance_q31 *S, const q31_t *pSrc, q31_t *pDst, q31_t  
                        *pScratchIn, uint32_t blockSize)
```

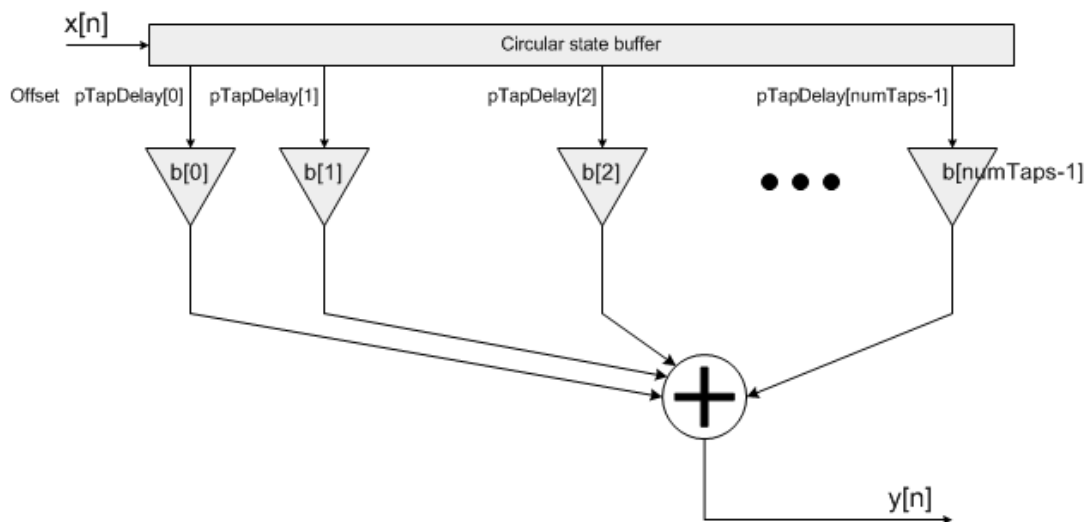
```
void riscv_fir_sparse_q7 (riscv_fir_sparse_instance_q7 *S, const q7_t *pSrc, q7_t *pDst, q7_t
                        *pScratchIn, q31_t *pScratchOut, uint32_t blockSize)
```

#### group **FIR\_Sparse**

This group of functions implements sparse FIR filters. Sparse FIR filters are equivalent to standard FIR filters except that most of the coefficients are equal to zero. Sparse filters are used for simulating reflections in communications and audio applications.

There are separate functions for Q7, Q15, Q31, and floating-point data types. The functions operate on blocks of input and output data and each call to the function processes `blockSize` samples through the filter. `pSrc` and `pDst` points to input and output arrays respectively containing `blockSize` values.

**Algorithm** The sparse filter instant structure contains an array of tap indices `pTapDelay` which specifies the locations of the non-zero coefficients. This is in addition to the coefficient array `b`. The implementation essentially skips the multiplications by zero and leads to an efficient realization.



`pCoeffs` points to a coefficient array of size `numTaps`; `pTapDelay` points to an array of nonzero indices and is also of size `numTaps`; `pState` points to a state array of size `maxDelay + blockSize`, where `maxDelay` is the largest offset value that is ever used in the `pTapDelay` array. Some of the processing functions also require temporary working buffers.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient and offset arrays may be shared among several instances while state variable arrays cannot be shared. There are separate instance structure declarations for each of the 4 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numTaps`, `pCoeffs`, `pTapDelay`, `maxDelay`, `stateIndex`, `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. Set the values in the state buffer to zeros before static initialization. The code below statically initializes each of the 4 different data type filter instance structures

**Fixed-Point Behavior** Care must be taken when using the fixed-point versions of the sparse FIR filter functions. In particular, the overflow and saturation behavior of the accumulator used in each function must be considered. Refer to the function specific documentation below for usage guidelines.

## Functions

void **riscv\_fir\_sparse\_f32** (riscv\_fir\_sparse\_instance\_f32 \*S, const float32\_t \*pSrc, float32\_t \*pDst, float32\_t \*pScratchIn, uint32\_t blockSize)

Processing function for the floating-point sparse FIR filter.

**Return** none

### Parameters

- [in] S: points to an instance of the floating-point sparse FIR structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] pScratchIn: points to a temporary buffer of size blockSize
- [in] blockSize: number of input samples to process

void **riscv\_fir\_sparse\_init\_f32** (riscv\_fir\_sparse\_instance\_f32 \*S, uint16\_t numTaps, const float32\_t \*pCoeffs, float32\_t \*pState, int32\_t \*pTapDelay, uint16\_t maxDelay, uint32\_t blockSize)

Initialization function for the floating-point sparse FIR filter.

**Return** none

**Details** pCoeffs holds the filter coefficients and has length numTaps. pState holds the filter's state variables and must be of length maxDelay + blockSize, where maxDelay is the maximum number of delay line values. blockSize is the number of samples processed by the riscv\_fir\_sparse\_f32() function.

### Parameters

- [inout] S: points to an instance of the floating-point sparse FIR structure
- [in] numTaps: number of nonzero coefficients in the filter
- [in] pCoeffs: points to the array of filter coefficients
- [in] pState: points to the state buffer
- [in] pTapDelay: points to the array of offset times
- [in] maxDelay: maximum offset time supported
- [in] blockSize: number of samples that will be processed per block

void **riscv\_fir\_sparse\_init\_q15** (riscv\_fir\_sparse\_instance\_q15 \*S, uint16\_t numTaps, const q15\_t \*pCoeffs, q15\_t \*pState, int32\_t \*pTapDelay, uint16\_t maxDelay, uint32\_t blockSize)

Initialization function for the Q15 sparse FIR filter.

**Return** none

**Details** `pCoeffs` holds the filter coefficients and has length `numTaps`. `pState` holds the filter's state variables and must be of length `maxDelay + blockSize`, where `maxDelay` is the maximum number of delay line values. `blockSize` is the number of words processed by `riscv_fir_sparse_q15()` function.

#### Parameters

- [inout] `S`: points to an instance of the Q15 sparse FIR structure
- [in] `numTaps`: number of nonzero coefficients in the filter
- [in] `pCoeffs`: points to the array of filter coefficients
- [in] `pState`: points to the state buffer
- [in] `pTapDelay`: points to the array of offset times
- [in] `maxDelay`: maximum offset time supported
- [in] `blockSize`: number of samples that will be processed per block

```
void riscv_fir_sparse_init_q31 (riscv_fir_sparse_instance_q31 *S, uint16_t numTaps, const
                               q31_t *pCoeffs, q31_t *pState, int32_t *pTapDelay, uint16_t
                               maxDelay, uint32_t blockSize)
```

Initialization function for the Q31 sparse FIR filter.

**Return** none

**Details** `pCoeffs` holds the filter coefficients and has length `numTaps`. `pState` holds the filter's state variables and must be of length `maxDelay + blockSize`, where `maxDelay` is the maximum number of delay line values. `blockSize` is the number of words processed by `riscv_fir_sparse_q31()` function.

#### Parameters

- [inout] `S`: points to an instance of the Q31 sparse FIR structure
- [in] `numTaps`: number of nonzero coefficients in the filter
- [in] `pCoeffs`: points to the array of filter coefficients
- [in] `pState`: points to the state buffer
- [in] `pTapDelay`: points to the array of offset times
- [in] `maxDelay`: maximum offset time supported
- [in] `blockSize`: number of samples that will be processed per block

```
void riscv_fir_sparse_init_q7 (riscv_fir_sparse_instance_q7 *S, uint16_t numTaps, const
                              q7_t *pCoeffs, q7_t *pState, int32_t *pTapDelay, uint16_t
                              maxDelay, uint32_t blockSize)
```

Initialization function for the Q7 sparse FIR filter.

**Return** none

**Details** `pCoeffs` holds the filter coefficients and has length `numTaps`. `pState` holds the filter's state variables and must be of length `maxDelay + blockSize`, where `maxDelay` is the maximum number of delay line values. `blockSize` is the number of samples processed by the `riscv_fir_sparse_q7()` function.

#### Parameters

- [inout] `S`: points to an instance of the Q7 sparse FIR structure

- [in] numTaps: number of nonzero coefficients in the filter
- [in] pCoeffs: points to the array of filter coefficients
- [in] pState: points to the state buffer
- [in] pTapDelay: points to the array of offset times
- [in] maxDelay: maximum offset time supported
- [in] blockSize: number of samples that will be processed per block

void **riscv\_fir\_sparse\_q15** (riscv\_fir\_sparse\_instance\_q15 \*S, **const** q15\_t \*pSrc, q15\_t \*pDst, q15\_t \*pScratchIn, q31\_t \*pScratchOut, uint32\_t blockSize)

Processing function for the Q15 sparse FIR filter.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 32-bit accumulator. The  $1.15 \times 1.15$  multiplications yield a 2.30 result and these are added to a 2.30 accumulator. Thus the full precision of the multiplications is maintained but there is only a single guard bit in the accumulator. If the accumulator result overflows it will wrap around rather than saturate. After all multiply-accumulates are performed, the 2.30 accumulator is truncated to 2.15 format and then saturated to 1.15 format. In order to avoid overflows the input signal or coefficients must be scaled down by  $\log_2(\text{numTaps})$  bits.

#### Parameters

- [in] S: points to an instance of the Q15 sparse FIR structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] pScratchIn: points to a temporary buffer of size blockSize
- [in] pScratchOut: points to a temporary buffer of size blockSize
- [in] blockSize: number of input samples to process per call

void **riscv\_fir\_sparse\_q31** (riscv\_fir\_sparse\_instance\_q31 \*S, **const** q31\_t \*pSrc, q31\_t \*pDst, q31\_t \*pScratchIn, uint32\_t blockSize)

Processing function for the Q31 sparse FIR filter.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 32-bit accumulator. The  $1.31 \times 1.31$  multiplications are truncated to 2.30 format. This leads to loss of precision on the intermediate multiplications and provides only a single guard bit. If the accumulator result overflows, it wraps around rather than saturate. In order to avoid overflows the input signal or coefficients must be scaled down by  $\log_2(\text{numTaps})$  bits.

#### Parameters

- [in] S: points to an instance of the Q31 sparse FIR structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] pScratchIn: points to a temporary buffer of size blockSize
- [in] blockSize: number of input samples to process



```
void riscv_fir_sparse_q7 (riscv_fir_sparse_instance_q7 *S, const q7_t *pSrc, q7_t *pDst, q7_t
                        *pScratchIn, q31_t *pScratchOut, uint32_t blockSize)
    Processing function for the Q7 sparse FIR filter.
```

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. Both coefficients and state variables are represented in 1.7 format and multiplications yield a 2.14 result. The 2.14 intermediate results are accumulated in a 32-bit accumulator in 18.14 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. The accumulator is then converted to 18.7 format by discarding the low 7 bits. Finally, the result is truncated to 1.7 format.

#### Parameters

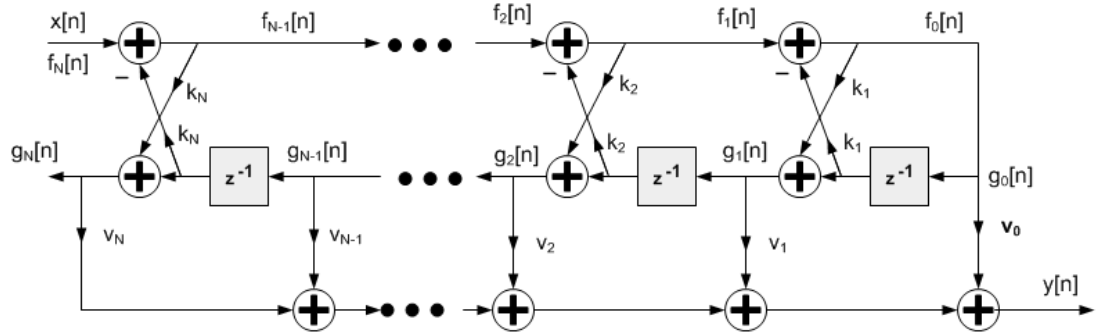
- [in] S: points to an instance of the Q7 sparse FIR structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] pScratchIn: points to a temporary buffer of size blockSize
- [in] pScratchOut: points to a temporary buffer of size blockSize
- [in] blockSize: number of input samples to process

### Infinite Impulse Response (IIR) Lattice Filters

```
void riscv_iir_lattice_f32 (const riscv_iir_lattice_instance_f32 *S, const float32_t *pSrc,
                          float32_t *pDst, uint32_t blockSize)
void riscv_iir_lattice_init_f32 (riscv_iir_lattice_instance_f32 *S, uint16_t numStages, float32_t
                                *pkCoeffs, float32_t *pvCoeffs, float32_t *pState, uint32_t block-
                                Size)
void riscv_iir_lattice_init_q15 (riscv_iir_lattice_instance_q15 *S, uint16_t numStages, q15_t
                                *pkCoeffs, q15_t *pvCoeffs, q15_t *pState, uint32_t blockSize)
void riscv_iir_lattice_init_q31 (riscv_iir_lattice_instance_q31 *S, uint16_t numStages, q31_t
                                *pkCoeffs, q31_t *pvCoeffs, q31_t *pState, uint32_t blockSize)
void riscv_iir_lattice_q15 (const riscv_iir_lattice_instance_q15 *S, const q15_t *pSrc, q15_t
                            *pDst, uint32_t blockSize)
void riscv_iir_lattice_q31 (const riscv_iir_lattice_instance_q31 *S, const q31_t *pSrc, q31_t
                            *pDst, uint32_t blockSize)
```

#### group IIR\_Lattice

This set of functions implements lattice filters for Q15, Q31 and floating-point data types. Lattice filters are used in a variety of adaptive filter applications. The filter structure has feedforward and feedback components and the net impulse response is infinite length. The functions operate on blocks of input and output data and each call to the function processes `blockSize` samples through the filter. `pSrc` and `pDst` point to input and output arrays containing `blockSize` values.



### Algorithm

`pkCoeffs` points to array of reflection coefficients of size `numStages`. Reflection Coefficients are stored in time-reversed order.

`pvCoeffs` points to the array of ladder coefficients of size `(numStages+1)`. Ladder coefficients are stored in time-reversed order.

`pState` points to a state array of size `numStages + blockSize`. The state variables shown in the figure above (the `g` values) are stored in the `pState` array. The state variables are updated after each block of data is processed; the coefficients are untouched.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient arrays may be shared among several instances while state variable arrays cannot be shared. There are separate instance structure declarations for each of the 3 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numStages`, `pkCoeffs`, `pvCoeffs`, `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. Set the values in the state buffer to zeros and then manually initialize the instance structure as follows:

where `numStages` is the number of stages in the filter; `pState` points to the state buffer array; `pkCoeffs` points to array of the reflection coefficients; `pvCoeffs` points to the array of ladder coefficients.

**Fixed-Point Behavior** Care must be taken when using the fixed-point versions of the IIR lattice filter functions. In particular, the overflow and saturation behavior of the accumulator used in each function must be considered. Refer to the function specific documentation below for usage guidelines.

### Functions

```
void riscv_iir_lattice_f32 (const riscv_iir_lattice_instance_f32 *S, const float32_t *pSrc,
                           float32_t *pDst, uint32_t blockSize)
    Processing function for the floating-point IIR lattice filter.
```

**Return** none

**Parameters**

- [in] *S*: points to an instance of the floating-point IIR lattice structure
- [in] *pSrc*: points to the block of input data
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process

```
void riscv_iir_lattice_init_f32 (riscv_iir_lattice_instance_f32 *S, uint16_t numStages,
                                float32_t *pkCoeffs, float32_t *pvCoeffs, float32_t *pState,
                                uint32_t blockSize)
```

Initialization function for the floating-point IIR lattice filter.

**Return** none

**Parameters**

- [in] *S*: points to an instance of the floating-point IIR lattice structure
- [in] *numStages*: number of stages in the filter
- [in] *pkCoeffs*: points to reflection coefficient buffer. The array is of length *numStages*
- [in] *pvCoeffs*: points to ladder coefficient buffer. The array is of length *numStages*+1
- [in] *pState*: points to state buffer. The array is of length *numStages*+*blockSize*
- [in] *blockSize*: number of samples to process

```
void riscv_iir_lattice_init_q15 (riscv_iir_lattice_instance_q15 *S, uint16_t numStages,
                                q15_t *pkCoeffs, q15_t *pvCoeffs, q15_t *pState, uint32_t
                                blockSize)
```

Initialization function for the Q15 IIR lattice filter.

**Return** none

**Parameters**

- [in] *S*: points to an instance of the Q15 IIR lattice structure
- [in] *numStages*: number of stages in the filter
- [in] *pkCoeffs*: points to reflection coefficient buffer. The array is of length *numStages*
- [in] *pvCoeffs*: points to ladder coefficient buffer. The array is of length *numStages*+1
- [in] *pState*: points to state buffer. The array is of length *numStages*+*blockSize*
- [in] *blockSize*: number of samples to process

```
void riscv_iir_lattice_init_q31 (riscv_iir_lattice_instance_q31 *S, uint16_t numStages,
                                q31_t *pkCoeffs, q31_t *pvCoeffs, q31_t *pState, uint32_t
                                blockSize)
```

Initialization function for the Q31 IIR lattice filter.

**Return** none

**Parameters**

- [in] *S*: points to an instance of the Q31 IIR lattice structure
- [in] *numStages*: number of stages in the filter
- [in] *pkCoeffs*: points to reflection coefficient buffer. The array is of length *numStages*

- [in] `pvCoeffs`: points to ladder coefficient buffer. The array is of length `numStages+1`
- [in] `pState`: points to state buffer. The array is of length `numStages+blockSize`
- [in] `blockSize`: number of samples to process

```
void riscv_iir_lattice_q15(const riscv_iir_lattice_instance_q15 *S, const q15_t *pSrc,  
                          q15_t *pDst, uint32_t blockSize)  
Processing function for the Q15 IIR lattice filter.
```

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. Both coefficients and state variables are represented in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. After all additions have been performed, the accumulator is truncated to 34.15 format by discarding low 15 bits. Lastly, the accumulator is saturated to yield a result in 1.15 format.

**Parameters**

- [in] `S`: points to an instance of the Q15 IIR lattice structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

```
void riscv_iir_lattice_q31(const riscv_iir_lattice_instance_q31 *S, const q31_t *pSrc,  
                          q31_t *pDst, uint32_t blockSize)  
Processing function for the Q31 IIR lattice filter.
```

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clip. In order to avoid overflows completely the input signal must be scaled down by  $2 \cdot \log_2(\text{numStages})$  bits. After all multiply-accumulates are performed, the 2.62 accumulator is saturated to 1.32 format and then truncated to 1.31 format.

**Parameters**

- [in] `S`: points to an instance of the Q31 IIR lattice structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

## Levinson Durbin Algorithm

```
void riscv_levinson_durbin_f16(const float16_t *phi, float16_t *a, float16_t *err, int nbCoefs)
```

```
void riscv_levinson_durbin_f32(const float32_t *phi, float32_t *a, float32_t *err, int nbCoefs)
```

```
void riscv_levinson_durbin_q31(const q31_t *phi, q31_t *a, q31_t *err, int nbCoefs)
```

group LD

## Functions

void **riscv\_levinson\_durbin\_f16**(**const** float16\_t \**phi*, float16\_t \**a*, float16\_t \**err*, int *nbCoefs*)

Levinson Durbin.

**Return** none

### Parameters

- [in] *phi*: autocovariance vector starting with lag 0 (length is nbCoefs + 1)
- [out] *a*: autoregressive coefficients
- [out] *err*: prediction error (variance)
- [in] *nbCoefs*: number of autoregressive coefficients

void **riscv\_levinson\_durbin\_f32**(**const** float32\_t \**phi*, float32\_t \**a*, float32\_t \**err*, int *nbCoefs*)

Levinson Durbin.

**Return** none

### Parameters

- [in] *phi*: autocovariance vector starting with lag 0 (length is nbCoefs + 1)
- [out] *a*: autoregressive coefficients
- [out] *err*: prediction error (variance)
- [in] *nbCoefs*: number of autoregressive coefficients

void **riscv\_levinson\_durbin\_q31**(**const** q31\_t \**phi*, q31\_t \**a*, q31\_t \**err*, int *nbCoefs*)

Levinson Durbin.

**Return** none

### Parameters

- [in] *phi*: autocovariance vector starting with lag 0 (length is nbCoefs + 1)
- [out] *a*: autoregressive coefficients
- [out] *err*: prediction error (variance)
- [in] *nbCoefs*: number of autoregressive coefficients

## Least Mean Square (LMS) Filters

void **riscv\_lms\_f32**(**const** riscv\_lms\_instance\_f32 \**S*, **const** float32\_t \**pSrc*, float32\_t \**pRef*, float32\_t \**pOut*, float32\_t \**pErr*, uint32\_t *blockSize*)

void **riscv\_lms\_init\_f32**(riscv\_lms\_instance\_f32 \**S*, uint16\_t *numTaps*, float32\_t \**pCoeffs*, float32\_t \**pState*, float32\_t *mu*, uint32\_t *blockSize*)

void **riscv\_lms\_init\_q15**(riscv\_lms\_instance\_q15 \**S*, uint16\_t *numTaps*, q15\_t \**pCoeffs*, q15\_t \**pState*, q15\_t *mu*, uint32\_t *blockSize*, uint32\_t *postShift*)

void **riscv\_lms\_init\_q31**(riscv\_lms\_instance\_q31 \**S*, uint16\_t *numTaps*, q31\_t \**pCoeffs*, q31\_t \**pState*, q31\_t *mu*, uint32\_t *blockSize*, uint32\_t *postShift*)

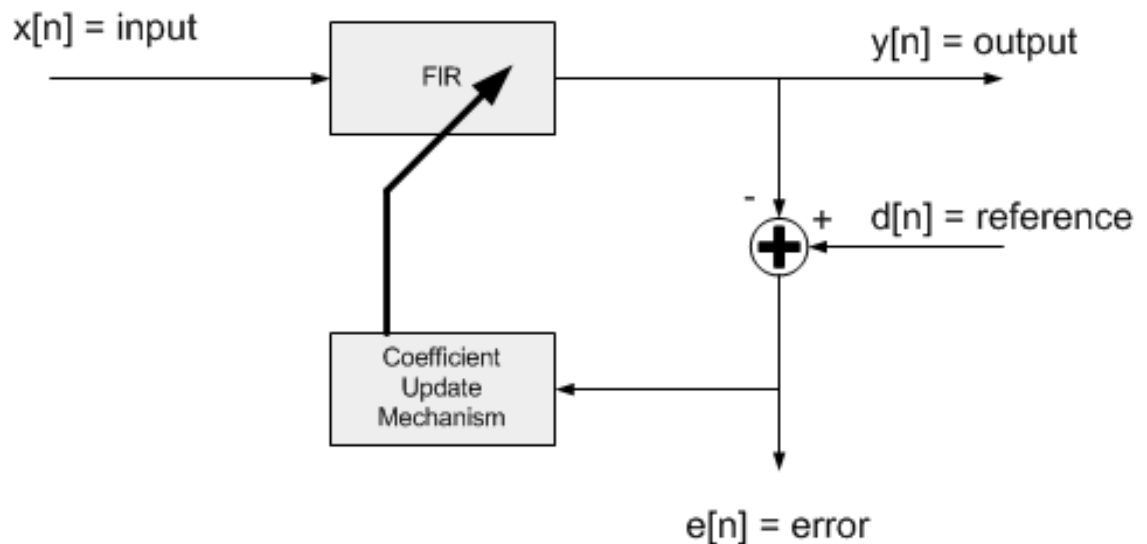
```
void riscv_lms_q15 (const riscv_lms_instance_q15 *S, const q15_t *pSrc, q15_t *pRef, q15_t *pOut,
                  q15_t *pErr, uint32_t blockSize)
```

```
void riscv_lms_q31 (const riscv_lms_instance_q31 *S, const q31_t *pSrc, q31_t *pRef, q31_t *pOut,
                  q31_t *pErr, uint32_t blockSize)
```

#### group LMS

LMS filters are a class of adaptive filters that are able to “learn” an unknown transfer functions. LMS filters use a gradient descent method in which the filter coefficients are updated based on the instantaneous error signal. Adaptive filters are often used in communication systems, equalizers, and noise removal. The NMSIS DSP Library contains LMS filter functions that operate on Q15, Q31, and floating-point data types. The library also contains normalized LMS filters in which the filter coefficient adaptation is independent of the level of the input signal.

An LMS filter consists of two components as shown below. The first component is a standard transversal or FIR filter. The second component is a coefficient update mechanism. The LMS filter has two input signals. The “input” feeds the FIR filter while the “reference input” corresponds to the desired output of the FIR filter. That is, the FIR filter coefficients are updated so that the output of the FIR filter matches the reference input. The filter coefficient update mechanism is based on the difference between the FIR filter output and the reference input. This “error signal” tends towards zero as the filter adapts. The LMS processing functions accept the input and reference input signals and generate the filter output and error signal. The functions operate on blocks of data and each call to the function processes `blockSize` samples through the filter. `pSrc` points to input signal, `pRef` points to reference signal, `pOut` points to output signal and `pErr` points to error signal. All arrays contain `blockSize`



values.

The functions operate on a block-by-block basis. Internally, the filter coefficients  $b[n]$  are updated on a sample-by-sample basis. The convergence of the LMS filter is slower compared to the normalized LMS algorithm.

**Algorithm** The output signal  $y[n]$  is computed by a standard FIR filter:

The error signal equals the difference between the reference signal  $d[n]$  and the filter output:

After each sample of the error signal is computed, the filter coefficients  $b[k]$  are updated on a sample-by-sample basis: where  $\mu$  is the step size and controls the rate of coefficient convergence.

In the APIs, `pCoeffs` points to a coefficient array of size `numTaps`. Coefficients are stored in time reversed order.

`pState` points to a state array of size `numTaps + blockSize - 1`. Samples in the state buffer are stored in the order:

Note that the length of the state buffer exceeds the length of the coefficient array by `blockSize-1` samples. The increased state buffer length allows circular addressing, which is traditionally used in FIR filters, to be avoided and yields a significant speed improvement. The state variables are updated after each block of data is processed.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter and coefficient and state arrays cannot be shared among instances. There are separate instance structure declarations for each of the 3 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numTaps`, `pCoeffs`, `mu`, `postShift` (not for `f32`), `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. Set the values in the state buffer to zeros before static initialization. The code below statically initializes each of the 3 different data type filter instance structures where `numTaps` is the number of filter coefficients in the filter; `pState` is the address of the state buffer; `pCoeffs` is the address of the coefficient buffer; `mu` is the step size parameter; and `postShift` is the shift applied to coefficients.

**Fixed-Point Behavior** Care must be taken when using the Q15 and Q31 versions of the LMS filter. The following issues must be considered:

- Scaling of coefficients
- Overflow and saturation

**Scaling of Coefficients** Filter coefficients are represented as fractional values and coefficients are restricted to lie in the range  $[-1 \ +1)$ . The fixed-point functions have an additional scaling parameter `postShift`. At the output of the filter's accumulator is a shift register which shifts the result by `postShift` bits. This essentially scales the filter coefficients by  $2^{\text{postShift}}$  and allows the filter coefficients to exceed the range  $[-1 \ +1)$ . The value of `postShift` is set by the user based on the expected gain through the system being modeled.

**Overflow and Saturation** Overflow and saturation behavior of the fixed-point Q15 and Q31 versions are described separately as part of the function specific documentation below.

## Functions

```
void riscv_lms_f32 (const riscv_lms_instance_f32 *S, const float32_t *pSrc, float32_t *pRef,
                   float32_t *pOut, float32_t *pErr, uint32_t blockSize)
Processing function for floating-point LMS filter.
```

**Return** none

**Parameters**

- `[in]` `S`: points to an instance of the floating-point LMS filter structure

- [in] `pSrc`: points to the block of input data
- [in] `pRef`: points to the block of reference data
- [out] `pOut`: points to the block of output data
- [out] `pErr`: points to the block of error data
- [in] `blockSize`: number of samples to process

void **riscv\_lms\_init\_f32** (riscv\_lms\_instance\_f32 \*S, uint16\_t numTaps, float32\_t \*pCoeffs, float32\_t \*pState, float32\_t mu, uint32\_t blockSize)

Initialization function for floating-point LMS filter.

**Return** none

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: The initial filter coefficients serve as a starting point for the adaptive filter. `pState` points to an array of length `numTaps+blockSize-1` samples, where `blockSize` is the number of input samples processed by each call to `riscv_lms_f32()`.

**Parameters**

- [in] `S`: points to an instance of the floating-point LMS filter structure
- [in] `numTaps`: number of filter coefficients
- [in] `pCoeffs`: points to coefficient buffer
- [in] `pState`: points to state buffer
- [in] `mu`: step size that controls filter coefficient updates
- [in] `blockSize`: number of samples to process

void **riscv\_lms\_init\_q15** (riscv\_lms\_instance\_q15 \*S, uint16\_t numTaps, q15\_t \*pCoeffs, q15\_t \*pState, q15\_t mu, uint32\_t blockSize, uint32\_t postShift)

Initialization function for the Q15 LMS filter.

**Return** none

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: The initial filter coefficients serve as a starting point for the adaptive filter. `pState` points to the array of state variables and size of array is `numTaps+blockSize-1` samples, where `blockSize` is the number of input samples processed by each call to `riscv_lms_q15()`.

**Parameters**

- [in] `S`: points to an instance of the Q15 LMS filter structure.
- [in] `numTaps`: number of filter coefficients.
- [in] `pCoeffs`: points to coefficient buffer.
- [in] `pState`: points to state buffer.
- [in] `mu`: step size that controls filter coefficient updates.
- [in] `blockSize`: number of samples to process.
- [in] `postShift`: bit shift applied to coefficients.

void **riscv\_lms\_init\_q31** (riscv\_lms\_instance\_q31 \*S, uint16\_t numTaps, q31\_t \*pCoeffs, q31\_t \*pState, q31\_t mu, uint32\_t blockSize, uint32\_t postShift)

Initialization function for Q31 LMS filter.



**Return** none

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order. The initial filter coefficients serve as a starting point for the adaptive filter. `pState` points to an array of length `numTaps+blockSize-1` samples, where `blockSize` is the number of input samples processed by each call to `riscv_lms_q31()`.

**Parameters**

- [in] `S`: points to an instance of the Q31 LMS filter structure
- [in] `numTaps`: number of filter coefficients
- [in] `pCoeffs`: points to coefficient buffer
- [in] `pState`: points to state buffer
- [in] `mu`: step size that controls filter coefficient updates
- [in] `blockSize`: number of samples to process
- [in] `postShift`: bit shift applied to coefficients

```
void riscv_lms_q15(const riscv_lms_instance_q15 *S, const q15_t *pSrc, q15_t *pRef, q15_t
                  *pOut, q15_t *pErr, uint32_t blockSize)
Processing function for Q15 LMS filter.
```

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. Both coefficients and state variables are represented in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. After all additions have been performed, the accumulator is truncated to 34.15 format by discarding low 15 bits. Lastly, the accumulator is saturated to yield a result in 1.15 format.

In this filter, filter coefficients are updated for each sample and the updation of filter coefficients are saturated.

**Parameters**

- [in] `S`: points to an instance of the Q15 LMS filter structure
- [in] `pSrc`: points to the block of input data
- [in] `pRef`: points to the block of reference data
- [out] `pOut`: points to the block of output data
- [out] `pErr`: points to the block of error data
- [in] `blockSize`: number of samples to process

```
void riscv_lms_q31(const riscv_lms_instance_q31 *S, const q31_t *pSrc, q31_t *pRef, q31_t
                  *pOut, q31_t *pErr, uint32_t blockSize)
Processing function for Q31 LMS filter.
```

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clips. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{numTaps})$  bits. The reference signal should not be scaled down. After all multiply-accumulates

are performed, the 2.62 accumulator is shifted and saturated to 1.31 format to yield the final result. The output signal and error signal are in 1.31 format.

In this filter, filter coefficients are updated for each sample and the updation of filter coefficients are saturated.

#### Parameters

- [in] *S*: points to an instance of the Q31 LMS filter structure.
- [in] *pSrc*: points to the block of input data.
- [in] *pRef*: points to the block of reference data.
- [out] *pOut*: points to the block of output data.
- [out] *pErr*: points to the block of error data.
- [in] *blockSize*: number of samples to process.

### Normalized LMS Filters

```
void riscv_lms_norm_f32 (riscv_lms_norm_instance_f32 *S, const float32_t *pSrc, float32_t *pRef,
                        float32_t *pOut, float32_t *pErr, uint32_t blockSize)
```

```
void riscv_lms_norm_init_f32 (riscv_lms_norm_instance_f32 *S, uint16_t numTaps, float32_t *pCoeffs,
                             float32_t *pState, float32_t mu, uint32_t blockSize)
```

```
void riscv_lms_norm_init_q15 (riscv_lms_norm_instance_q15 *S, uint16_t numTaps, q15_t *pCoeffs,
                             q15_t *pState, q15_t mu, uint32_t blockSize, uint8_t postShift)
```

```
void riscv_lms_norm_init_q31 (riscv_lms_norm_instance_q31 *S, uint16_t numTaps, q31_t *pCoeffs,
                             q31_t *pState, q31_t mu, uint32_t blockSize, uint8_t postShift)
```

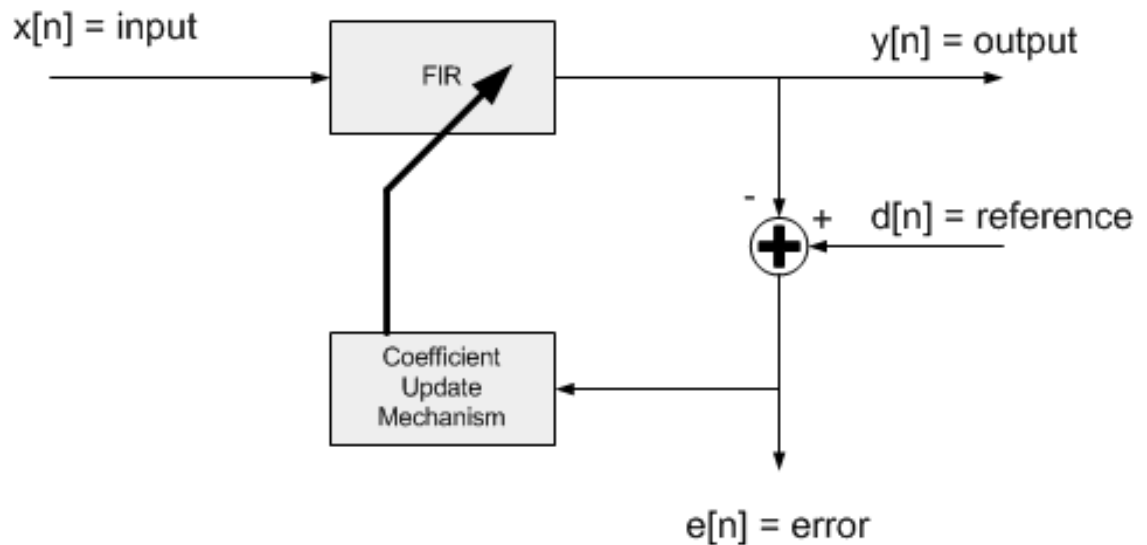
```
void riscv_lms_norm_q15 (riscv_lms_norm_instance_q15 *S, const q15_t *pSrc, q15_t *pRef, q15_t
                        *pOut, q15_t *pErr, uint32_t blockSize)
```

```
void riscv_lms_norm_q31 (riscv_lms_norm_instance_q31 *S, const q31_t *pSrc, q31_t *pRef, q31_t
                        *pOut, q31_t *pErr, uint32_t blockSize)
```

#### group **LMS\_NORM**

This set of functions implements a commonly used adaptive filter. It is related to the Least Mean Square (LMS) adaptive filter and includes an additional normalization factor which increases the adaptation rate of the filter. The NMSIS DSP Library contains normalized LMS filter functions that operate on Q15, Q31, and floating-point data types.

A normalized least mean square (NLMS) filter consists of two components as shown below. The first component is a standard transversal or FIR filter. The second component is a coefficient update mechanism. The NLMS filter has two input signals. The “input” feeds the FIR filter while the “reference input” corresponds to the desired output of the FIR filter. That is, the FIR filter coefficients are updated so that the output of the FIR filter matches the reference input. The filter coefficient update mechanism is based on the difference between the FIR filter output and the reference input. This “error signal” tends towards zero as the filter adapts. The NLMS processing functions accept the input and reference input signals and generate the filter output and error signal. The functions operate on blocks of data and each call to the function processes *blockSize* samples through the filter. *pSrc* points to input signal, *pRef* points to reference signal, *pOut* points to output signal and *pErr* points to error signal. All arrays contain *blockSize*



values.

The functions operate on a block-by-block basis. Internally, the filter coefficients  $b[n]$  are updated on a sample-by-sample basis. The convergence of the LMS filter is slower compared to the normalized LMS algorithm.

**Algorithm** The output signal  $y[n]$  is computed by a standard FIR filter:

The error signal equals the difference between the reference signal  $d[n]$  and the filter output:

After each sample of the error signal is computed the instantaneous energy of the filter state variables is calculated: The filter coefficients  $b[k]$  are then updated on a sample-by-sample basis: where  $\mu$  is the step size and controls the rate of coefficient convergence.

In the APIs, `pCoeffs` points to a coefficient array of size `numTaps`. Coefficients are stored in time reversed order.

`pState` points to a state array of size `numTaps + blockSize - 1`. Samples in the state buffer are stored in the order:

Note that the length of the state buffer exceeds the length of the coefficient array by `blockSize-1` samples.

The increased state buffer length allows circular addressing, which is traditionally used in FIR filters, to be avoided and yields a significant speed improvement. The state variables are updated after each block of data is processed.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter and coefficient and state arrays cannot be shared among instances. There are separate instance structure declarations for each of the 3 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer. To do this manually without calling the init function, assign the follow subfields of the instance structure: `numTaps`, `pCoeffs`, `mu`, `energy`, `x0`, `pState`. Also set all of the values in `pState` to zero. For Q7, Q15, and Q31 the following fields must also be initialized; `recipTable`, `postShift`

Instance structure cannot be placed into a const data section and it is recommended to use the initialization function.

**Fixed-Point Behavior** Care must be taken when using the Q15 and Q31 versions of the normalised LMS filter. The following issues must be considered:

- Scaling of coefficients
- Overflow and saturation

**Scaling of Coefficients (fixed point versions)** Filter coefficients are represented as fractional values and coefficients are restricted to lie in the range  $[-1 \ +1)$ . The fixed-point functions have an additional scaling parameter `postShift`. At the output of the filter's accumulator is a shift register which shifts the result by `postShift` bits. This essentially scales the filter coefficients by  $2^{\text{postShift}}$  and allows the filter coefficients to exceed the range  $[+1 \ -1)$ . The value of `postShift` is set by the user based on the expected gain through the system being modeled.

**Overflow and Saturation (fixed point versions)** Overflow and saturation behavior of the fixed-point Q15 and Q31 versions are described separately as part of the function specific documentation below.

## Functions

void **riscv\_lms\_norm\_f32** (riscv\_lms\_norm\_instance\_f32 \*S, const float32\_t \*pSrc, float32\_t \*pRef, float32\_t \*pOut, float32\_t \*pErr, uint32\_t blockSize)  
Processing function for floating-point normalized LMS filter.

**Return** none

### Parameters

- [in] S: points to an instance of the floating-point normalized LMS filter structure
- [in] pSrc: points to the block of input data
- [in] pRef: points to the block of reference data
- [out] pOut: points to the block of output data
- [out] pErr: points to the block of error data
- [in] blockSize: number of samples to process

void **riscv\_lms\_norm\_init\_f32** (riscv\_lms\_norm\_instance\_f32 \*S, uint16\_t numTaps, float32\_t \*pCoeffs, float32\_t \*pState, float32\_t mu, uint32\_t blockSize)  
Initialization function for floating-point normalized LMS filter.

**Return** none

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: The initial filter coefficients serve as a starting point for the adaptive filter. `pState` points to an array of length `numTaps+blockSize-1` samples, where `blockSize` is the number of input samples processed by each call to `riscv_lms_norm_f32()`.

### Parameters

- [in] S: points to an instance of the floating-point LMS filter structure
- [in] numTaps: number of filter coefficients
- [in] pCoeffs: points to coefficient buffer
- [in] pState: points to state buffer

- [in] `mu`: step size that controls filter coefficient updates
- [in] `blockSize`: number of samples to process

```
void riscv_lms_norm_init_q15(riscv_lms_norm_instance_q15 *S, uint16_t numTaps, q15_t
                           *pCoeffs, q15_t *pState, q15_t mu, uint32_t blockSize, uint8_t
                           postShift)
```

Initialization function for Q15 normalized LMS filter.

**Return** none

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: The initial filter coefficients serve as a starting point for the adaptive filter. `pState` points to the array of state variables and size of array is `numTaps+blockSize-1` samples, where `blockSize` is the number of input samples processed by each call to `riscv_lms_norm_q15()`.

#### Parameters

- [in] `S`: points to an instance of the Q15 normalized LMS filter structure.
- [in] `numTaps`: number of filter coefficients.
- [in] `pCoeffs`: points to coefficient buffer.
- [in] `pState`: points to state buffer.
- [in] `mu`: step size that controls filter coefficient updates.
- [in] `blockSize`: number of samples to process.
- [in] `postShift`: bit shift applied to coefficients.

```
void riscv_lms_norm_init_q31(riscv_lms_norm_instance_q31 *S, uint16_t numTaps, q31_t
                           *pCoeffs, q31_t *pState, q31_t mu, uint32_t blockSize, uint8_t
                           postShift)
```

Initialization function for Q31 normalized LMS filter.

**Return** none

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: The initial filter coefficients serve as a starting point for the adaptive filter. `pState` points to an array of length `numTaps+blockSize-1` samples, where `blockSize` is the number of input samples processed by each call to `riscv_lms_norm_q31()`.

#### Parameters

- [in] `S`: points to an instance of the Q31 normalized LMS filter structure.
- [in] `numTaps`: number of filter coefficients.
- [in] `pCoeffs`: points to coefficient buffer.
- [in] `pState`: points to state buffer.
- [in] `mu`: step size that controls filter coefficient updates.
- [in] `blockSize`: number of samples to process.
- [in] `postShift`: bit shift applied to coefficients.

```
void riscv_lms_norm_q15(riscv_lms_norm_instance_q15 *S, const q15_t *pSrc, q15_t *pRef,
                       q15_t *pOut, q15_t *pErr, uint32_t blockSize)
```

Processing function for Q15 normalized LMS filter.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both coefficients and state variables are represented in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. After all additions have been performed, the accumulator is truncated to 34.15 format by discarding low 15 bits. Lastly, the accumulator is saturated to yield a result in 1.15 format.

In this filter, filter coefficients are updated for each sample and the updation of filter coefficients are saturated.

**Parameters**

- [in] *S*: points to an instance of the Q15 normalized LMS filter structure
- [in] *pSrc*: points to the block of input data
- [in] *pRef*: points to the block of reference data
- [out] *pOut*: points to the block of output data
- [out] *pErr*: points to the block of error data
- [in] *blockSize*: number of samples to process

```
void riscv_lms_norm_q31 (riscv_lms_norm_instance_q31 *S, const q31_t *pSrc, q31_t *pRef,  
                        q31_t *pOut, q31_t *pErr, uint32_t blockSize)
```

Processing function for Q31 normalized LMS filter.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clip. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{numTaps})$  bits. The reference signal should not be scaled down. After all multiply-accumulates are performed, the 2.62 accumulator is shifted and saturated to 1.31 format to yield the final result. The output signal and error signal are in 1.31 format.

In this filter, filter coefficients are updated for each sample and the updation of filter coefficients are saturated.

**Parameters**

- [in] *S*: points to an instance of the Q31 normalized LMS filter structure
- [in] *pSrc*: points to the block of input data
- [in] *pRef*: points to the block of reference data
- [out] *pOut*: points to the block of output data
- [out] *pErr*: points to the block of error data
- [in] *blockSize*: number of samples to process

## Finite Impulse Response (FIR) Interpolator

```
void riscv_fir_interpolate_f32 (const riscv_fir_interpolate_instance_f32 *S, const float32_t  
                               *pSrc, float32_t *pDst, uint32_t blockSize)
```

```

riscv_status riscv_fir_interpolate_init_f32 (riscv_fir_interpolate_instance_f32 *S, uint8_t L,
                                             uint16_t numTaps, const float32_t *pCoeffs,
                                             float32_t *pState, uint32_t blockSize)

riscv_status riscv_fir_interpolate_init_q15 (riscv_fir_interpolate_instance_q15 *S, uint8_t L,
                                             uint16_t numTaps, const q15_t *pCoeffs, q15_t
                                             *pState, uint32_t blockSize)

riscv_status riscv_fir_interpolate_init_q31 (riscv_fir_interpolate_instance_q31 *S, uint8_t L,
                                             uint16_t numTaps, const q31_t *pCoeffs, q31_t
                                             *pState, uint32_t blockSize)

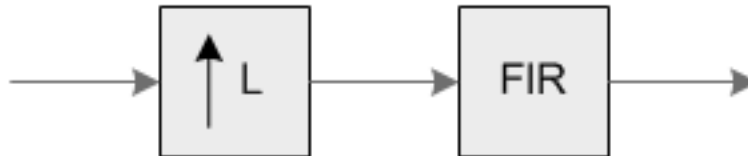
void riscv_fir_interpolate_q15 (const riscv_fir_interpolate_instance_q15 *S, const q15_t
                              *pSrc, q15_t *pDst, uint32_t blockSize)

void riscv_fir_interpolate_q31 (const riscv_fir_interpolate_instance_q31 *S, const q31_t
                              *pSrc, q31_t *pDst, uint32_t blockSize)

```

#### group **FIR\_Interpolate**

These functions combine an upsampler (zero stuffer) and an FIR filter. They are used in multirate systems for increasing the sample rate of a signal without introducing high frequency images. Conceptually, the functions are equivalent to the block diagram below: After upsampling by a factor of  $L$ , the signal should be filtered by a lowpass filter with a normalized cutoff frequency of  $1/L$  in order to eliminate high frequency copies of the spectrum. The user of the function is responsible for providing the filter



coefficients.

The FIR interpolator functions provided in the NMSIS DSP Library combine the upsampler and FIR filter in an efficient manner. The upsampler inserts  $L-1$  zeros between each sample. Instead of multiplying by these zero values, the FIR filter is designed to skip them. This leads to an efficient implementation without any wasted effort. The functions operate on blocks of input and output data. `pSrc` points to an array of `blockSize` input values and `pDst` points to an array of `blockSize*L` output values.

The library provides separate functions for Q15, Q31, and floating-point data types.

**Algorithm** The functions use a polyphase filter structure: This approach is more efficient than straightforward upsample-then-filter algorithms. With this method the computation is reduced by a factor of  $1/L$  when compared to using a standard FIR filter.

`pCoeffs` points to a coefficient array of size `numTaps`. `numTaps` must be a multiple of the interpolation factor  $L$  and this is checked by the initialization functions. Internally, the function divides the FIR filter's impulse response into shorter filters of length `phaseLength=numTaps/L`. Coefficients are stored in time reversed order.

`pState` points to a state array of size `blockSize + phaseLength - 1`. Samples in the state buffer are stored in the order:

The state variables are updated after each block of data is processed, the coefficients are untouched.

**Instance Structure** The coefficients and state variables for a filter are stored together in an instance data structure. A separate instance structure must be defined for each filter. Coefficient arrays may be shared among several instances while state variable array should be allocated separately. There are separate instance structure declarations for each of the 3 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Zeros out the values in the state buffer.
- Checks to make sure that the length of the filter is a multiple of the interpolation factor. To do this manually without calling the init function, assign the follow subfields of the instance structure: `L` (interpolation factor), `pCoeffs`, `phaseLength` (`numTaps / L`), `pState`. Also set all of the values in `pState` to zero.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. The code below statically initializes each of the 3 different data type filter instance structures

where `L` is the interpolation factor; `phaseLength=numTaps/L` is the length of each of the shorter FIR filters used internally, `pCoeffs` is the address of the coefficient buffer; `pState` is the address of the state buffer. Be sure to set the values in the state buffer to zeros when doing static initialization.

**Fixed-Point Behavior** Care must be taken when using the fixed-point versions of the FIR interpolate filter functions. In particular, the overflow and saturation behavior of the accumulator used in each function must be considered. Refer to the function specific documentation below for usage guidelines.

## Functions

```
void riscv_fir_interpolate_f32 (const  riscv_fir_interpolate_instance_f32  *S,  const
                                float32_t *pSrc, float32_t *pDst, uint32_t blockSize)
```

Processing function for floating-point FIR interpolator.

Processing function for the floating-point FIR interpolator.

**Return** none

### Parameters

- [in] `S`: points to an instance of the floating-point FIR interpolator structure
- [in] `pSrc`: points to the block of input data
- [out] `pDst`: points to the block of output data
- [in] `blockSize`: number of samples to process

```
riscv_status riscv_fir_interpolate_init_f32 (riscv_fir_interpolate_instance_f32 *S, uint8_t
                                              L, uint16_t numTaps, const float32_t *pCo-
                                              effs, float32_t *pState, uint32_t blockSize)
```

Initialization function for the floating-point FIR interpolator.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_ARGUMENT_ERROR` : filter length `numTaps` is not a multiple of the interpolation factor `L`

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order:

The length of the filter `numTaps` must be a multiple of the interpolation factor `L`.



`pState` points to the array of state variables. `pState` is of length  $(\text{numTaps}/L) + \text{blockSize} - 1$  words where `blockSize` is the number of input samples processed by each call to `riscv_fir_interpolate_f32()`.

#### Parameters

- [inout] `S`: points to an instance of the floating-point FIR interpolator structure
- [in] `L`: upsample factor
- [in] `numTaps`: number of filter coefficients in the filter
- [in] `pCoeffs`: points to the filter coefficient buffer
- [in] `pState`: points to the state buffer
- [in] `blockSize`: number of input samples to process per call

`riscv_status riscv_fir_interpolate_init_q15` (`riscv_fir_interpolate_instance_q15 *S`, `uint8_t L`, `uint16_t numTaps`, **const** `q15_t *pCoeffs`, `q15_t *pState`, `uint32_t blockSize`)

Initialization function for the Q15 FIR interpolator.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_ARGUMENT_ERROR` : filter length `numTaps` is not a multiple of the interpolation factor `L`

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: The length of the filter `numTaps` must be a multiple of the interpolation factor `L`.

`pState` points to the array of state variables. `pState` is of length  $(\text{numTaps}/L) + \text{blockSize} - 1$  words where `blockSize` is the number of input samples processed by each call to `riscv_fir_interpolate_q15()`.

#### Parameters

- [inout] `S`: points to an instance of the Q15 FIR interpolator structure
- [in] `L`: upsample factor
- [in] `numTaps`: number of filter coefficients in the filter
- [in] `pCoeffs`: points to the filter coefficient buffer
- [in] `pState`: points to the state buffer
- [in] `blockSize`: number of input samples to process per call

`riscv_status riscv_fir_interpolate_init_q31` (`riscv_fir_interpolate_instance_q31 *S`, `uint8_t L`, `uint16_t numTaps`, **const** `q31_t *pCoeffs`, `q31_t *pState`, `uint32_t blockSize`)

Initialization function for the Q31 FIR interpolator.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_ARGUMENT_ERROR` : filter length `numTaps` is not a multiple of the interpolation factor `L`

**Details** `pCoeffs` points to the array of filter coefficients stored in time reversed order: The length of the filter `numTaps` must be a multiple of the interpolation factor `L`.

pState points to the array of state variables. pState is of length  $(\text{numTaps}/L) + \text{blockSize} - 1$  words where blockSize is the number of input samples processed by each call to `riscv_fir_interpolate_q31()`.

#### Parameters

- [inout] S: points to an instance of the Q31 FIR interpolator structure
- [in] L: upsample factor
- [in] numTaps: number of filter coefficients in the filter
- [in] pCoeffs: points to the filter coefficient buffer
- [in] pState: points to the state buffer
- [in] blockSize: number of input samples to process per call

```
void riscv_fir_interpolate_q15 (const riscv_fir_interpolate_instance_q15 *S, const q15_t  
                               *pSrc, q15_t *pDst, uint32_t blockSize)
```

Processing function for the Q15 FIR interpolator.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. Both coefficients and state variables are represented in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. There is no risk of internal overflow with this approach and the full precision of intermediate multiplications is preserved. After all additions have been performed, the accumulator is truncated to 34.15 format by discarding low 15 bits. Lastly, the accumulator is saturated to yield a result in 1.15 format.

#### Parameters

- [in] S: points to an instance of the Q15 FIR interpolator structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

```
void riscv_fir_interpolate_q31 (const riscv_fir_interpolate_instance_q31 *S, const q31_t  
                               *pSrc, q31_t *pDst, uint32_t blockSize)
```

Processing function for the Q31 FIR interpolator.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. Thus, if the accumulator result overflows it wraps around rather than clip. In order to avoid overflows completely the input signal must be scaled down by  $1/(\text{numTaps}/L)$ . since  $\text{numTaps}/L$  additions occur per output sample. After all multiply-accumulates are performed, the 2.62 accumulator is truncated to 1.32 format and then saturated to 1.31 format.

#### Parameters

- [in] S: points to an instance of the Q31 FIR interpolator structure
- [in] pSrc: points to the block of input data
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process

*group* **groupFilters**

### 3.3.9 Interpolation Functions

#### Bilinear Interpolation

```
float32_t riscv_bilinear_interp_f32 (const riscv_bilinear_interp_instance_f32 *S, float32_t X,
                                     float32_t Y)
q31_t riscv_bilinear_interp_q31 (riscv_bilinear_interp_instance_q31 *S, q31_t X, q31_t Y)
q15_t riscv_bilinear_interp_q15 (riscv_bilinear_interp_instance_q15 *S, q31_t X, q31_t Y)
q7_t riscv_bilinear_interp_q7 (riscv_bilinear_interp_instance_q7 *S, q31_t X, q31_t Y)
float16_t riscv_bilinear_interp_f16 (const riscv_bilinear_interp_instance_f16 *S, float16_t X,
                                     float16_t Y)
```

#### *group* **BilinearInterpolate**

Bilinear interpolation is an extension of linear interpolation applied to a two dimensional grid. The underlying function  $f(x, y)$  is sampled on a regular grid and the interpolation process determines values between the grid points. Bilinear interpolation is equivalent to two step linear interpolation, first in the x-dimension and then in the y-dimension. Bilinear interpolation is often used in image processing to rescale images. The NMSIS DSP library provides bilinear interpolation functions for Q7, Q15, Q31, and floating-point data types.

**Algorithm** Bilinear interpolation is an extension of linear interpolation applied to a two dimensional grid. The underlying function  $f(x, y)$  is sampled on a regular grid and the interpolation process determines values between the grid points. Bilinear interpolation is equivalent to two step linear interpolation, first in the x-dimension and then in the y-dimension. Bilinear interpolation is often used in image processing to rescale images. The NMSIS DSP library provides bilinear interpolation functions for Q7, Q15, Q31, and floating-point data types.

The instance structure used by the bilinear interpolation functions describes a two dimensional data table. For floating-point, the instance structure is defined as:

where `numRows` specifies the number of rows in the table; `numCols` specifies the number of columns in the table; and `pData` points to an array of size `numRows*numCols` values. The data table `pTable` is organized in row order and the supplied data values fall on integer indexes. That is, table element (x,y) is located at `pTable[x + y*numCols]` where x and y are integers.

Let (x, y) specify the desired interpolation point. Then define:

The interpolated output point is computed as: Note that the coordinates (x, y) contain integer and fractional components. The integer components specify which portion of the table to use while the fractional components control the interpolation processor.

if (x,y) are outside of the table boundary, Bilinear interpolation returns zero output.

#### **Algorithm** end of LinearInterpolate group

The instance structure used by the bilinear interpolation functions describes a two dimensional data table. For floating-point, the instance structure is defined as:

where `numRows` specifies the number of rows in the table; `numCols` specifies the number of columns in the table; and `pData` points to an array of size `numRows*numCols` values. The data table `pTable` is organized in row order and the supplied data values fall on integer indexes. That is, table element (x,y) is located at `pTable[x + y*numCols]` where x and y are integers.

Let (x, y) specify the desired interpolation point. Then define:

The interpolated output point is computed as: Note that the coordinates (x, y) contain integer and fractional components. The integer components specify which portion of the table to use while the fractional components control the interpolation processor.

if (x,y) are outside of the table boundary, Bilinear interpolation returns zero output.

## Functions

`float32_t riscv_bilinear_interp_f32 (const riscv_bilinear_interp_instance_f32 *S, float32_t X, float32_t Y)`

Floating-point bilinear interpolation.

**Return** out interpolated value.

### Parameters

- [inout] S: points to an instance of the interpolation structure.
- [in] X: interpolation coordinate.
- [in] Y: interpolation coordinate.

`q31_t riscv_bilinear_interp_q31 (riscv_bilinear_interp_instance_q31 *S, q31_t X, q31_t Y)`  
Q31 bilinear interpolation.

**Return** out interpolated value.

### Parameters

- [inout] S: points to an instance of the interpolation structure.
- [in] X: interpolation coordinate in 12.20 format.
- [in] Y: interpolation coordinate in 12.20 format.

`q15_t riscv_bilinear_interp_q15 (riscv_bilinear_interp_instance_q15 *S, q31_t X, q31_t Y)`  
Q15 bilinear interpolation.

**Return** out interpolated value.

### Parameters

- [inout] S: points to an instance of the interpolation structure.
- [in] X: interpolation coordinate in 12.20 format.
- [in] Y: interpolation coordinate in 12.20 format.

`q7_t riscv_bilinear_interp_q7 (riscv_bilinear_interp_instance_q7 *S, q31_t X, q31_t Y)`  
Q7 bilinear interpolation.

**Return** out interpolated value.

### Parameters

- [inout] S: points to an instance of the interpolation structure.
- [in] X: interpolation coordinate in 12.20 format.
- [in] Y: interpolation coordinate in 12.20 format.

float16\_t **riscv\_bilinear\_interp\_f16** (const riscv\_bilinear\_interp\_instance\_f16 \*S, float16\_t X, float16\_t Y)

Floating-point bilinear interpolation.

**Return** out interpolated value.

**Parameters**

- [inout] S: points to an instance of the interpolation structure.
- [in] X: interpolation coordinate.
- [in] Y: interpolation coordinate.

## Linear Interpolation

float32\_t **riscv\_linear\_interp\_f32** (riscv\_linear\_interp\_instance\_f32 \*S, float32\_t x)

q31\_t **riscv\_linear\_interp\_q31** (q31\_t \*pYData, q31\_t x, uint32\_t nValues)

q15\_t **riscv\_linear\_interp\_q15** (q15\_t \*pYData, q31\_t x, uint32\_t nValues)

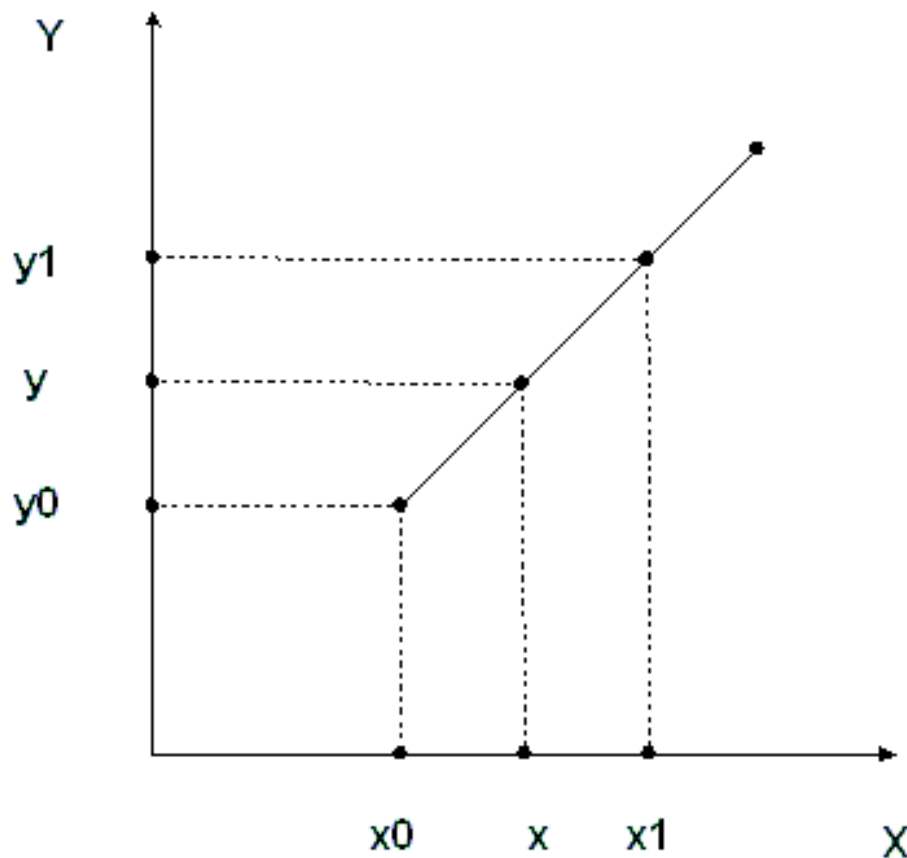
q7\_t **riscv\_linear\_interp\_q7** (q7\_t \*pYData, q31\_t x, uint32\_t nValues)

float16\_t **riscv\_linear\_interp\_f16** (riscv\_linear\_interp\_instance\_f16 \*S, float16\_t x)

*group* **LinearInterpolate**

Linear interpolation is a method of curve fitting using linear polynomials. Linear interpolation works by effectively drawing a straight line between two neighboring samples and returning the appropriate point along that line

end of SplineInterpolate group



A Linear Interpolate function calculates an output value( $y$ ), for the input( $x$ ) using linear interpolation of the input values  $x_0$ ,  $x_1$  (nearest input values) and the output values  $y_0$  and  $y_1$  (nearest output values)

**Algorithm:**

This set of functions implements Linear interpolation process for Q7, Q15, Q31, and floating-point data types. The functions operate on a single sample of data and each call to the function returns a single processed value.  $S$  points to an instance of the Linear Interpolate function data structure.  $x$  is the input sample value. The functions returns the output value.

if  $x$  is outside of the table boundary, Linear interpolation returns first value of the table if  $x$  is below input range and returns last value of table if  $x$  is above range.

**Functions**

`float32_t riscv_linear_interp_f32 (riscv_linear_interp_instance_f32 *S, float32_t x)`  
Process function for the floating-point Linear Interpolation Function.

**Return**  $y$  processed output sample.

**Parameters**

- [inout]  $S$ : is an instance of the floating-point Linear Interpolation structure
- [in]  $x$ : input sample to process

**q31\_t riscv\_linear\_interp\_q31** (q31\_t \*pYData, q31\_t x, uint32\_t nValues)

Process function for the Q31 Linear Interpolation Function.

**Return** y processed output sample.

Input sample x is in 12.20 format which contains 12 bits for table index and 20 bits for fractional part.

This function can support maximum of table size  $2^{12}$ .

**Parameters**

- [in] pYData: pointer to Q31 Linear Interpolation table
- [in] x: input sample to process
- [in] nValues: number of table values

**q15\_t riscv\_linear\_interp\_q15** (q15\_t \*pYData, q31\_t x, uint32\_t nValues)

Process function for the Q15 Linear Interpolation Function.

**Return** y processed output sample.

Input sample x is in 12.20 format which contains 12 bits for table index and 20 bits for fractional part.

This function can support maximum of table size  $2^{12}$ .

**Parameters**

- [in] pYData: pointer to Q15 Linear Interpolation table
- [in] x: input sample to process
- [in] nValues: number of table values

**q7\_t riscv\_linear\_interp\_q7** (q7\_t \*pYData, q31\_t x, uint32\_t nValues)

Process function for the Q7 Linear Interpolation Function.

**Return** y processed output sample.

Input sample x is in 12.20 format which contains 12 bits for table index and 20 bits for fractional part.

This function can support maximum of table size  $2^{12}$ .

**Parameters**

- [in] pYData: pointer to Q7 Linear Interpolation table
- [in] x: input sample to process
- [in] nValues: number of table values

**float16\_t riscv\_linear\_interp\_f16** (riscv\_linear\_interp\_instance\_f16 \*S, float16\_t x)

Process function for the floating-point Linear Interpolation Function.

**Return** y processed output sample.

**Parameters**

- [inout] S: is an instance of the floating-point Linear Interpolation structure
- [in] x: input sample to process

## Cubic Spline Interpolation

```
void riscv_spline_f32 (riscv_spline_instance_f32 *S, const float32_t *xq, float32_t *pDst, uint32_t
                      blockSize)
```

```
void riscv_spline_init_f32 (riscv_spline_instance_f32 *S, riscv_spline_type type, const float32_t
                           *x, const float32_t *y, uint32_t n, float32_t *coeffs, float32_t *temp-
                           Buffer)
```

### group **SplineInterpolate**

Spline interpolation is a method of interpolation where the interpolant is a piecewise-defined polynomial called “spline”.

Given a function  $f$  defined on the interval  $[a, b]$ , a set of  $n$  nodes  $x(i)$  where  $a = x(1) < x(2) < \dots < x(n) = b$  and a set of  $n$  values  $y(i) = f(x(i))$ , a cubic spline interpolant  $S(x)$  is defined as:

#### Introduction

where

Having defined  $h(i) = x(i+1) - x(i)$

#### Algorithm

It is possible to write the previous conditions in matrix form ( $Ax=B$ ). In order to solve the system two boundary conditions are needed.

- Natural spline:  $S''(x_1) = 2 * c(1) = 0$  ;  $S''(x_n) = 2 * c(n) = 0$  In matrix form:
- Parabolic runout spline:  $S''(x_1) = 2 * c(1) = S''(x_2) = 2 * c(2)$  ;  $S''(x_{n-1}) = 2 * c(n-1) = S''(x_n) = 2 * c(n)$  In matrix form:

$A$  is a tridiagonal matrix (a band matrix of bandwidth 3) of size  $N = n+1$ . The factorization algorithms ( $A=LU$ ) can be simplified considerably because a large number of zeros appear in regular patterns. The Crout method has been used: 1) Solve  $LZ=B$

2) Solve  $UX=Z$

$c(i)$  for  $i=1, \dots, n-1$  are needed to compute the  $n-1$  polynomials.  $b(i)$  and  $d(i)$  are computed as:

- $b(i) = [y(i+1) - y(i)] / h(i) - h(i) * [c(i+1) + 2 * c(i)] / 3$
- $d(i) = [c(i+1) - c(i)] / [3 * h(i)]$  Moreover,  $a(i) = y(i)$ .

It is possible to compute the interpolated vector for  $x$  values outside the input range ( $x_q < x(1)$ ;  $x_q > x(n)$ ). The coefficients used to compute the  $y$  values for  $x_q < x(1)$  are going to be the ones used for the first interval, while for  $x_q > x(n)$  the coefficients used for the last interval.

#### Behaviour outside the given intervals

The initialization function takes as input two arrays that the user has to allocate: `coeffs` will contain the  $b$ ,  $c$ , and  $d$  coefficients for the  $(n-1)$  intervals ( $n$  is the number of known points), hence its size must be  $3 * (n-1)$ ; `tempBuffer` is temporally used for internal computations and its size is  $n+n-1$ .

#### Initialization function

The  $x$  input array must be strictly sorted in ascending order and it must not contain twice the same value ( $x(i) < x(i+1)$ ).



## Functions

void **riscv\_spline\_f32** (riscv\_spline\_instance\_f32 \*S, **const** float32\_t \*xq, float32\_t \*pDst, uint32\_t blockSize)

Processing function for the floating-point cubic spline interpolation.

### Parameters

- [in] S: points to an instance of the floating-point spline structure.
- [in] xq: points to the x values of the interpolated data points.
- [out] pDst: points to the block of output data.
- [in] blockSize: number of samples of output data.
- [in] S: points to an instance of the floating-point spline structure.
- [in] xq: points to the x values of the interpolated data points.
- [out] pDst: points to the block of output data.
- [in] blockSize: number of samples of output data.

void **riscv\_spline\_init\_f32** (riscv\_spline\_instance\_f32 \*S, riscv\_spline\_type type, **const** float32\_t \*x, **const** float32\_t \*y, uint32\_t n, float32\_t \*coeffs, float32\_t \*tempBuffer)

Initialization function for the floating-point cubic spline interpolation.

### Parameters

- [inout] S: points to an instance of the floating-point spline structure.
- [in] type: type of cubic spline interpolation (boundary conditions)
- [in] x: points to the x values of the known data points.
- [in] y: points to the y values of the known data points.
- [in] n: number of known data points.
- [in] coeffs: coefficients array for b, c, and d
- [in] tempBuffer: buffer array for internal computations

### group **groupInterpolation**

These functions perform 1- and 2-dimensional interpolation of data. Linear interpolation is used for 1-dimensional data and bilinear interpolation is used for 2-dimensional data.

## 3.3.10 Matrix Functions

### Matrix Addition

riscv\_status **riscv\_mat\_add\_f16** (**const** riscv\_matrix\_instance\_f16 \*pSrcA, **const** riscv\_matrix\_instance\_f16 \*pSrcB, riscv\_matrix\_instance\_f16 \*pDst)

riscv\_status **riscv\_mat\_add\_f32** (**const** riscv\_matrix\_instance\_f32 \*pSrcA, **const** riscv\_matrix\_instance\_f32 \*pSrcB, riscv\_matrix\_instance\_f32 \*pDst)

riscv\_status **riscv\_mat\_add\_q15** (**const** riscv\_matrix\_instance\_q15 \*pSrcA, **const** riscv\_matrix\_instance\_q15 \*pSrcB, riscv\_matrix\_instance\_q15 \*pDst)

riscv\_status **riscv\_mat\_add\_q31** (**const** riscv\_matrix\_instance\_q31 \*pSrcA, **const** riscv\_matrix\_instance\_q31 \*pSrcB, riscv\_matrix\_instance\_q31 \*pDst)

**group MatrixAdd**

Adds two matrices. The functions check to make sure that pSrcA, pSrcB, and pDst have the same number of rows and columns.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} \end{bmatrix}$$

**Functions**

riscv\_status **riscv\_mat\_add\_f16** (const riscv\_matrix\_instance\_f16 \*pSrcA, const riscv\_matrix\_instance\_f16 \*pSrcB, riscv\_matrix\_instance\_f16 \*pDst)

Floating-point matrix addition.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrcA: points to first input matrix structure
- [in] pSrcB: points to second input matrix structure
- [out] pDst: points to output matrix structure

riscv\_status **riscv\_mat\_add\_f32** (const riscv\_matrix\_instance\_f32 \*pSrcA, const riscv\_matrix\_instance\_f32 \*pSrcB, riscv\_matrix\_instance\_f32 \*pDst)

Floating-point matrix addition.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrcA: points to first input matrix structure
- [in] pSrcB: points to second input matrix structure
- [out] pDst: points to output matrix structure

riscv\_status **riscv\_mat\_add\_q15** (const riscv\_matrix\_instance\_q15 \*pSrcA, const riscv\_matrix\_instance\_q15 \*pSrcB, riscv\_matrix\_instance\_q15 \*pDst)

Q15 matrix addition.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful

- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

#### Parameters

- [in] pSrcA: points to first input matrix structure
- [in] pSrcB: points to second input matrix structure
- [out] pDst: points to output matrix structure

```
riscv_status riscv_mat_add_q31 (const   riscv_matrix_instance_q31   *pSrcA,   const
                                riscv_matrix_instance_q31   *pSrcB,   riscv_matrix_instance_q31
                                *pDst)
```

Q31 matrix addition.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range [0x80000000 0x7FFFFFFF] are saturated.

#### Parameters

- [in] pSrcA: points to first input matrix structure
- [in] pSrcB: points to second input matrix structure
- [out] pDst: points to output matrix structure

### Cholesky and LDLT decompositions

```
riscv_status riscv_mat_cholesky_f16 (const   riscv_matrix_instance_f16   *pSrc,
                                      riscv_matrix_instance_f16 *pDst)
```

```
riscv_status riscv_mat_cholesky_f32 (const   riscv_matrix_instance_f32   *pSrc,
                                      riscv_matrix_instance_f32 *pDst)
```

```
riscv_status riscv_mat_cholesky_f64 (const   riscv_matrix_instance_f64   *pSrc,
                                      riscv_matrix_instance_f64 *pDst)
```

```
riscv_status riscv_mat_ldlt_f32 (const riscv_matrix_instance_f32 *pSrc, riscv_matrix_instance_f32
                                *pl, riscv_matrix_instance_f32 *pd, uint16_t *pp)
```

```
riscv_status riscv_mat_ldlt_f64 (const riscv_matrix_instance_f64 *pSrc, riscv_matrix_instance_f64
                                *pl, riscv_matrix_instance_f64 *pd, uint16_t *pp)
```

*group* **MatrixChol**

Computes the Cholesky or LDL<sup>t</sup> decomposition of a matrix.

If the input matrix does not have a decomposition, then the algorithm terminates and returns error status RISC\_V\_MATH\_DECOMPOSITION\_FAILURE.

#### Functions

```
riscv_status riscv_mat_cholesky_f16 (const   riscv_matrix_instance_f16   *pSrc,
                                      riscv_matrix_instance_f16 *pDst)
```

Floating-point Cholesky decomposition of positive-definite matrix.

Floating-point Cholesky decomposition of Symmetric Positive Definite Matrix.

**Return** The function returns `RISCV_MATH_SIZE_MISMATCH`, if the dimensions do not match.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_SIZE_MISMATCH` : Matrix size check failed
- `RISCV_MATH_DECOMPOSITION_FAILURE` : Input matrix cannot be decomposed

If the matrix is ill conditioned or only semi-definite, then it is better using the LDL<sup>t</sup> decomposition. The decomposition of A is returning a lower triangular matrix U such that  $A = U U^t$

**Parameters**

- [in] `pSrc`: points to the instance of the input floating-point matrix structure.
- [out] `pDst`: points to the instance of the output floating-point matrix structure.

riscv\_status **riscv\_mat\_cholesky\_f32** (**const** riscv\_matrix\_instance\_f32 \**pSrc*,  
riscv\_matrix\_instance\_f32 \**pDst*)

Floating-point Cholesky decomposition of positive-definite matrix.

Floating-point Cholesky decomposition of Symmetric Positive Definite Matrix.

**Return** The function returns `RISCV_MATH_SIZE_MISMATCH`, if the dimensions do not match.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_SIZE_MISMATCH` : Matrix size check failed
- `RISCV_MATH_DECOMPOSITION_FAILURE` : Input matrix cannot be decomposed

If the matrix is ill conditioned or only semi-definite, then it is better using the LDL<sup>t</sup> decomposition. The decomposition of A is returning a lower triangular matrix U such that  $A = U U^t$

**Parameters**

- [in] `pSrc`: points to the instance of the input floating-point matrix structure.
- [out] `pDst`: points to the instance of the output floating-point matrix structure.

riscv\_status **riscv\_mat\_cholesky\_f64** (**const** riscv\_matrix\_instance\_f64 \**pSrc*,  
riscv\_matrix\_instance\_f64 \**pDst*)

Floating-point Cholesky decomposition of positive-definite matrix.

Floating-point Cholesky decomposition of Symmetric Positive Definite Matrix.

**Return** The function returns `RISCV_MATH_SIZE_MISMATCH`, if the dimensions do not match.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_SIZE_MISMATCH` : Matrix size check failed
- `RISCV_MATH_DECOMPOSITION_FAILURE` : Input matrix cannot be decomposed

If the matrix is ill conditioned or only semi-definite, then it is better using the LDL<sup>t</sup> decomposition. The decomposition of A is returning a lower triangular matrix U such that  $A = U U^t$

**Parameters**

- [in] pSrc: points to the instance of the input floating-point matrix structure.
- [out] pDst: points to the instance of the output floating-point matrix structure.

```
riscv_status riscv_mat_ldlt_f32 (const riscv_matrix_instance_f32 *pSrc,
                                riscv_matrix_instance_f32 *pl, riscv_matrix_instance_f32
                                *pd, uint16_t *pp)
```

Floating-point LDL<sup>t</sup> decomposition of positive semi-definite matrix.

Floating-point LDL decomposition of Symmetric Positive Semi-Definite Matrix.

**Return** The function returns RISCV\_MATH\_SIZE\_MISMATCH, if the dimensions do not match.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_SIZE\_MISMATCH : Matrix size check failed
- RISCV\_MATH\_DECOMPOSITION\_FAILURE : Input matrix cannot be decomposed

Computes the LDL<sup>t</sup> decomposition of a matrix A such that  $P A P^t = L D L^t$ .

#### Parameters

- [in] pSrc: points to the instance of the input floating-point matrix structure.
- [out] pl: points to the instance of the output floating-point triangular matrix structure.
- [out] pd: points to the instance of the output floating-point diagonal matrix structure.
- [out] pp: points to the instance of the output floating-point permutation vector.

```
riscv_status riscv_mat_ldlt_f64 (const riscv_matrix_instance_f64 *pSrc,
                                riscv_matrix_instance_f64 *pl, riscv_matrix_instance_f64
                                *pd, uint16_t *pp)
```

Floating-point LDL<sup>t</sup> decomposition of positive semi-definite matrix.

Floating-point LDL decomposition of Symmetric Positive Semi-Definite Matrix.

**Return** The function returns RISCV\_MATH\_SIZE\_MISMATCH, if the dimensions do not match.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_SIZE\_MISMATCH : Matrix size check failed
- RISCV\_MATH\_DECOMPOSITION\_FAILURE : Input matrix cannot be decomposed

Computes the LDL<sup>t</sup> decomposition of a matrix A such that  $P A P^t = L D L^t$ .

#### Parameters

- [in] pSrc: points to the instance of the input floating-point matrix structure.
- [out] pl: points to the instance of the output floating-point triangular matrix structure.
- [out] pd: points to the instance of the output floating-point diagonal matrix structure.
- [out] pp: points to the instance of the output floating-point permutation vector.

## Complex Matrix Multiplication

```
riscv_status riscv_mat_cmplx_mult_f16 (const      riscv_matrix_instance_f16    *pSrcA,  
                                         const      riscv_matrix_instance_f16    *pSrcB,  
                                         riscv_matrix_instance_f16 *pDst)  
  
riscv_status riscv_mat_cmplx_mult_f32 (const      riscv_matrix_instance_f32    *pSrcA,  
                                         const      riscv_matrix_instance_f32    *pSrcB,  
                                         riscv_matrix_instance_f32 *pDst)  
  
riscv_status riscv_mat_cmplx_mult_q15 (const      riscv_matrix_instance_q15    *pSrcA,  
                                         const      riscv_matrix_instance_q15    *pSrcB,  
                                         riscv_matrix_instance_q15 *pDst, q15_t *pScratch)  
  
riscv_status riscv_mat_cmplx_mult_q31 (const      riscv_matrix_instance_q31    *pSrcA,  
                                         const      riscv_matrix_instance_q31    *pSrcB,  
                                         riscv_matrix_instance_q31 *pDst)
```

### group **CmplxMatrixMult**

Complex Matrix multiplication is only defined if the number of columns of the first matrix equals the number of rows of the second matrix. Multiplying an  $M \times N$  matrix with an  $N \times P$  matrix results in an  $M \times P$  matrix.

When matrix size checking is enabled, the functions check:

- that the inner dimensions of pSrcA and pSrcB are equal;
- that the size of the output matrix equals the outer dimensions of pSrcA and pSrcB.

## Functions

```
riscv_status riscv_mat_cmplx_mult_f16 (const      riscv_matrix_instance_f16    *pSrcA,  
                                         const      riscv_matrix_instance_f16    *pSrcB,  
                                         riscv_matrix_instance_f16 *pDst)
```

Floating-point Complex matrix multiplication.

Floating-point, complex, matrix multiplication.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrcA: points to first input complex matrix structure
- [in] pSrcB: points to second input complex matrix structure
- [out] pDst: points to output complex matrix structure

```
riscv_status riscv_mat_cmplx_mult_f32 (const      riscv_matrix_instance_f32    *pSrcA,  
                                         const      riscv_matrix_instance_f32    *pSrcB,  
                                         riscv_matrix_instance_f32 *pDst)
```

Floating-point Complex matrix multiplication.

Floating-point, complex, matrix multiplication.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrcA: points to first input complex matrix structure
- [in] pSrcB: points to second input complex matrix structure
- [out] pDst: points to output complex matrix structure

```
riscv_status riscv_mat_cmplx_mult_q15 (const   riscv_matrix_instance_q15   *pSrcA,
                                         const   riscv_matrix_instance_q15   *pSrcB,
                                         riscv_matrix_instance_q15 *pDst, q15_t *pScratch)
```

Q15 Complex matrix multiplication.

Q15, complex, matrix multiplication.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Conditions for optimum performance** Input, output and state buffers should be aligned by 32-bit

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The inputs to the multiplications are in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. This approach provides 33 guard bits and there is no risk of overflow. The 34.30 result is then truncated to 34.15 format by discarding the low 15 bits and then saturated to 1.15 format.

**Parameters**

- [in] pSrcA: points to first input complex matrix structure
- [in] pSrcB: points to second input complex matrix structure
- [out] pDst: points to output complex matrix structure
- [in] pScratch: points to an array for storing intermediate results

```
riscv_status riscv_mat_cmplx_mult_q31 (const   riscv_matrix_instance_q31   *pSrcA,
                                         const   riscv_matrix_instance_q31   *pSrcB,
                                         riscv_matrix_instance_q31 *pDst)
```

Q31 Complex matrix multiplication.

Q31, complex, matrix multiplication.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. There is no saturation on intermediate additions. Thus, if the accumulator overflows it wraps around and distorts the result. The input signals should be scaled down to avoid intermediate overflows. The input is thus scaled down by  $\log_2(\text{numColsA})$  bits to avoid overflows, as a total of numColsA additions are performed internally. The 2.62 accumulator is right shifted by 31 bits and saturated to 1.31 format to yield the final result.

**Parameters**

- [in] pSrcA: points to first input complex matrix structure

- [in] pSrcB: points to second input complex matrix structure
- [out] pDst: points to output complex matrix structure

### Complex Matrix Transpose

```
riscv_status riscv_mat_cmplx_trans_f16 (const      riscv_matrix_instance_f16      *pSrc,
                                           riscv_matrix_instance_f16 *pDst)
riscv_status riscv_mat_cmplx_trans_f32 (const      riscv_matrix_instance_f32      *pSrc,
                                           riscv_matrix_instance_f32 *pDst)
riscv_status riscv_mat_cmplx_trans_q15 (const      riscv_matrix_instance_q15      *pSrc,
                                           riscv_matrix_instance_q15 *pDst)
riscv_status riscv_mat_cmplx_trans_q31 (const      riscv_matrix_instance_q31      *pSrc,
                                           riscv_matrix_instance_q31 *pDst)
```

**group MatrixComplexTrans**  
 Transposes a complex matrix.

Transposing an  $M \times N$  matrix flips it around the center diagonal and results in an  $N \times M$  matrix.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

### Functions

```
riscv_status riscv_mat_cmplx_trans_f16 (const      riscv_matrix_instance_f16      *pSrc,
                                           riscv_matrix_instance_f16 *pDst)
```

Floating-point matrix transpose.

Floating-point complex matrix transpose.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

```
riscv_status riscv_mat_cmplx_trans_f32 (const      riscv_matrix_instance_f32      *pSrc,
                                           riscv_matrix_instance_f32 *pDst)
```

Floating-point matrix transpose.



Floating-point complex matrix transpose.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

riscv\_status **riscv\_mat\_cmplx\_trans\_q15** (const riscv\_matrix\_instance\_q15 \*pSrc,  
riscv\_matrix\_instance\_q15 \*pDst)

Q15 complex matrix transpose.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

riscv\_status **riscv\_mat\_cmplx\_trans\_q31** (const riscv\_matrix\_instance\_q31 \*pSrc,  
riscv\_matrix\_instance\_q31 \*pDst)

Q31 complex matrix transpose.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

## Matrix Initialization

void **riscv\_mat\_init\_f16** (riscv\_matrix\_instance\_f16 \*S, uint16\_t nRows, uint16\_t nColumns, float16\_t \*pData)

void **riscv\_mat\_init\_f32** (riscv\_matrix\_instance\_f32 \*S, uint16\_t nRows, uint16\_t nColumns, float32\_t \*pData)

void **riscv\_mat\_init\_f64** (riscv\_matrix\_instance\_f64 \*S, uint16\_t nRows, uint16\_t nColumns, float64\_t \*pData)

void **riscv\_mat\_init\_q15** (riscv\_matrix\_instance\_q15 \*S, uint16\_t nRows, uint16\_t nColumns, q15\_t \*pData)

void **riscv\_mat\_init\_q31** (riscv\_matrix\_instance\_q31 \*S, uint16\_t nRows, uint16\_t nColumns, q31\_t \*pData)

**group MatrixInit**

Initializes the underlying matrix data structure. The functions set the `numRows`, `numCols`, and `pData` fields of the matrix data structure.

**Functions**

void **riscv\_mat\_init\_f16** (riscv\_matrix\_instance\_f16 \*S, uint16\_t nRows, uint16\_t nColumns, float16\_t \*pData)  
Floating-point matrix initialization.

**Return** none

**Parameters**

- [inout] S: points to an instance of the floating-point matrix structure
- [in] nRows: number of rows in the matrix
- [in] nColumns: number of columns in the matrix
- [in] pData: points to the matrix data array

void **riscv\_mat\_init\_f32** (riscv\_matrix\_instance\_f32 \*S, uint16\_t nRows, uint16\_t nColumns, float32\_t \*pData)  
Floating-point matrix initialization.

**Return** none

**Parameters**

- [inout] S: points to an instance of the floating-point matrix structure
- [in] nRows: number of rows in the matrix
- [in] nColumns: number of columns in the matrix
- [in] pData: points to the matrix data array

void **riscv\_mat\_init\_f64** (riscv\_matrix\_instance\_f64 \*S, uint16\_t nRows, uint16\_t nColumns, float32\_t \*pData)  
Floating-point matrix initialization.

**Return** none

**Parameters**

- [inout] S: points to an instance of the floating-point matrix structure
- [in] nRows: number of rows in the matrix
- [in] nColumns: number of columns in the matrix
- [in] pData: points to the matrix data array

void **riscv\_mat\_init\_q15** (riscv\_matrix\_instance\_q15 \*S, uint16\_t nRows, uint16\_t nColumns, q15\_t \*pData)  
Q15 matrix initialization.

**Return** none

**Parameters**

- [inout] *S*: points to an instance of the floating-point matrix structure
- [in] *nRows*: number of rows in the matrix
- [in] *nColumns*: number of columns in the matrix
- [in] *pData*: points to the matrix data array

void **riscv\_mat\_init\_q31** (riscv\_matrix\_instance\_q31 \**S*, uint16\_t *nRows*, uint16\_t *nColumns*,  
q31\_t \**pData*)  
Q31 matrix initialization.

**Return** none

**Parameters**

- [inout] *S*: points to an instance of the Q31 matrix structure
- [in] *nRows*: number of rows in the matrix
- [in] *nColumns*: number of columns in the matrix
- [in] *pData*: points to the matrix data array

## Matrix Inverse

```
riscv_status riscv_mat_inverse_f16 (const      riscv_matrix_instance_f16      *pSrc,
                                     riscv_matrix_instance_f16 *pDst)

riscv_status riscv_mat_inverse_f32 (const      riscv_matrix_instance_f32      *pSrc,
                                     riscv_matrix_instance_f32 *pDst)

riscv_status riscv_mat_inverse_f64 (const      riscv_matrix_instance_f64      *pSrc,
                                     riscv_matrix_instance_f64 *pDst)

riscv_status riscv_mat_solve_lower_triangular_f16 (const  riscv_matrix_instance_f16 *lt,
                                                     const  riscv_matrix_instance_f16 *a,
                                                     riscv_matrix_instance_f16 *dst)

riscv_status riscv_mat_solve_lower_triangular_f32 (const  riscv_matrix_instance_f32 *lt,
                                                     const  riscv_matrix_instance_f32 *a,
                                                     riscv_matrix_instance_f32 *dst)

riscv_status riscv_mat_solve_lower_triangular_f64 (const  riscv_matrix_instance_f64 *lt,
                                                     const  riscv_matrix_instance_f64 *a,
                                                     riscv_matrix_instance_f64 *dst)

riscv_status riscv_mat_solve_upper_triangular_f16 (const  riscv_matrix_instance_f16 *ut,
                                                     const  riscv_matrix_instance_f16 *a,
                                                     riscv_matrix_instance_f16 *dst)

riscv_status riscv_mat_solve_upper_triangular_f32 (const  riscv_matrix_instance_f32 *ut,
                                                     const  riscv_matrix_instance_f32 *a,
                                                     riscv_matrix_instance_f32 *dst)

riscv_status riscv_mat_solve_upper_triangular_f64 (const  riscv_matrix_instance_f64 *ut,
                                                     const  riscv_matrix_instance_f64 *a,
                                                     riscv_matrix_instance_f64 *dst)
```

### group **MatrixInv**

Computes the inverse of a matrix.

The inverse is defined only if the input matrix is square and non-singular (the determinant is non-zero). The function checks that the input and output matrices are square and of the same size.

Matrix inversion is numerically sensitive and the NMSIS DSP library only supports matrix inversion of floating-point matrices.

**Algorithm** The Gauss-Jordan method is used to find the inverse. The algorithm performs a sequence of elementary row-operations until it reduces the input matrix to an identity matrix. Applying the same sequence of elementary row-operations to an identity matrix yields the inverse matrix. If the input matrix is singular, then the algorithm terminates and returns error status `RISCV_MATH_SINGULAR`.

$$\begin{bmatrix} \overline{a_{11}} & \overline{a_{12}} & \overline{a_{13}} & | & 1 & 0 & 0 \\ \overline{a_{21}} & \overline{a_{22}} & \overline{a_{23}} & | & 0 & 1 & 0 \\ \overline{a_{31}} & \overline{a_{32}} & \overline{a_{33}} & | & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & | & X_{11} & X_{21} & X_{31} \\ 0 & 1 & 0 & | & X_{12} & X_{22} & X_{32} \\ 0 & 0 & 1 & | & X_{13} & X_{23} & X_{33} \end{bmatrix}$$

A is a 3 x 3 matrix and its inverse is X

## Functions

riscv\_status **riscv\_mat\_inverse\_f16** (const riscv\_matrix\_instance\_f16 \*pSrc,  
riscv\_matrix\_instance\_f16 \*pDst)

Floating-point matrix inverse.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_SIZE_MISMATCH` : Matrix size check failed
- `RISCV_MATH_SINGULAR` : Input matrix is found to be singular (non-invertible)

### Parameters

- [in] pSrc: points to input matrix structure. The source matrix is modified by the function.
- [out] pDst: points to output matrix structure

riscv\_status **riscv\_mat\_inverse\_f32** (const riscv\_matrix\_instance\_f32 \*pSrc,  
riscv\_matrix\_instance\_f32 \*pDst)

Floating-point matrix inverse.

**Return** execution status

- `RISCV_MATH_SUCCESS` : Operation successful
- `RISCV_MATH_SIZE_MISMATCH` : Matrix size check failed
- `RISCV_MATH_SINGULAR` : Input matrix is found to be singular (non-invertible)

### Parameters

- [in] pSrc: points to input matrix structure. The source matrix is modified by the function.
- [out] pDst: points to output matrix structure

riscv\_status **riscv\_mat\_inverse\_f64** (const riscv\_matrix\_instance\_f64 \*pSrc,  
riscv\_matrix\_instance\_f64 \*pDst)

Floating-point (64 bit) matrix inverse.

Floating-point matrix inverse.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed
- RISC\_V\_MATH\_SINGULAR : Input matrix is found to be singular (non-invertible)

**Parameters**

- [in] pSrc: points to input matrix structure. The source matrix is modified by the function.
- [out] pDst: points to output matrix structure

```
riscv_status riscv_mat_solve_lower_triangular_f16 (const riscv_matrix_instance_f16
                                                    *lt,                                const
                                                    riscv_matrix_instance_f16      *a,
                                                    riscv_matrix_instance_f16 *dst)
```

Solve  $LT \cdot X = A$  where  $LT$  is a lower triangular matrix.

**Return** The function returns RISC\_V\_MATH\_SINGULAR, if the system can't be solved.

**Parameters**

- [in] lt: The lower triangular matrix
- [in] a: The matrix a
- [out] dst: The solution X of  $LT \cdot X = A$

```
riscv_status riscv_mat_solve_lower_triangular_f32 (const riscv_matrix_instance_f32
                                                    *lt,                                const
                                                    riscv_matrix_instance_f32      *a,
                                                    riscv_matrix_instance_f32 *dst)
```

Solve  $LT \cdot X = A$  where  $LT$  is a lower triangular matrix.

**Return** The function returns RISC\_V\_MATH\_SINGULAR, if the system can't be solved.

**Parameters**

- [in] lt: The lower triangular matrix
- [in] a: The matrix a
- [out] dst: The solution X of  $LT \cdot X = A$

```
riscv_status riscv_mat_solve_lower_triangular_f64 (const riscv_matrix_instance_f64
                                                    *lt,                                const
                                                    riscv_matrix_instance_f64      *a,
                                                    riscv_matrix_instance_f64 *dst)
```

Solve  $LT \cdot X = A$  where  $LT$  is a lower triangular matrix.

**Return** The function returns RISC\_V\_MATH\_SINGULAR, if the system can't be solved.

**Parameters**

- [in] lt: The lower triangular matrix
- [in] a: The matrix a
- [out] dst: The solution X of  $LT \cdot X = A$

```
riscv_status riscv_mat_solve_upper_triangular_f16 (const riscv_matrix_instance_f16
                                                    *ut, const
                                                    riscv_matrix_instance_f16 *a,
                                                    riscv_matrix_instance_f16 *dst)
```

Solve  $UT \cdot X = A$  where  $UT$  is an upper triangular matrix.

**Return** The function returns `RISCV_MATH_SINGULAR`, if the system can't be solved.

**Parameters**

- [in] `ut`: The upper triangular matrix
- [in] `a`: The matrix  $a$
- [out] `dst`: The solution  $X$  of  $UT \cdot X = A$

```
riscv_status riscv_mat_solve_upper_triangular_f32 (const riscv_matrix_instance_f32
                                                    *ut, const
                                                    riscv_matrix_instance_f32 *a,
                                                    riscv_matrix_instance_f32 *dst)
```

Solve  $UT \cdot X = A$  where  $UT$  is an upper triangular matrix.

**Return** The function returns `RISCV_MATH_SINGULAR`, if the system can't be solved.

**Parameters**

- [in] `ut`: The upper triangular matrix
- [in] `a`: The matrix  $a$
- [out] `dst`: The solution  $X$  of  $UT \cdot X = A$

```
riscv_status riscv_mat_solve_upper_triangular_f64 (const riscv_matrix_instance_f64
                                                    *ut, const
                                                    riscv_matrix_instance_f64 *a,
                                                    riscv_matrix_instance_f64 *dst)
```

Solve  $UT \cdot X = A$  where  $UT$  is an upper triangular matrix.

**Return** The function returns `RISCV_MATH_SINGULAR`, if the system can't be solved.

**Parameters**

- [in] `ut`: The upper triangular matrix
- [in] `a`: The matrix  $a$
- [out] `dst`: The solution  $X$  of  $UT \cdot X = A$

## Matrix Multiplication

```
riscv_status riscv_mat_mult_f16 (const riscv_matrix_instance_f16 *pSrcA, const
                                riscv_matrix_instance_f16 *pSrcB, riscv_matrix_instance_f16
                                *pDst)
```

```
riscv_status riscv_mat_mult_f32 (const riscv_matrix_instance_f32 *pSrcA, const
                                riscv_matrix_instance_f32 *pSrcB, riscv_matrix_instance_f32
                                *pDst)
```

```
riscv_status riscv_mat_mult_f64 (const riscv_matrix_instance_f64 *pSrcA, const
                                riscv_matrix_instance_f64 *pSrcB, riscv_matrix_instance_f64
                                *pDst)
```

```

riscv_status riscv_mat_mult_fast_q15 (const      riscv_matrix_instance_q15      *pSrcA,
                                         const      riscv_matrix_instance_q15      *pSrcB,
                                         riscv_matrix_instance_q15 *pDst, q15_t *pState)

riscv_status riscv_mat_mult_fast_q31 (const      riscv_matrix_instance_q31      *pSrcA,
                                         const      riscv_matrix_instance_q31      *pSrcB,
                                         riscv_matrix_instance_q31 *pDst)

riscv_status riscv_mat_mult_q15 (const      riscv_matrix_instance_q15      *pSrcA, const
                                riscv_matrix_instance_q15 *pSrcB, riscv_matrix_instance_q15
                                *pDst, q15_t *pState)

riscv_status riscv_mat_mult_q31 (const      riscv_matrix_instance_q31      *pSrcA, const
                                riscv_matrix_instance_q31 *pSrcB, riscv_matrix_instance_q31
                                *pDst)

riscv_status riscv_mat_mult_q7 (const      riscv_matrix_instance_q7      *pSrcA, const
                                riscv_matrix_instance_q7 *pSrcB, riscv_matrix_instance_q7
                                *pDst, q7_t *pState)

```

#### group **MatrixMult**

Multiplies two matrices.

Matrix multiplication is only defined if the number of columns of the first matrix equals the number of rows of the second matrix. Multiplying an  $M \times N$  matrix with an  $N \times P$  matrix results in an  $M \times P$  matrix. When matrix size checking is enabled, the functions check: (1) that the inner dimensions of pSrcA and pSrcB are equal; and (2) that the size of the output matrix equals the outer dimensions of pSrcA and pSrcB.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11} \times b_{11} + a_{12} \times b_{21} + a_{13} \times b_{31} & a_{11} \times b_{12} + a_{12} \times b_{22} + a_{13} \times b_{32} & a_{11} \times b_{13} + a_{12} \times b_{23} + a_{13} \times b_{33} \\ a_{21} \times b_{11} + a_{22} \times b_{21} + a_{23} \times b_{31} & a_{21} \times b_{12} + a_{22} \times b_{22} + a_{23} \times b_{32} & a_{21} \times b_{13} + a_{22} \times b_{23} + a_{23} \times b_{33} \\ a_{31} \times b_{11} + a_{32} \times b_{21} + a_{33} \times b_{31} & a_{31} \times b_{12} + a_{32} \times b_{22} + a_{33} \times b_{32} & a_{31} \times b_{13} + a_{32} \times b_{23} + a_{33} \times b_{33} \end{bmatrix}$$

## Functions

```

riscv_status riscv_mat_mult_f16 (const      riscv_matrix_instance_f16      *pSrcA, const
                                riscv_matrix_instance_f16 *pSrcB, riscv_matrix_instance_f16
                                *pDst)

```

Floating-point matrix multiplication.

**Return** The function returns either RISCVC\_MATH\_SIZE\_MISMATCH or RISCVC\_MATH\_SUCCESS based on the outcome of size checking.

#### Parameters

- [in] \*pSrcA: points to the first input matrix structure
- [in] \*pSrcB: points to the second input matrix structure
- [out] \*pDst: points to output matrix structure

```

riscv_status riscv_mat_mult_f32 (const      riscv_matrix_instance_f32      *pSrcA, const
                                riscv_matrix_instance_f32 *pSrcB, riscv_matrix_instance_f32
                                *pDst)

```

Floating-point matrix multiplication.

**Return** The function returns either RISCVC\_MATH\_SIZE\_MISMATCH or RISCVC\_MATH\_SUCCESS based on the outcome of size checking.

**Parameters**

- [in] \*pSrcA: points to the first input matrix structure
- [in] \*pSrcB: points to the second input matrix structure
- [out] \*pDst: points to output matrix structure

```
riscv_status riscv_mat_mult_f64 (const   riscv_matrix_instance_f64   *pSrcA,   const  
                                riscv_matrix_instance_f64   *pSrcB, riscv_matrix_instance_f64  
                                *pDst)
```

Floating-point matrix multiplication.

**Return** The function returns either RISC\_V\_MATH\_SIZE\_MISMATCH or RISC\_V\_MATH\_SUCCESS based on the outcome of size checking.

**Parameters**

- [in] \*pSrcA: points to the first input matrix structure
- [in] \*pSrcB: points to the second input matrix structure
- [out] \*pDst: points to output matrix structure

```
riscv_status riscv_mat_mult_fast_q15 (const   riscv_matrix_instance_q15   *pSrcA,  
                                      const   riscv_matrix_instance_q15   *pSrcB,  
                                      riscv_matrix_instance_q15 *pDst, q15_t *pState)
```

Q15 matrix multiplication (fast variant).

Q15 matrix multiplication (fast variant) for RISC-V Core with DSP enabled.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The difference between the function riscv\_mat\_mult\_q15() and this fast variant is that the fast variant use a 32-bit rather than a 64-bit accumulator. The result of each 1.15 x 1.15 multiplication is truncated to 2.30 format. These intermediate results are accumulated in a 32-bit register in 2.30 format. Finally, the accumulator is saturated and converted to a 1.15 result.

The fast version has the same overflow behavior as the standard version but provides less precision since it discards the low 16 bits of each multiplication result. In order to avoid overflows completely the input signals must be scaled down. Scale down one of the input matrices by log2(numColsA) bits to avoid overflows, as a total of numColsA additions are computed internally for each output element.

**Remark** Refer to riscv\_mat\_mult\_q15() for a slower implementation of this function which uses 64-bit accumulation to provide higher precision.

**Parameters**

- [in] pSrcA: points to the first input matrix structure
- [in] pSrcB: points to the second input matrix structure
- [out] pDst: points to output matrix structure
- [in] pState: points to the array for storing intermediate results



```
riscv_status riscv_mat_mult_fast_q31 (const riscv_matrix_instance_q31 *pSrcA,
                                     const riscv_matrix_instance_q31 *pSrcB,
                                     riscv_matrix_instance_q31 *pDst)
```

Q31 matrix multiplication (fast variant).

Q31 matrix multiplication (fast variant) for RISC-V Core with DSP enabled.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The difference between the function `riscv_mat_mult_q31()` and this fast variant is that the fast variant use a 32-bit rather than a 64-bit accumulator. The result of each 1.31 x 1.31 multiplication is truncated to 2.30 format. These intermediate results are accumulated in a 32-bit register in 2.30 format. Finally, the accumulator is saturated and converted to a 1.31 result.

The fast version has the same overflow behavior as the standard version but provides less precision since it discards the low 32 bits of each multiplication result. In order to avoid overflows completely the input signals must be scaled down. Scale down one of the input matrices by  $\log_2(\text{numColsA})$  bits to avoid overflows, as a total of `numColsA` additions are computed internally for each output element.

**Remark** Refer to `riscv_mat_mult_q31()` for a slower implementation of this function which uses 64-bit accumulation to provide higher precision.

**Parameters**

- [in] `pSrcA`: points to the first input matrix structure
- [in] `pSrcB`: points to the second input matrix structure
- [out] `pDst`: points to output matrix structure

```
riscv_status riscv_mat_mult_q15 (const riscv_matrix_instance_q15 *pSrcA, const
                                riscv_matrix_instance_q15 *pSrcB, riscv_matrix_instance_q15
                                *pDst, q15_t *pState)
```

Q15 matrix multiplication.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The inputs to the multiplications are in 1.15 format and multiplications yield a 2.30 result. The 2.30 intermediate results are accumulated in a 64-bit accumulator in 34.30 format. This approach provides 33 guard bits and there is no risk of overflow. The 34.30 result is then truncated to 34.15 format by discarding the low 15 bits and then saturated to 1.15 format.

Refer to `riscv_mat_mult_fast_q15()` for a faster but less precise version of this function.

**Parameters**

- [in] `pSrcA`: points to the first input matrix structure
- [in] `pSrcB`: points to the second input matrix structure
- [out] `pDst`: points to output matrix structure
- [in] `pState`: points to the array for storing intermediate results (Unused)

```
riscv_status riscv_mat_mult_q31(const riscv_matrix_instance_q31 *pSrcA, const
                                riscv_matrix_instance_q31 *pSrcB, riscv_matrix_instance_q31
                                *pDst)
```

Q31 matrix multiplication.

**Return** execution status

- RISCVMATH\_SUCCESS : Operation successful
- RISCVMATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The accumulator has a 2.62 format and maintains full precision of the intermediate multiplication results but provides only a single guard bit. There is no saturation on intermediate additions. Thus, if the accumulator overflows it wraps around and distorts the result. The input signals should be scaled down to avoid intermediate overflows. The input is thus scaled down by  $\log_2(\text{numColsA})$  bits to avoid overflows, as a total of  $\text{numColsA}$  additions are performed internally. The 2.62 accumulator is right shifted by 31 bits and saturated to 1.31 format to yield the final result.

**Remark** Refer to `riscv_mat_mult_fast_q31()` for a faster but less precise implementation of this function.

**Parameters**

- [in] pSrcA: points to the first input matrix structure
- [in] pSrcB: points to the second input matrix structure
- [out] pDst: points to output matrix structure

```
riscv_status riscv_mat_mult_q7(const riscv_matrix_instance_q7 *pSrcA, const
                                riscv_matrix_instance_q7 *pSrcB, riscv_matrix_instance_q7
                                *pDst, q7_t *pState)
```

Q7 matrix multiplication.

**Scaling and Overflow Behavior:**

**Return** The function returns either RISCVMATH\_SIZE\_MISMATCH or RISCVMATH\_SUCCESS based on the outcome of size checking.

**Parameters**

- [in] \*pSrcA: points to the first input matrix structure
- [in] \*pSrcB: points to the second input matrix structure
- [out] \*pDst: points to output matrix structure
- [in] \*pState: points to the array for storing intermediate results (Unused in some versions)

The function is implemented using a 32-bit internal accumulator saturated to 1.7 format.

## Matrix Scale

```
riscv_status riscv_mat_scale_f16(const riscv_matrix_instance_f16 *pSrc, float16_t scale,
                                riscv_matrix_instance_f16 *pDst)
```

```
riscv_status riscv_mat_scale_f32(const riscv_matrix_instance_f32 *pSrc, float32_t scale,
                                riscv_matrix_instance_f32 *pDst)
```

```
riscv_status riscv_mat_scale_q15(const riscv_matrix_instance_q15 *pSrc, q15_t scaleFract, int32_t
                                shift, riscv_matrix_instance_q15 *pDst)
```

```
riscv_status riscv_mat_scale_q31 (const riscv_matrix_instance_q31 *pSrc, q31_t scaleFract, int32_t
                                shift, riscv_matrix_instance_q31 *pDst)
```

*group* **MatrixScale**

Multiplies a matrix by a scalar. This is accomplished by multiplying each element in the matrix by the scalar. For example: The function checks to make sure that the input and output matrices are of the same size.

$$\begin{bmatrix} \overline{a_{11}} & a_{12} & \overline{a_{13}} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & \overline{a_{33}} \end{bmatrix} \times K = \begin{bmatrix} \overline{a_{11} \times K} & a_{12 \times K} & \overline{a_{13} \times K} \\ a_{21 \times K} & a_{22 \times K} & a_{23 \times K} \\ a_{31 \times K} & a_{32 \times K} & \overline{a_{33 \times K}} \end{bmatrix}$$

In the fixed-point Q15 and Q31 functions, *scale* is represented by a fractional multiplication *scaleFract* and an arithmetic shift *shift*. The shift allows the gain of the scaling operation to exceed 1.0. The overall scale factor applied to the fixed-point data is

## Functions

```
riscv_status riscv_mat_scale_f16 (const riscv_matrix_instance_f16 *pSrc, float16_t scale,
                                riscv_matrix_instance_f16 *pDst)
```

Floating-point matrix scaling.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [in] scale: scale factor to be applied
- [out] pDst: points to output matrix structure

```
riscv_status riscv_mat_scale_f32 (const riscv_matrix_instance_f32 *pSrc, float32_t scale,
                                riscv_matrix_instance_f32 *pDst)
```

Floating-point matrix scaling.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [in] scale: scale factor to be applied
- [out] pDst: points to output matrix structure

```
riscv_status riscv_mat_scale_q15 (const riscv_matrix_instance_q15 *pSrc, q15_t scaleFract,
                                int32_t shift, riscv_matrix_instance_q15 *pDst)
```

Q15 matrix scaling.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The input data *\*pSrc* and *scaleFract* are in 1.15 format. These are multiplied to yield a 2.30 intermediate result and this is shifted with saturation to 1.15 format.

**Parameters**

- [in] *pSrc*: points to input matrix
- [in] *scaleFract*: fractional portion of the scale factor
- [in] *shift*: number of bits to shift the result by
- [out] *pDst*: points to output matrix structure

```
riscv_status riscv_mat_scale_q31 (const riscv_matrix_instance_q31 *pSrc, q31_t scaleFract,  
                                     int32_t shift, riscv_matrix_instance_q31 *pDst)
```

Q31 matrix scaling.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The input data *\*pSrc* and *scaleFract* are in 1.31 format. These are multiplied to yield a 2.62 intermediate result which is shifted with saturation to 1.31 format.

**Parameters**

- [in] *pSrc*: points to input matrix
- [in] *scaleFract*: fractional portion of the scale factor
- [in] *shift*: number of bits to shift the result by
- [out] *pDst*: points to output matrix structure

## Matrix Subtraction

```
riscv_status riscv_mat_sub_f16 (const      riscv_matrix_instance_f16      *pSrcA,      const  
                                   riscv_matrix_instance_f16 *pSrcB, riscv_matrix_instance_f16 *pDst)
```

```
riscv_status riscv_mat_sub_f32 (const      riscv_matrix_instance_f32      *pSrcA,      const  
                                   riscv_matrix_instance_f32 *pSrcB, riscv_matrix_instance_f32 *pDst)
```

```
riscv_status riscv_mat_sub_f64 (const      riscv_matrix_instance_f64      *pSrcA,      const  
                                   riscv_matrix_instance_f64 *pSrcB, riscv_matrix_instance_f64 *pDst)
```

```
riscv_status riscv_mat_sub_q15 (const      riscv_matrix_instance_q15      *pSrcA,      const  
                                   riscv_matrix_instance_q15 *pSrcB, riscv_matrix_instance_q15 *pDst)
```

```
riscv_status riscv_mat_sub_q31 (const      riscv_matrix_instance_q31      *pSrcA,      const  
                                   riscv_matrix_instance_q31 *pSrcB, riscv_matrix_instance_q31 *pDst)
```

*group* **MatrixSub**

Subtract two matrices. The functions check to make sure that *pSrcA*,

`pSrcB`, and `pDst` have the same number of rows and columns.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}-b_{11} & a_{12}-b_{12} & a_{13}-b_{13} \\ a_{21}-b_{21} & a_{22}-b_{22} & a_{23}-b_{23} \\ a_{31}-b_{31} & a_{32}-b_{32} & a_{33}-b_{33} \end{bmatrix}$$

## Functions

riscv\_status **riscv\_mat\_sub\_f16** (const riscv\_matrix\_instance\_f16 \**pSrcA*, const riscv\_matrix\_instance\_f16 \**pSrcB*, riscv\_matrix\_instance\_f16 \**pDst*)

Floating-point matrix subtraction.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] `pSrcA`: points to the first input matrix structure
- [in] `pSrcB`: points to the second input matrix structure
- [out] `pDst`: points to output matrix structure

riscv\_status **riscv\_mat\_sub\_f32** (const riscv\_matrix\_instance\_f32 \**pSrcA*, const riscv\_matrix\_instance\_f32 \**pSrcB*, riscv\_matrix\_instance\_f32 \**pDst*)

Floating-point matrix subtraction.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] `pSrcA`: points to the first input matrix structure
- [in] `pSrcB`: points to the second input matrix structure
- [out] `pDst`: points to output matrix structure

riscv\_status **riscv\_mat\_sub\_f64** (const riscv\_matrix\_instance\_f64 \**pSrcA*, const riscv\_matrix\_instance\_f64 \**pSrcB*, riscv\_matrix\_instance\_f64 \**pDst*)

Floating-point matrix subtraction.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrcA: points to the first input matrix structure
- [in] pSrcB: points to the second input matrix structure
- [out] pDst: points to output matrix structure

```
riscv_status riscv_mat_sub_q15 (const      riscv_matrix_instance_q15  *pSrcA, const  
                                riscv_matrix_instance_q15  *pSrcB, riscv_matrix_instance_q15  
                                *pDst)
```

Q15 matrix subtraction.

**Return** execution status

- RISCVMATH\_SUCCESS : Operation successful
- RISCVMATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

**Parameters**

- [in] pSrcA: points to the first input matrix structure
- [in] pSrcB: points to the second input matrix structure
- [out] pDst: points to output matrix structure

```
riscv_status riscv_mat_sub_q31 (const      riscv_matrix_instance_q31  *pSrcA, const  
                                riscv_matrix_instance_q31  *pSrcB, riscv_matrix_instance_q31  
                                *pDst)
```

Q31 matrix subtraction.

**Return** execution status

- RISCVMATH\_SUCCESS : Operation successful
- RISCVMATH\_SIZE\_MISMATCH : Matrix size check failed

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range [0x80000000 0x7FFFFFFF] are saturated.

**Parameters**

- [in] pSrcA: points to the first input matrix structure
- [in] pSrcB: points to the second input matrix structure
- [out] pDst: points to output matrix structure

## Matrix Transpose

```
riscv_status riscv_mat_trans_f16 (const riscv_matrix_instance_f16 *pSrc, riscv_matrix_instance_f16  
                                *pDst)
```

```
riscv_status riscv_mat_trans_f32 (const riscv_matrix_instance_f32 *pSrc, riscv_matrix_instance_f32  
                                *pDst)
```

```
riscv_status riscv_mat_trans_f64 (const riscv_matrix_instance_f64 *pSrc, riscv_matrix_instance_f64  
                                *pDst)
```

```
riscv_status riscv_mat_trans_q15 (const      riscv_matrix_instance_q15  *pSrc,  
                                riscv_matrix_instance_q15 *pDst)
```

```
riscv_status riscv_mat_trans_q31 (const riscv_matrix_instance_q31 *pSrc,
                                     riscv_matrix_instance_q31 *pDst)
```

```
riscv_status riscv_mat_trans_q7 (const riscv_matrix_instance_q7 *pSrc, riscv_matrix_instance_q7
                                     *pDst)
```

*group* **MatrixTrans**

Tranposes a matrix.

Transposing an  $M \times N$  matrix flips it around the center diagonal and results in an  $N \times M$  matrix.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

## Functions

```
riscv_status riscv_mat_trans_f16 (const riscv_matrix_instance_f16 *pSrc,
                                     riscv_matrix_instance_f16 *pDst)
```

Floating-point matrix transpose.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

```
riscv_status riscv_mat_trans_f32 (const riscv_matrix_instance_f32 *pSrc,
                                     riscv_matrix_instance_f32 *pDst)
```

Floating-point matrix transpose.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

```
riscv_status riscv_mat_trans_f64 (const riscv_matrix_instance_f64 *pSrc,
                                     riscv_matrix_instance_f64 *pDst)
```

Floating-point matrix transpose.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

riscv\_status **riscv\_mat\_trans\_q15** (const riscv\_matrix\_instance\_q15 \*pSrc,  
riscv\_matrix\_instance\_q15 \*pDst)  
Q15 matrix transpose.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

riscv\_status **riscv\_mat\_trans\_q31** (const riscv\_matrix\_instance\_q31 \*pSrc,  
riscv\_matrix\_instance\_q31 \*pDst)  
Q31 matrix transpose.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix

riscv\_status **riscv\_mat\_trans\_q7** (const riscv\_matrix\_instance\_q7 \*pSrc,  
riscv\_matrix\_instance\_q7 \*pDst)  
Q7 matrix transpose.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_SIZE\_MISMATCH : Matrix size check failed

**Parameters**

- [in] pSrc: points to input matrix
- [out] pDst: points to output matrix



## Matrix Vector Multiplication

```
void riscv_mat_vec_mult_f16 (const riscv_matrix_instance_f16 *pSrcMat, const float16_t *pVec,
                             float16_t *pDst)
```

```
void riscv_mat_vec_mult_f32 (const riscv_matrix_instance_f32 *pSrcMat, const float32_t *pVec,
                             float32_t *pDst)
```

```
void riscv_mat_vec_mult_q15 (const riscv_matrix_instance_q15 *pSrcMat, const q15_t *pVec,
                             q15_t *pDst)
```

```
void riscv_mat_vec_mult_q31 (const riscv_matrix_instance_q31 *pSrcMat, const q31_t *pVec,
                             q31_t *pDst)
```

```
void riscv_mat_vec_mult_q7 (const riscv_matrix_instance_q7 *pSrcMat, const q7_t *pVec, q7_t
                             *pDst)
```

group **MatrixVectMult**

Multiplies a matrix and a vector.

## Functions

```
void riscv_mat_vec_mult_f16 (const riscv_matrix_instance_f16 *pSrcMat, const float16_t
                             *pVec, float16_t *pDst)
```

Floating-point matrix and vector multiplication.

### Parameters

- [in] \*pSrcMat: points to the input matrix structure
- [in] \*pVec: points to input vector
- [out] \*pDst: points to output vector

```
void riscv_mat_vec_mult_f32 (const riscv_matrix_instance_f32 *pSrcMat, const float32_t
                             *pVec, float32_t *pDst)
```

Floating-point matrix and vector multiplication.

### Parameters

- [in] \*pSrcMat: points to the input matrix structure
- [in] \*pVec: points to input vector
- [out] \*pDst: points to output vector

```
void riscv_mat_vec_mult_q15 (const riscv_matrix_instance_q15 *pSrcMat, const q15_t
                             *pVec, q15_t *pDst)
```

Q15 matrix and vector multiplication.

### Parameters

- [in] \*pSrcMat: points to the input matrix structure
- [in] \*pVec: points to input vector
- [out] \*pDst: points to output vector

```
void riscv_mat_vec_mult_q31 (const riscv_matrix_instance_q31 *pSrcMat, const q31_t
                             *pVec, q31_t *pDst)
```

Q31 matrix and vector multiplication.

**Parameters**

- [in] *\*pSrcMat*: points to the input matrix structure
- [in] *\*pVec*: points to the input vector
- [out] *\*pDst*: points to the output vector

void **riscv\_mat\_vec\_mult\_q7** (**const** riscv\_matrix\_instance\_q7 *\*pSrcMat*, **const** q7\_t *\*pVec*,  
q7\_t *\*pDst*)  
Q7 matrix and vector multiplication.

**Parameters**

- [in] *\*pSrcMat*: points to the input matrix structure
- [in] *\*pVec*: points to the input vector
- [out] *\*pDst*: points to the output vector

*group* **groupMatrix**

This set of functions provides basic matrix math operations. The functions operate on matrix data structures. For example, the type definition for the floating-point matrix structure is shown below: There are similar definitions for Q15 and Q31 data types.

The structure specifies the size of the matrix and then points to an array of data. The array is of size `numRows X numCols` and the values are arranged in row order. That is, the matrix element  $(i, j)$  is stored at:

**Init Functions** There is an associated initialization function for each type of matrix data structure. The initialization function sets the values of the internal structure fields. Refer to `riscv_mat_init_f32()`, `riscv_mat_init_q31()` and `riscv_mat_init_q15()` for floating-point, Q31 and Q15 types, respectively.

Use of the initialization function is optional. However, if initialization function is used then the instance structure cannot be placed into a `const` data section. To place the instance structure in a `const` data section, manually initialize the data structure. For example: where `nRows` specifies the number of rows, `nColumns` specifies the number of columns, and `pData` points to the data array.

**Size Checking** By default all of the matrix functions perform size checking on the input and output matrices. For example, the matrix addition function verifies that the two input matrices and the output matrix all have the same number of rows and columns. If the size check fails the functions return: Otherwise the functions return `There is some overhead associated with this matrix size checking. The matrix size checking is enabled via the #define within the library project settings. By default this macro is defined and size checking is enabled. By changing the project settings and undefining this macro size checking is eliminated and the functions run a bit faster. With size checking disabled the functions always return RISCV_MATH_SUCCESS.`

### 3.3.11 Quaternion Math Functions

**Quaternion conversions****Quaternion to Rotation**

void **riscv\_quaternion2rotation\_f32** (**const** float32\_t *\*pInputQuaternions*, float32\_t *\*pOutputRotations*, uint32\_t *nbQuaternions*)

*group* **QuatRot**

Conversions from quaternion to rotation.

## Functions

void **riscv\_quaternion2rotation\_f32** (const float32\_t \*pInputQuaternions, float32\_t \*pOutputRotations, uint32\_t nbQuaternions)

Conversion of quaternion to equivalent rotation matrix.

The quaternion  $a + ib + jc + kd$  is converted into rotation matrix: Rotation matrix is saved in row order :  
R00 R01 R02 R10 R11 R12 R20 R21 R22

**Return** none.

Format of rotation matrix

### Parameters

- [in] pInputQuaternions: points to an array of normalized quaternions
- [out] pOutputRotations: points to an array of 3x3 rotations (in row order)
- [in] nbQuaternions: number of quaternions in the array

## Rotation to Quaternion

void **riscv\_rotation2quaternion\_f32** (const float32\_t \*pInputRotations, float32\_t \*pOutputQuaternions, uint32\_t nbQuaternions)

group **RotQuat**

Conversions from rotation to quaternion.

## Functions

void **riscv\_rotation2quaternion\_f32** (const float32\_t \*pInputRotations, float32\_t \*pOutputQuaternions, uint32\_t nbQuaternions)

Conversion of a rotation matrix to an equivalent quaternion.

Conversion of a rotation matrix to equivalent quaternion.

$q$  and  $-q$  are representing the same rotation. This ambiguity must be taken into account when using the output of this function.

**Return** none.

### Parameters

- [in] pInputRotations: points to an array 3x3 rotation matrix (in row order)
- [out] pOutputQuaternions: points to an array quaternions
- [in] nbQuaternions: number of quaternions in the array

group **QuatConv**

Conversions between quaternion and rotation representations.

## Quaternion Conjugate

void **riscv\_quaternion\_conjugate\_f32** (const float32\_t \*pInputQuaternions, float32\_t \*pConjugateQuaternions, uint32\_t nbQuaternions)

group **QuatConjugate**

Compute the conjugate of a quaternion.

## Functions

void **riscv\_quaternion\_conjugate\_f32** (**const** float32\_t \*pInputQuaternions, float32\_t \*pConjugateQuaternions, uint32\_t nbQuaternions)

Floating-point quaternion conjugates.

**Return** none

### Parameters

- [in] pInputQuaternions: points to the input vector of quaternions
- [out] pConjugateQuaternions: points to the output vector of conjugate quaternions
- [in] nbQuaternions: number of quaternions in each vector

## Quaternion Inverse

void **riscv\_quaternion\_inverse\_f32** (**const** float32\_t \*pInputQuaternions, float32\_t \*pInverseQuaternions, uint32\_t nbQuaternions)

group **QuatInverse**

Compute the inverse of a quaternion.

## Functions

void **riscv\_quaternion\_inverse\_f32** (**const** float32\_t \*pInputQuaternions, float32\_t \*pInverseQuaternions, uint32\_t nbQuaternions)

Floating-point quaternion inverse.

**Return** none

### Parameters

- [in] pInputQuaternions: points to the input vector of quaternions
- [out] pInverseQuaternions: points to the output vector of inverse quaternions
- [in] nbQuaternions: number of quaternions in each vector

## Quaternion Norm

void **riscv\_quaternion\_norm\_f32** (**const** float32\_t \*pInputQuaternions, float32\_t \*pNorms, uint32\_t nbQuaternions)

group **QuatNorm**

Compute the norm of a quaternion.

## Functions

void **riscv\_quaternion\_norm\_f32** (**const** float32\_t \*pInputQuaternions, float32\_t \*pNorms, uint32\_t nbQuaternions)

Floating-point quaternion Norm.

**Return** none

### Parameters

- [in] `pInputQuaternions`: points to the input vector of quaternions
- [out] `pNorms`: points to the output vector of norms
- [in] `nbQuaternions`: number of quaternions in the input vector

### Quaternion normalization

void **riscv\_quaternion\_normalize\_f32** (**const** float32\_t \**pInputQuaternions*, float32\_t \**pNormalizedQuaternions*, uint32\_t *nbQuaternions*)

group **QuatNormalized**

Compute a normalized quaternion.

### Functions

void **riscv\_quaternion\_normalize\_f32** (**const** float32\_t \**pInputQuaternions*, float32\_t \**pNormalizedQuaternions*, uint32\_t *nbQuaternions*)

Floating-point normalization of quaternions.

**Return** none

**Parameters**

- [in] `pInputQuaternions`: points to the input vector of quaternions
- [out] `pNormalizedQuaternions`: points to the output vector of normalized quaternions
- [in] `nbQuaternions`: number of quaternions in each vector

### Quaternion Product

#### Elementwise Quaternion Product

void **riscv\_quaternion\_product\_f32** (**const** float32\_t \**qa*, **const** float32\_t \**qb*, float32\_t \**qr*, uint32\_t *nbQuaternions*)

group **QuatProdVect**

Compute the elementwise product of quaternions.

### Functions

void **riscv\_quaternion\_product\_f32** (**const** float32\_t \**qa*, **const** float32\_t \**qb*, float32\_t \**qr*, uint32\_t *nbQuaternions*)

Floating-point elementwise product two quaternions.

**Return** none

**Parameters**

- [in] `qa`: first array of quaternions
- [in] `qb`: second array of quaternions
- [out] `qr`: elementwise product of quaternions
- [in] `nbQuaternions`: number of quaternions in the array

## Quaternion Product

```
void riscv_quaternion_product_single_f32 (const float32_t *qa, const float32_t *qb,  
                                          float32_t *qr)
```

*group* **QuatProdSingle**

Compute the product of two quaternions.

## Functions

```
void riscv_quaternion_product_single_f32 (const float32_t *qa, const float32_t *qb,  
                                          float32_t *qr)
```

Floating-point product of two quaternions.

**Return** none

**Parameters**

- [in] qa: first quaternion
- [in] qb: second quaternion
- [out] qr: product of two quaternions

*group* **QuatProd**

Compute the product of quaternions.

*group* **groupQuaternionMath**

Functions to operates on quaternions and convert between a rotation and quaternion representation.

## 3.3.12 Statistics Functions

### Absolute Maximum

```
void riscv_absmax_f16 (const float16_t *pSrc, uint32_t blockSize, float16_t *pResult, uint32_t *pIndex)
```

```
void riscv_absmax_f32 (const float32_t *pSrc, uint32_t blockSize, float32_t *pResult, uint32_t *pIndex)
```

```
void riscv_absmax_q15 (const q15_t *pSrc, uint32_t blockSize, q15_t *pResult, uint32_t *pIndex)
```

```
void riscv_absmax_q31 (const q31_t *pSrc, uint32_t blockSize, q31_t *pResult, uint32_t *pIndex)
```

```
void riscv_absmax_q7 (const q7_t *pSrc, uint32_t blockSize, q7_t *pResult, uint32_t *pIndex)
```

*group* **AbsMax**

Computes the maximum value of absolute values of an array of data. The function returns both the maximum value and its position within the array. There are separate functions for floating-point, Q31, Q15, and Q7 data types.

## Functions

```
void riscv_absmax_f16 (const float16_t *pSrc, uint32_t blockSize, float16_t *pResult, uint32_t  
                      *pIndex)
```

Maximum value of absolute values of a floating-point vector.

**Return** none

**Parameters**

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: maximum value returned here
- [out] `pIndex`: index of maximum value returned here

void **riscv\_absmax\_f32** (**const** float32\_t \**pSrc*, uint32\_t *blockSize*, float32\_t \**pResult*, uint32\_t \**pIndex*)

Maximum value of absolute values of a floating-point vector.

**Return** none

**Parameters**

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: maximum value returned here
- [out] `pIndex`: index of maximum value returned here

void **riscv\_absmax\_q15** (**const** q15\_t \**pSrc*, uint32\_t *blockSize*, q15\_t \**pResult*, uint32\_t \**pIndex*)

Maximum value of absolute values of a Q15 vector.

**Return** none

**Parameters**

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: maximum value returned here
- [out] `pIndex`: index of maximum value returned here

void **riscv\_absmax\_q31** (**const** q31\_t \**pSrc*, uint32\_t *blockSize*, q31\_t \**pResult*, uint32\_t \**pIndex*)

Maximum value of absolute values of a Q31 vector.

**Return** none

**Parameters**

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: maximum value returned here
- [out] `pIndex`: index of maximum value returned here

void **riscv\_absmax\_q7** (**const** q7\_t \**pSrc*, uint32\_t *blockSize*, q7\_t \**pResult*, uint32\_t \**pIndex*)

Maximum value of absolute values of a Q7 vector.

**Return** none

**Parameters**

- [in] `pSrc`: points to the input vector

- [in] blockSize: number of samples in input vector
- [out] pResult: maximum value returned here
- [out] pIndex: index of maximum value returned here

## Absolute Minimum

void **riscv\_absmin\_f16**(const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult, uint32\_t \*pIndex)

void **riscv\_absmin\_f32**(const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult, uint32\_t \*pIndex)

void **riscv\_absmin\_q15**(const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult, uint32\_t \*pIndex)

void **riscv\_absmin\_q31**(const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult, uint32\_t \*pIndex)

void **riscv\_absmin\_q7**(const q7\_t \*pSrc, uint32\_t blockSize, q7\_t \*pResult, uint32\_t \*pIndex)

### group AbsMin

Computes the minimum value of absolute values of an array of data. The function returns both the minimum value and its position within the array. There are separate functions for floating-point, Q31, Q15, and Q7 data types.

## Functions

void **riscv\_absmin\_f16**(const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult, uint32\_t \*pIndex)

Minimum value of absolute values of a floating-point vector.

**Return** none

### Parameters

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: minimum value returned here
- [out] pIndex: index of minimum value returned here

void **riscv\_absmin\_f32**(const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult, uint32\_t \*pIndex)

Minimum value of absolute values of a floating-point vector.

**Return** none

### Parameters

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: minimum value returned here
- [out] pIndex: index of minimum value returned here

void **riscv\_absmin\_q15**(const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult, uint32\_t \*pIndex)

Minimum value of absolute values of a Q15 vector.



**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: minimum value returned here
- [out] *pIndex*: index of minimum value returned here

void **riscv\_absmin\_q31** (**const** q31\_t \**pSrc*, uint32\_t *blockSize*, q31\_t \**pResult*, uint32\_t \**pIndex*)  
Minimum value of absolute values of a Q31 vector.

**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: minimum value returned here
- [out] *pIndex*: index of minimum value returned here

void **riscv\_absmin\_q7** (**const** q7\_t \**pSrc*, uint32\_t *blockSize*, q7\_t \**pResult*, uint32\_t \**pIndex*)  
Minimum value of absolute values of a Q7 vector.

**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: minimum value returned here
- [out] *pIndex*: index of minimum value returned here

## Entropy

float16\_t **riscv\_entropy\_f16** (**const** float16\_t \**pSrcA*, uint32\_t *blockSize*)

float32\_t **riscv\_entropy\_f32** (**const** float32\_t \**pSrcA*, uint32\_t *blockSize*)

float64\_t **riscv\_entropy\_f64** (**const** float64\_t \**pSrcA*, uint32\_t *blockSize*)

*group* **Entropy**

Computes the entropy of a distribution

## Functions

float16\_t **riscv\_entropy\_f16** (**const** float16\_t \**pSrcA*, uint32\_t *blockSize*)  
Entropy.

**Return** Entropy -Sum( $p \ln p$ )

**Parameters**

- [in] pSrcA: Array of input values.
- [in] blockSize: Number of samples in the input array.

float32\_t **riscv\_entropy\_f32** (const float32\_t \*pSrcA, uint32\_t blockSize)  
Entropy.

**Return** Entropy -Sum( $p \ln p$ )

**Parameters**

- [in] pSrcA: Array of input values.
- [in] blockSize: Number of samples in the input array.

float64\_t **riscv\_entropy\_f64** (const float64\_t \*pSrcA, uint32\_t blockSize)  
Entropy.

**Return** Entropy -Sum( $p \ln p$ )

**Parameters**

- [in] pSrcA: Array of input values.
- [in] blockSize: Number of samples in the input array.

## Kullback-Leibler divergence

float16\_t **riscv\_kullback\_leibler\_f16** (const float16\_t \*pSrcA, const float16\_t \*pSrcB, uint32\_t blockSize)

float32\_t **riscv\_kullback\_leibler\_f32** (const float32\_t \*pSrcA, const float32\_t \*pSrcB, uint32\_t blockSize)

float64\_t **riscv\_kullback\_leibler\_f64** (const float64\_t \*pSrcA, const float64\_t \*pSrcB, uint32\_t blockSize)

*group* **Kullback-Leibler**

Computes the Kullback-Leibler divergence between two distributions

## Functions

float16\_t **riscv\_kullback\_leibler\_f16** (const float16\_t \*pSrcA, const float16\_t \*pSrcB, uint32\_t blockSize)

Kullback-Leibler.

Distribution A may contain 0 with Neon version. Result will be right but some exception flags will be set.

Distribution B must not contain 0 probability.

**Return** Kullback-Leibler divergence  $D(A \parallel B)$

**Parameters**

- [in] \*pSrcA: points to an array of input values for probability distribution A.
- [in] \*pSrcB: points to an array of input values for probability distribution B.
- [in] blockSize: number of samples in the input array.

`float32_t riscv_kullback_leibler_f32 (const float32_t *pSrcA, const float32_t *pSrcB, uint32_t blockSize)`

Kullback-Leibler.

Distribution A may contain 0 with Neon version. Result will be right but some exception flags will be set.

Distribution B must not contain 0 probability.

**Return** Kullback-Leibler divergence  $D(A \parallel B)$

**Parameters**

- [in] `*pSrcA`: points to an array of input values for probability distribution A.
- [in] `*pSrcB`: points to an array of input values for probability distribution B.
- [in] `blockSize`: number of samples in the input array.

`float64_t riscv_kullback_leibler_f64 (const float64_t *pSrcA, const float64_t *pSrcB, uint32_t blockSize)`

Kullback-Leibler.

**Return** Kullback-Leibler divergence  $D(A \parallel B)$

**Parameters**

- [in] `*pSrcA`: points to an array of input values for probability distribution A.
- [in] `*pSrcB`: points to an array of input values for probability distribution B.
- [in] `blockSize`: number of samples in the input array.

## LogSumExp

`float16_t riscv_logsumexp_dot_prod_f16 (const float16_t *pSrcA, const float16_t *pSrcB, uint32_t blockSize, float16_t *pTmpBuffer)`

`float32_t riscv_logsumexp_dot_prod_f32 (const float32_t *pSrcA, const float32_t *pSrcB, uint32_t blockSize, float32_t *pTmpBuffer)`

`float16_t riscv_logsumexp_f16 (const float16_t *in, uint32_t blockSize)`

`float32_t riscv_logsumexp_f32 (const float32_t *in, uint32_t blockSize)`

*group* **LogSumExp**

LogSumExp optimizations to compute sum of probabilities with Gaussian distributions

## Functions

`float16_t riscv_logsumexp_dot_prod_f16 (const float16_t *pSrcA, const float16_t *pSrcB, uint32_t blockSize, float16_t *pTmpBuffer)`

Dot product with log arithmetic.

Vectors are containing the log of the samples

**Return** The log of the dot product.

**Parameters**

- [in] `*pSrcA`: points to the first input vector
- [in] `*pSrcB`: points to the second input vector

- [in] blockSize: number of samples in each vector
- [in] \*pTmpBuffer: temporary buffer of length blockSize

float32\_t **riscv\_logsumexp\_dot\_prod\_f32** (const float32\_t \*pSrcA, const float32\_t \*pSrcB,  
uint32\_t blockSize, float32\_t \*pTmpBuffer)

Dot product with log arithmetic.

Vectors are containing the log of the samples

**Return** The log of the dot product.

**Parameters**

- [in] \*pSrcA: points to the first input vector
- [in] \*pSrcB: points to the second input vector
- [in] blockSize: number of samples in each vector
- [in] \*pTmpBuffer: temporary buffer of length blockSize

float16\_t **riscv\_logsumexp\_f16** (const float16\_t \*in, uint32\_t blockSize)

Computation of the LogSumExp.

In probabilistic computations, the dynamic of the probability values can be very wide because they come from gaussian functions. To avoid underflow and overflow issues, the values are represented by their log. In this representation, multiplying the original exp values is easy : their logs are added. But adding the original exp values is requiring some special handling and it is the goal of the LogSumExp function.

If the values are  $x_1 \dots x_n$ , the function is computing:

$\ln(\exp(x_1) + \dots + \exp(x_n))$  and the computation is done in such a way that rounding issues are minimised.

The max  $x_m$  of the values is extracted and the function is computing:  $x_m + \ln(\exp(x_1 - x_m) + \dots + \exp(x_n - x_m))$

**Return** LogSumExp

**Parameters**

- [in] \*in: Pointer to an array of input values.
- [in] blockSize: Number of samples in the input array.

float32\_t **riscv\_logsumexp\_f32** (const float32\_t \*in, uint32\_t blockSize)

Computation of the LogSumExp.

In probabilistic computations, the dynamic of the probability values can be very wide because they come from gaussian functions. To avoid underflow and overflow issues, the values are represented by their log. In this representation, multiplying the original exp values is easy : their logs are added. But adding the original exp values is requiring some special handling and it is the goal of the LogSumExp function.

If the values are  $x_1 \dots x_n$ , the function is computing:

$\ln(\exp(x_1) + \dots + \exp(x_n))$  and the computation is done in such a way that rounding issues are minimised.

The max  $x_m$  of the values is extracted and the function is computing:  $x_m + \ln(\exp(x_1 - x_m) + \dots + \exp(x_n - x_m))$

**Return** LogSumExp

**Parameters**

- [in] *\*in*: Pointer to an array of input values.
- [in] *blockSize*: Number of samples in the input array.

## Maximum

```
void riscv_max_f16 (const float16_t *pSrc, uint32_t blockSize, float16_t *pResult, uint32_t *pIndex)
void riscv_max_f32 (const float32_t *pSrc, uint32_t blockSize, float32_t *pResult, uint32_t *pIndex)
void riscv_max_no_idx_f16 (const float16_t *pSrc, uint32_t blockSize, float16_t *pResult)
void riscv_max_no_idx_f32 (const float32_t *pSrc, uint32_t blockSize, float32_t *pResult)
void riscv_max_q15 (const q15_t *pSrc, uint32_t blockSize, q15_t *pResult, uint32_t *pIndex)
void riscv_max_q31 (const q31_t *pSrc, uint32_t blockSize, q31_t *pResult, uint32_t *pIndex)
void riscv_max_q7 (const q7_t *pSrc, uint32_t blockSize, q7_t *pResult, uint32_t *pIndex)
```

### group **Max**

Computes the maximum value of an array of data. The function returns both the maximum value and its position within the array. There are separate functions for floating-point, Q31, Q15, and Q7 data types.

## Functions

```
void riscv_max_f16 (const float16_t *pSrc, uint32_t blockSize, float16_t *pResult, uint32_t *pIndex)
                        dex)
Maximum value of a floating-point vector.
```

**Return** none

### Parameters

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: maximum value returned here
- [out] *pIndex*: index of maximum value returned here

```
void riscv_max_f32 (const float32_t *pSrc, uint32_t blockSize, float32_t *pResult, uint32_t *pIndex)
                        dex)
Maximum value of a floating-point vector.
```

**Return** none

### Parameters

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: maximum value returned here
- [out] *pIndex*: index of maximum value returned here

```
void riscv_max_no_idx_f16 (const float16_t *pSrc, uint32_t blockSize, float16_t *pResult)
Maximum value of a floating-point vector.
```

**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: maximum value returned here

void **riscv\_max\_no\_idx\_f32** (**const** float32\_t \**pSrc*, uint32\_t *blockSize*, float32\_t \**pResult*)  
Maximum value of a floating-point vector.

**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: maximum value returned here

void **riscv\_max\_q15** (**const** q15\_t \**pSrc*, uint32\_t *blockSize*, q15\_t \**pResult*, uint32\_t \**pIndex*)  
Maximum value of a Q15 vector.

**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: maximum value returned here
- [out] *pIndex*: index of maximum value returned here

void **riscv\_max\_q31** (**const** q31\_t \**pSrc*, uint32\_t *blockSize*, q31\_t \**pResult*, uint32\_t \**pIndex*)  
Maximum value of a Q31 vector.

**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: maximum value returned here
- [out] *pIndex*: index of maximum value returned here

void **riscv\_max\_q7** (**const** q7\_t \**pSrc*, uint32\_t *blockSize*, q7\_t \**pResult*, uint32\_t \**pIndex*)  
Maximum value of a Q7 vector.

**Return** none

**Parameters**

- [in] *pSrc*: points to the input vector
- [in] *blockSize*: number of samples in input vector
- [out] *pResult*: maximum value returned here
- [out] *pIndex*: index of maximum value returned here

## Mean

void **riscv\_mean\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult)

void **riscv\_mean\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult)

void **riscv\_mean\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult)

void **riscv\_mean\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult)

void **riscv\_mean\_q7** (const q7\_t \*pSrc, uint32\_t blockSize, q7\_t \*pResult)

### group mean

Calculates the mean of the input vector. Mean is defined as the average of the elements in the vector. The underlying algorithm is used:

There are separate functions for floating-point, Q31, Q15, and Q7 data types.

## Functions

void **riscv\_mean\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult)

Mean value of a floating-point vector.

**Return** none

### Parameters

- [in] pSrc: points to the input vector.
- [in] blockSize: number of samples in input vector.
- [out] pResult: mean value returned here.

void **riscv\_mean\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult)

Mean value of a floating-point vector.

**Return** none

### Parameters

- [in] pSrc: points to the input vector.
- [in] blockSize: number of samples in input vector.
- [out] pResult: mean value returned here.

void **riscv\_mean\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult)

Mean value of a Q15 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. The input is represented in 1.15 format and is accumulated in a 32-bit accumulator in 17.15 format. There is no risk of internal overflow with this approach, and the full precision of intermediate result is preserved. Finally, the accumulator is truncated to yield a result of 1.15 format.

### Parameters

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: mean value returned here

void **riscv\_mean\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult)  
Mean value of a Q31 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. The input is represented in 1.31 format and is accumulated in a 64-bit accumulator in 33.31 format. There is no risk of internal overflow with this approach, and the full precision of intermediate result is preserved. Finally, the accumulator is truncated to yield a result of 1.31 format.

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: mean value returned here

void **riscv\_mean\_q7** (const q7\_t \*pSrc, uint32\_t blockSize, q7\_t \*pResult)  
Mean value of a Q7 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. The input is represented in 1.7 format and is accumulated in a 32-bit accumulator in 25.7 format. There is no risk of internal overflow with this approach, and the full precision of intermediate result is preserved. Finally, the accumulator is truncated to yield a result of 1.7 format.

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: mean value returned here

## Minimum

void **riscv\_min\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult, uint32\_t \*pIndex)

void **riscv\_min\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult, uint32\_t \*pIndex)

void **riscv\_min\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult, uint32\_t \*pIndex)

void **riscv\_min\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult, uint32\_t \*pIndex)

void **riscv\_min\_q7** (const q7\_t \*pSrc, uint32\_t blockSize, q7\_t \*pResult, uint32\_t \*pIndex)

*group* **Min**

Computes the minimum value of an array of data. The function returns both the minimum value and its position within the array. There are separate functions for floating-point, Q31, Q15, and Q7 data types.

## Functions

void **riscv\_min\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult, uint32\_t \*pIndex)  
*dex*

Minimum value of a floating-point vector.

**Return** none

**Parameters**



- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: minimum value returned here
- [out] pIndex: index of minimum value returned here

void **riscv\_min\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult, uint32\_t \*pIndex)  
 Minimum value of a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: minimum value returned here
- [out] pIndex: index of minimum value returned here

void **riscv\_min\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult, uint32\_t \*pIndex)  
 Minimum value of a Q15 vector.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: minimum value returned here
- [out] pIndex: index of minimum value returned here

void **riscv\_min\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult, uint32\_t \*pIndex)  
 Minimum value of a Q31 vector.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: minimum value returned here
- [out] pIndex: index of minimum value returned here

void **riscv\_min\_q7** (const q7\_t \*pSrc, uint32\_t blockSize, q7\_t \*pResult, uint32\_t \*pIndex)  
 Minimum value of a Q7 vector.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector

- [out] `pResult`: minimum value returned here
- [out] `pIndex`: index of minimum value returned here

## Power

void **riscv\_power\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult)

void **riscv\_power\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult)

void **riscv\_power\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q63\_t \*pResult)

void **riscv\_power\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q63\_t \*pResult)

void **riscv\_power\_q7** (const q7\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult)

### group power

Calculates the sum of the squares of the elements in the input vector. The underlying algorithm is used:

There are separate functions for floating point, Q31, Q15, and Q7 data types.

Since the result is not divided by the length, those functions are in fact computing something which is more an energy than a power.

## Functions

void **riscv\_power\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult)

Sum of the squares of the elements of a floating-point vector.

**Return** none

### Parameters

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: sum of the squares value returned here

void **riscv\_power\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult)

Sum of the squares of the elements of a floating-point vector.

**Return** none

### Parameters

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: sum of the squares value returned here

void **riscv\_power\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q63\_t \*pResult)

Sum of the squares of the elements of a Q15 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. The input is represented in 1.15 format. Intermediate multiplication yields a 2.30 format, and this result is

added without saturation to a 64-bit accumulator in 34.30 format. With 33 guard bits in the accumulator, there is no risk of overflow, and the full precision of the intermediate multiplication is preserved. Finally, the return result is in 34.30 format.

#### Parameters

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: sum of the squares value returned here

void **riscv\_power\_q31** (const q31\_t \**pSrc*, uint32\_t *blockSize*, q63\_t \**pResult*)  
Sum of the squares of the elements of a Q31 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. The input is represented in 1.31 format. Intermediate multiplication yields a 2.62 format, and this result is truncated to 2.48 format by discarding the lower 14 bits. The 2.48 result is then added without saturation to a 64-bit accumulator in 16.48 format. With 15 guard bits in the accumulator, there is no risk of overflow, and the full precision of the intermediate multiplication is preserved. Finally, the return result is in 16.48 format.

#### Parameters

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: sum of the squares value returned here

void **riscv\_power\_q7** (const q7\_t \**pSrc*, uint32\_t *blockSize*, q31\_t \**pResult*)  
Sum of the squares of the elements of a Q7 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 32-bit internal accumulator. The input is represented in 1.7 format. Intermediate multiplication yields a 2.14 format, and this result is added without saturation to an accumulator in 18.14 format. With 17 guard bits in the accumulator, there is no risk of overflow, and the full precision of the intermediate multiplication is preserved. Finally, the return result is in 18.14 format.

#### Parameters

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: sum of the squares value returned here

### Root mean square (RMS)

void **riscv\_rms\_f16** (const float16\_t \**pSrc*, uint32\_t *blockSize*, float16\_t \**pResult*)

void **riscv\_rms\_f32** (const float32\_t \**pSrc*, uint32\_t *blockSize*, float32\_t \**pResult*)

void **riscv\_rms\_q15** (const q15\_t \**pSrc*, uint32\_t *blockSize*, q15\_t \**pResult*)

void **riscv\_rms\_q31** (const q31\_t \**pSrc*, uint32\_t *blockSize*, q31\_t \**pResult*)

*group* **RMS**

Calculates the Root Mean Square of the elements in the input vector. The underlying algorithm is used:

There are separate functions for floating point, Q31, and Q15 data types.

**Functions**

void **riscv\_rms\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult)

Root Mean Square of the elements of a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: root mean square value returned here

void **riscv\_rms\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult)

Root Mean Square of the elements of a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: root mean square value returned here

void **riscv\_rms\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult)

Root Mean Square of the elements of a Q15 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. The input is represented in 1.15 format. Intermediate multiplication yields a 2.30 format, and this result is added without saturation to a 64-bit accumulator in 34.30 format. With 33 guard bits in the accumulator, there is no risk of overflow, and the full precision of the intermediate multiplication is preserved. Finally, the 34.30 result is truncated to 34.15 format by discarding the lower 15 bits, and then saturated to yield a result in 1.15 format.

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: root mean square value returned here

void **riscv\_rms\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult)

Root Mean Square of the elements of a Q31 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The input is represented in 1.31 format, and intermediate multiplication yields a 2.62 format. The accumulator maintains full precision of the intermediate multiplication results, but provides only a single guard bit. There is no saturation on intermediate additions. If the accumulator overflows, it wraps around and distorts the result. In order to avoid overflows completely, the input signal must be scaled down by  $\log_2(\text{blockSize})$  bits, as a total of  $\text{blockSize}$  additions are performed internally. Finally, the 2.62 accumulator is right shifted by 31 bits to yield a 1.31 format value.

#### Parameters

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: root mean square value returned here

### Standard deviation

```
void riscv_std_f16 (const float16_t *pSrc, uint32_t blockSize, float16_t *pResult)
```

```
void riscv_std_f32 (const float32_t *pSrc, uint32_t blockSize, float32_t *pResult)
```

```
void riscv_std_q15 (const q15_t *pSrc, uint32_t blockSize, q15_t *pResult)
```

```
void riscv_std_q31 (const q31_t *pSrc, uint32_t blockSize, q31_t *pResult)
```

#### group STD

Calculates the standard deviation of the elements in the input vector.

The float implementation is relying on `riscv_var_f32` which is using a two-pass algorithm to avoid problem of numerical instabilities and cancellation errors.

Fixed point versions are using the standard textbook algorithm since the fixed point numerical behavior is different from the float one.

Algorithm for fixed point versions is summarized below:

There are separate functions for floating point, Q31, and Q15 data types.

### Functions

```
void riscv_std_f16 (const float16_t *pSrc, uint32_t blockSize, float16_t *pResult)
```

Standard deviation of the elements of a floating-point vector.

**Return** none

#### Parameters

- [in] `pSrc`: points to the input vector
- [in] `blockSize`: number of samples in input vector
- [out] `pResult`: standard deviation value returned here

```
void riscv_std_f32 (const float32_t *pSrc, uint32_t blockSize, float32_t *pResult)
```

Standard deviation of the elements of a floating-point vector.

**Return** none

#### Parameters

- [in] `pSrc`: points to the input vector

- [in] blockSize: number of samples in input vector
- [out] pResult: standard deviation value returned here

void **riscv\_std\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult)  
Standard deviation of the elements of a Q15 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. The input is represented in 1.15 format. Intermediate multiplication yields a 2.30 format, and this result is added without saturation to a 64-bit accumulator in 34.30 format. With 33 guard bits in the accumulator, there is no risk of overflow, and the full precision of the intermediate multiplication is preserved. Finally, the 34.30 result is truncated to 34.15 format by discarding the lower 15 bits, and then saturated to yield a result in 1.15 format.

#### Parameters

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: standard deviation value returned here

void **riscv\_std\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult)  
Standard deviation of the elements of a Q31 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The input is represented in 1.31 format, which is then downshifted by 8 bits which yields 1.23, and intermediate multiplication yields a 2.46 format. The accumulator maintains full precision of the intermediate multiplication results, but provides only a 16 guard bits. There is no saturation on intermediate additions. If the accumulator overflows it wraps around and distorts the result. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{blockSize})-8$  bits, as a total of blockSize additions are performed internally. After division, internal variables should be Q18.46. Finally, the 18.46 accumulator is right shifted by 15 bits to yield a 1.31 format value.

#### Parameters

- [in] pSrc: points to the input vector.
- [in] blockSize: number of samples in input vector.
- [out] pResult: standard deviation value returned here.

## Variance

void **riscv\_var\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult)

void **riscv\_var\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult)

void **riscv\_var\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult)

void **riscv\_var\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult)

#### group variance

Calculates the variance of the elements in the input vector. The underlying algorithm used is the direct method sometimes referred to as the two-pass method:

There are separate functions for floating point, Q31, and Q15 data types.

## Functions

void **riscv\_var\_f16** (const float16\_t \*pSrc, uint32\_t blockSize, float16\_t \*pResult)  
 Variance of the elements of a floating-point vector.

**Return** none

### Parameters

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: variance value returned here

void **riscv\_var\_f32** (const float32\_t \*pSrc, uint32\_t blockSize, float32\_t \*pResult)  
 Variance of the elements of a floating-point vector.

**Return** none

### Parameters

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: variance value returned here

void **riscv\_var\_q15** (const q15\_t \*pSrc, uint32\_t blockSize, q15\_t \*pResult)  
 Variance of the elements of a Q15 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using a 64-bit internal accumulator. The input is represented in 1.15 format. Intermediate multiplication yields a 2.30 format, and this result is added without saturation to a 64-bit accumulator in 34.30 format. With 33 guard bits in the accumulator, there is no risk of overflow, and the full precision of the intermediate multiplication is preserved. Finally, the 34.30 result is truncated to 34.15 format by discarding the lower 15 bits, and then saturated to yield a result in 1.15 format.

### Parameters

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: variance value returned here

void **riscv\_var\_q31** (const q31\_t \*pSrc, uint32\_t blockSize, q31\_t \*pResult)  
 Variance of the elements of a Q31 vector.

**Return** none

**Scaling and Overflow Behavior** The function is implemented using an internal 64-bit accumulator. The input is represented in 1.31 format, which is then downshifted by 8 bits which yields 1.23, and intermediate multiplication yields a 2.46 format. The accumulator maintains full precision of the intermediate multiplication results, and as a consequence has only 16 guard bits. There is no saturation on intermediate additions. If the accumulator overflows it wraps around and distorts the result. In order to avoid overflows completely the input signal must be scaled down by  $\log_2(\text{blockSize}) - 8$  bits, as a total of blockSize additions are performed internally. After division, internal variables should be Q18.46. Finally, the 18.46 accumulator is right shifted by 15 bits to yield a 1.31 format value.

**Parameters**

- [in] pSrc: points to the input vector
- [in] blockSize: number of samples in input vector
- [out] pResult: variance value returned here

*group* **groupStats**

### 3.3.13 Support Functions

**Barycenter**

void **riscv\_barycenter\_f16** (**const** float16\_t \*in, **const** float16\_t \*weights, float16\_t \*out, uint32\_t nbVectors, uint32\_t vecDim)

void **riscv\_barycenter\_f32** (**const** float32\_t \*in, **const** float32\_t \*weights, float32\_t \*out, uint32\_t nbVectors, uint32\_t vecDim)

*group* **barycenter**

Barycenter of weighted vectors

**Functions**

void **riscv\_barycenter\_f16** (**const** float16\_t \*in, **const** float16\_t \*weights, float16\_t \*out, uint32\_t nbVectors, uint32\_t vecDim)

Barycenter.

**Return** None

**Parameters**

- [in] \*in: List of vectors
- [in] \*weights: Weights of the vectors
- [out] \*out: Barycenter
- [in] nbVectors: Number of vectors
- [in] vecDim: Dimension of space (vector dimension)

void **riscv\_barycenter\_f32** (**const** float32\_t \*in, **const** float32\_t \*weights, float32\_t \*out, uint32\_t nbVectors, uint32\_t vecDim)

Barycenter.

**Return** None

**Parameters**

- [in] \*in: List of vectors
- [in] \*weights: Weights of the vectors
- [out] \*out: Barycenter
- [in] nbVectors: Number of vectors
- [in] vecDim: Dimension of space (vector dimension)



## Vector sorting algorithms

```
void riscv_merge_sort_f32 (const riscv_merge_sort_instance_f32 *S, float32_t *pSrc, float32_t
                          *pDst, uint32_t blockSize)
```

```
void riscv_merge_sort_init_f32 (riscv_merge_sort_instance_f32 *S, riscv_sort_dir dir, float32_t
                              *buffer)
```

```
void riscv_sort_f32 (const riscv_sort_instance_f32 *S, float32_t *pSrc, float32_t *pDst, uint32_t block-
                    Size)
```

```
void riscv_sort_init_f32 (riscv_sort_instance_f32 *S, riscv_sort_alg alg, riscv_sort_dir dir)
```

### group **Sorting**

Sort the elements of a vector

There are separate functions for floating-point, Q31, Q15, and Q7 data types.

## Functions

```
void riscv_bitonic_sort_f32 (const riscv_sort_instance_f32 *S, float32_t *pSrc, float32_t
                             *pDst, uint32_t blockSize)
```

### Parameters

- [in] S: points to an instance of the sorting structure.
- [in] pSrc: points to the block of input data.
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process.

```
void riscv_bubble_sort_f32 (const riscv_sort_instance_f32 *S, float32_t *pSrc, float32_t
                             *pDst, uint32_t blockSize)
```

**Algorithm** The bubble sort algorithm is a simple comparison algorithm that reads the elements of a vector from the beginning to the end, compares the adjacent ones and swaps them if they are in the wrong order. The procedure is repeated until there is nothing left to swap. Bubble sort is fast for input vectors that are nearly sorted.

**It's an in-place algorithm. In order to obtain an out-of-place** function, a memcpy of the source vector is performed

### Parameters

- [in] S: points to an instance of the sorting structure.
- [in] pSrc: points to the block of input data.
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process.

```
void riscv_heap_sort_f32 (const riscv_sort_instance_f32 *S, float32_t *pSrc, float32_t *pDst,
                          uint32_t blockSize)
```

**Algorithm** The heap sort algorithm is a comparison algorithm that divides the input array into a sorted and an unsorted region, and shrinks the unsorted region by extracting the largest element and moving it to the sorted region. A heap data structure is used to find the maximum.

**It's an in-place algorithm. In order to obtain an out-of-place** function, a memcpy of the source vector is performed.

### Parameters

- [in] *S*: points to an instance of the sorting structure.
- [in] *pSrc*: points to the block of input data.
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process.

```
void riscv_insertion_sort_f32 (const riscv_sort_instance_f32 *S, float32_t *pSrc, float32_t  
                             *pDst, uint32_t blockSize)
```

**Algorithm** The insertion sort is a simple sorting algorithm that reads all the element of the input array and removes one element at a time, finds the location it belongs in the final sorted list, and inserts it there.

**It's an in-place algorithm. In order to obtain an out-of-place** function, a memcpy of the source vector is performed.

**Parameters**

- [in] *S*: points to an instance of the sorting structure.
- [in] *pSrc*: points to the block of input data.
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process.

```
void riscv_merge_sort_f32 (const riscv_merge_sort_instance_f32 *S, float32_t *pSrc, float32_t  
                           *pDst, uint32_t blockSize)
```

**Algorithm** The merge sort algorithm is a comparison algorithm that divide the input array in sublists and merge them to produce longer sorted sublists until there is only one list remaining.

**A work array is always needed. It must be allocated by the user** linked to the instance at initialization time.

**It's an in-place algorithm. In order to obtain an out-of-place** function, a memcpy of the source vector is performed

**Parameters**

- [in] *S*: points to an instance of the sorting structure.
- [in] *pSrc*: points to the block of input data.
- [out] *pDst*: points to the block of output data
- [in] *blockSize*: number of samples to process.

```
void riscv_merge_sort_init_f32 (riscv_merge_sort_instance_f32 *S, riscv_sort_dir dir,  
                                float32_t *buffer)
```

**Parameters**

- [inout] *S*: points to an instance of the sorting structure.
- [in] *dir*: Sorting order.
- [in] *buffer*: Working buffer.

```
void riscv_quick_sort_f32 (const riscv_sort_instance_f32 *S, float32_t *pSrc, float32_t *pDst,  
                           uint32_t blockSize)
```

**Algorithm** The quick sort algorithm is a comparison algorithm that divides the input array into two smaller sub-arrays and recursively sort them. An element of the array (the pivot) is chosen, all the elements with values smaller than the pivot are moved before the pivot, while all elements with values greater than the pivot are moved after it (partition).

In this implementation the Hoare partition scheme has been used [Hoare, C. A. R. (1 January 1962). “Quicksort”. The Computer Journal. 5 (1): 10...16.] The first element has always been chosen as the pivot. The partition algorithm guarantees that the returned pivot is never placed outside the vector, since it is returned only when the pointers crossed each other. In this way it isn’t possible to obtain empty partitions and infinite recursion is avoided.

It’s an in-place algorithm. In order to obtain an out-of-place function, a memcpy of the source vector is performed.

#### Parameters

- [in] S: points to an instance of the sorting structure.
- [inout] pSrc: points to the block of input data.
- [out] pDst: points to the block of output data.
- [in] blockSize: number of samples to process.

```
void riscv_selection_sort_f32 (const riscv_sort_instance_f32 *S, float32_t *pSrc, float32_t *pDst, uint32_t blockSize)
```

**Algorithm** The Selection sort algorithm is a comparison algorithm that divides the input array into a sorted and an unsorted sublist (initially the sorted sublist is empty and the unsorted sublist is the input array), looks for the smallest (or biggest) element in the unsorted sublist, swapping it with the leftmost one, and moving the sublists boundary one element to the right.

**It’s an in-place algorithm. In order to obtain an out-of-place** function, a memcpy of the source vector is performed.

#### Parameters

- [in] S: points to an instance of the sorting structure.
- [in] pSrc: points to the block of input data.
- [out] pDst: points to the block of output data
- [in] blockSize: number of samples to process.

```
void riscv_sort_f32 (const riscv_sort_instance_f32 *S, float32_t *pSrc, float32_t *pDst, uint32_t blockSize)
```

Generic sorting function.

#### Parameters

- [in] S: points to an instance of the sorting structure.
- [in] pSrc: points to the block of input data.
- [out] pDst: points to the block of output data.
- [in] blockSize: number of samples to process.

```
void riscv_sort_init_f32 (riscv_sort_instance_f32 *S, riscv_sort_alg alg, riscv_sort_dir dir)
```

#### Parameters

- [inout] S: points to an instance of the sorting structure.
- [in] alg: Selected algorithm.
- [in] dir: Sorting order.

## Vector Copy

void **riscv\_copy\_f16** (const float16\_t \*pSrc, float16\_t \*pDst, uint32\_t blockSize)

void **riscv\_copy\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_copy\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_copy\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t blockSize)

void **riscv\_copy\_q7** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

*group* **copy**

Copies sample by sample from source vector to destination vector.

There are separate functions for floating point, Q31, Q15, and Q7 data types.

## Functions

void **riscv\_copy\_f16** (const float16\_t \*pSrc, float16\_t \*pDst, uint32\_t blockSize)

Copies the elements of a f16 vector.

Copies the elements of a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_copy\_f32** (const float32\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)

Copies the elements of a floating-point vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_copy\_q15** (const q15\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

Copies the elements of a Q15 vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_copy\_q31** (const q31\_t \*pSrc, q31\_t \*pDst, uint32\_t blockSize)

Copies the elements of a Q31 vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_copy\_q7** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)  
Copies the elements of a Q7 vector.

**Return** none

**Parameters**

- [in] pSrc: points to input vector
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

### Convert 16-bit floating point value

void **riscv\_f16\_to\_float** (const float16\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_f16\_to\_q15** (const float16\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

group **f16\_to\_x**

### Functions

void **riscv\_f16\_to\_float** (const float16\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the f16 vector to f32 vector.

Converts the elements of the floating-point vector to Q31 vector.

**Return** none

**Parameters**

- [in] pSrc: points to the f16 input vector
- [out] pDst: points to the f32 output vector
- [in] blockSize: number of samples in each vector

void **riscv\_f16\_to\_q15** (const float16\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the f16 vector to Q15 vector.

Converts the elements of the floating-point vector to Q31 vector.

**Return** none

**Details** The equation used for the conversion process is:

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

**Note** In order to apply rounding in scalar version, the library should be rebuilt with the ROUNDING macro defined in the preprocessor section of project options.

**Parameters**

- [in] pSrc: points to the f16 input vector
- [out] pDst: points to the Q15 output vector
- [in] blockSize: number of samples in each vector

**Vector Fill**

void **riscv\_fill\_f16** (float16\_t *value*, float16\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_fill\_f32** (float32\_t *value*, float32\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_fill\_q15** (q15\_t *value*, q15\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_fill\_q31** (q31\_t *value*, q31\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_fill\_q7** (q7\_t *value*, q7\_t \**pDst*, uint32\_t *blockSize*)

**group Fill**

Fills the destination vector with a constant value.

There are separate functions for floating point, Q31, Q15, and Q7 data types.

**Functions**

void **riscv\_fill\_f16** (float16\_t *value*, float16\_t \**pDst*, uint32\_t *blockSize*)

Fills a constant value into a f16 vector.

Fills a constant value into a floating-point vector.

**Return** none

**Parameters**

- [in] value: input value to be filled
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_fill\_f32** (float32\_t *value*, float32\_t \**pDst*, uint32\_t *blockSize*)

Fills a constant value into a floating-point vector.

**Return** none

**Parameters**

- [in] value: input value to be filled
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_fill\_q15** (q15\_t *value*, q15\_t \**pDst*, uint32\_t *blockSize*)

Fills a constant value into a Q15 vector.

**Return** none

**Parameters**

- [in] value: input value to be filled
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_fill\_q31** (q31\_t *value*, q31\_t \**pDst*, uint32\_t *blockSize*)

Fills a constant value into a Q31 vector.

**Return** none

**Parameters**

- [in] value: input value to be filled
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

void **riscv\_fill\_q7** (q7\_t *value*, q7\_t \**pDst*, uint32\_t *blockSize*)

Fills a constant value into a Q7 vector.

**Return** none

**Parameters**

- [in] value: input value to be filled
- [out] pDst: points to output vector
- [in] blockSize: number of samples in each vector

### Convert 32-bit floating point value

void **riscv\_float\_to\_f16** (const float32\_t \**pSrc*, float16\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_float\_to\_q15** (const float32\_t \**pSrc*, q15\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_float\_to\_q31** (const float32\_t \**pSrc*, q31\_t \**pDst*, uint32\_t *blockSize*)

void **riscv\_float\_to\_q7** (const float32\_t \**pSrc*, q7\_t \**pDst*, uint32\_t *blockSize*)

*group* **float\_to\_x**

### Functions

void **riscv\_float\_to\_f16** (const float32\_t \**pSrc*, float16\_t \**pDst*, uint32\_t *blockSize*)

Converts the elements of the floating-point vector to f16 vector.

Converts the elements of the floating-point vector to Q31 vector.

**Return** none

**Parameters**

- [in] pSrc: points to the f32 input vector
- [out] pDst: points to the f16 output vector
- [in] blockSize: number of samples in each vector

void **riscv\_float\_to\_q15** (const float32\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

Converts the elements of the floating-point vector to Q15 vector.

**Return** none

**Details** The equation used for the conversion process is:

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] are saturated.

**Note** In order to apply rounding, the library should be rebuilt with the ROUNDING macro defined in the preprocessor section of project options.

**Parameters**

- [in] pSrc: points to the floating-point input vector
- [out] pDst: points to the Q15 output vector
- [in] blockSize: number of samples in each vector

void **riscv\_float\_to\_q31** (const float32\_t \*pSrc, q31\_t \*pDst, uint32\_t blockSize)

Converts the elements of the floating-point vector to Q31 vector.

**Return** none

**Details** The equation used for the conversion process is:

**Scaling and Overflow Behavior** The function uses saturating arithmetic. Results outside of the allowable Q31 range [0x80000000 0x7FFFFFFF] are saturated.

**Note** In order to apply rounding, the library should be rebuilt with the ROUNDING macro defined in the preprocessor section of project options.

**Parameters**

- [in] pSrc: points to the floating-point input vector
- [out] pDst: points to the Q31 output vector
- [in] blockSize: number of samples in each vector

void **riscv\_float\_to\_q7** (const float32\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

Converts the elements of the floating-point vector to Q7 vector.

**Return** none.

**Description:**

The equation used for the conversion process is:

**Scaling and Overflow Behavior:**

The function uses saturating arithmetic. Results outside of the allowable Q7 range [0x80 0x7F] will be saturated.

**Note** In order to apply rounding, the library should be rebuilt with the ROUNDING macro defined in the preprocessor section of project options.

**Parameters**

- [in] \*pSrc: points to the floating-point input vector
- [out] \*pDst: points to the Q7 output vector



- [in] blockSize: length of the input vector

### Convert 16-bit Integer value

```
void riscv_q15_to_f16 (const q15_t *pSrc, float16_t *pDst, uint32_t blockSize)
void riscv_q15_to_float (const q15_t *pSrc, float32_t *pDst, uint32_t blockSize)
void riscv_q15_to_q31 (const q15_t *pSrc, q31_t *pDst, uint32_t blockSize)
void riscv_q15_to_q7 (const q15_t *pSrc, q7_t *pDst, uint32_t blockSize)
group q15_to_x
```

### Functions

void **riscv\_q15\_to\_f16** (const q15\_t \*pSrc, float16\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the Q15 vector to f16 vector.

Converts the elements of the floating-point vector to Q31 vector.

**Return** none

**Details** The equation used for the conversion process is:

#### Parameters

- [in] pSrc: points to the Q15 input vector
- [out] pDst: points to the f16 output vector
- [in] blockSize: number of samples in each vector

void **riscv\_q15\_to\_float** (const q15\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the Q15 vector to floating-point vector.

**Return** none

**Details** The equation used for the conversion process is:

#### Parameters

- [in] pSrc: points to the Q15 input vector
- [out] pDst: points to the floating-point output vector
- [in] blockSize: number of samples in each vector

void **riscv\_q15\_to\_q31** (const q15\_t \*pSrc, q31\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the Q15 vector to Q31 vector.

**Return** none

**Details** The equation used for the conversion process is:

#### Parameters

- [in] pSrc: points to the Q15 input vector
- [out] pDst: points to the Q31 output vector
- [in] blockSize: number of samples in each vector

void **riscv\_q15\_to\_q7** (**const** q15\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the Q15 vector to Q7 vector.

**Return** none

**Details** The equation used for the conversion process is:

**Parameters**

- [in] pSrc: points to the Q15 input vector
- [out] pDst: points to the Q7 output vector
- [in] blockSize: number of samples in each vector

### Convert 32-bit Integer value

void **riscv\_q31\_to\_float** (**const** q31\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)

void **riscv\_q31\_to\_q15** (**const** q31\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_q31\_to\_q7** (**const** q31\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

group **q31\_to\_x**

### Functions

void **riscv\_q31\_to\_float** (**const** q31\_t \*pSrc, float32\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the Q31 vector to floating-point vector.

**Return** none

**Details** The equation used for the conversion process is:

**Parameters**

- [in] pSrc: points to the Q31 input vector
- [out] pDst: points to the floating-point output vector
- [in] blockSize: number of samples in each vector

void **riscv\_q31\_to\_q15** (**const** q31\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the Q31 vector to Q15 vector.

**Return** none

**Details** The equation used for the conversion process is:

**Parameters**

- [in] pSrc: points to the Q31 input vector
- [out] pDst: points to the Q15 output vector
- [in] blockSize: number of samples in each vector

void **riscv\_q31\_to\_q7** (**const** q31\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)  
Converts the elements of the Q31 vector to Q7 vector.

**Return** none

**Details** The equation used for the conversion process is:

**Parameters**

- [in] pSrc: points to the Q31 input vector
- [out] pDst: points to the Q7 output vector
- [in] blockSize: number of samples in each vector

### Convert 8-bit Integer value

```
void riscv_q7_to_float (const q7_t *pSrc, float32_t *pDst, uint32_t blockSize)
```

```
void riscv_q7_to_q15 (const q7_t *pSrc, q15_t *pDst, uint32_t blockSize)
```

```
void riscv_q7_to_q31 (const q7_t *pSrc, q31_t *pDst, uint32_t blockSize)
```

group q7\_to\_x

### Functions

```
void riscv_q7_to_float (const q7_t *pSrc, float32_t *pDst, uint32_t blockSize)
```

Converts the elements of the Q7 vector to floating-point vector.

**Return** none

**Details** The equation used for the conversion process is:

**Parameters**

- [in] pSrc: points to the Q7 input vector
- [out] pDst: points to the floating-point output vector
- [in] blockSize: number of samples in each vector

```
void riscv_q7_to_q15 (const q7_t *pSrc, q15_t *pDst, uint32_t blockSize)
```

Converts the elements of the Q7 vector to Q15 vector.

**Return** none

**Details** The equation used for the conversion process is:

**Parameters**

- [in] pSrc: points to the Q7 input vector
- [out] pDst: points to the Q15 output vector
- [in] blockSize: number of samples in each vector

```
void riscv_q7_to_q31 (const q7_t *pSrc, q31_t *pDst, uint32_t blockSize)
```

Converts the elements of the Q7 vector to Q31 vector.

**Return** none

**Details** The equation used for the conversion process is:

**Parameters**

- [in] pSrc: points to the Q7 input vector

- [out] `pDst`: points to the Q31 output vector
- [in] `blockSize`: number of samples in each vector

## Weighted Sum

`float16_t riscv_weighted_sum_f16 (const float16_t *in, const float16_t *weights, uint32_t blockSize)`

`float32_t riscv_weighted_sum_f32 (const float32_t *in, const float32_t *weights, uint32_t blockSize)`

*group* **weightedsum**  
Weighted sum of values

## Functions

`float16_t riscv_weighted_sum_f16 (const float16_t *in, const float16_t *weights, uint32_t blockSize)`

Weighted sum.

**Return** Weighted sum

### Parameters

- [in] `*in`: Array of input values.
- [in] `*weights`: Weights
- [in] `blockSize`: Number of samples in the input array.

`float32_t riscv_weighted_sum_f32 (const float32_t *in, const float32_t *weights, uint32_t blockSize)`

Weighted sum.

**Return** Weighted sum

### Parameters

- [in] `*in`: Array of input values.
- [in] `*weights`: Weights
- [in] `blockSize`: Number of samples in the input array.

*group* **groupSupport**

## 3.3.14 SVM Functions

### Linear SVM

`void riscv_svm_linear_init_f16 (riscv_svm_linear_instance_f16 *S, uint32_t nbOfSupportVectors, uint32_t vectorDimension, float16_t intercept, const float16_t *dualCoefficients, const float16_t *supportVectors, const int32_t *classes)`

`void riscv_svm_linear_init_f32 (riscv_svm_linear_instance_f32 *S, uint32_t nbOfSupportVectors, uint32_t vectorDimension, float32_t intercept, const float32_t *dualCoefficients, const float32_t *supportVectors, const int32_t *classes)`

```
void riscv_svm_linear_predict_f16 (const riscv_svm_linear_instance_f16 *S, const float16_t
                                     *in, int32_t *pResult)
```

```
void riscv_svm_linear_predict_f32 (const riscv_svm_linear_instance_f32 *S, const float32_t
                                     *in, int32_t *pResult)
```

*group* **linearsvm**

Linear SVM classifier

## Functions

```
void riscv_svm_linear_init_f16 (riscv_svm_linear_instance_f16 *S, uint32_t nbOfSupportVec-
                                tors, uint32_t vectorDimension, float16_t intercept, const
                                float16_t *dualCoefficients, const float16_t *supportVec-
                                tors, const int32_t *classes)
```

SVM linear instance init function.

Classes are integer used as output of the function (instead of having -1,1 as class values).

**Return** none.

### Parameters

- [in] S: Parameters for the SVM function
- [in] nbOfSupportVectors: Number of support vectors
- [in] vectorDimension: Dimension of vector space
- [in] intercept: Intercept
- [in] dualCoefficients: Array of dual coefficients
- [in] supportVectors: Array of support vectors
- [in] classes: Array of 2 classes ID

```
void riscv_svm_linear_init_f32 (riscv_svm_linear_instance_f32 *S, uint32_t nbOfSupportVec-
                                tors, uint32_t vectorDimension, float32_t intercept, const
                                float32_t *dualCoefficients, const float32_t *supportVec-
                                tors, const int32_t *classes)
```

SVM linear instance init function.

Classes are integer used as output of the function (instead of having -1,1 as class values).

**Return** none.

### Parameters

- [in] S: Parameters for the SVM function
- [in] nbOfSupportVectors: Number of support vectors
- [in] vectorDimension: Dimension of vector space
- [in] intercept: Intercept
- [in] dualCoefficients: Array of dual coefficients
- [in] supportVectors: Array of support vectors
- [in] classes: Array of 2 classes ID

```
void riscv_svm_linear_predict_f16(const riscv_svm_linear_instance_f16 *S, const
                                float16_t *in, int32_t *pResult)
```

SVM linear prediction.

**Return** none.

**Parameters**

- [in] S: Pointer to an instance of the linear SVM structure.
- [in] in: Pointer to input vector
- [out] pResult: Decision value

```
void riscv_svm_linear_predict_f32(const riscv_svm_linear_instance_f32 *S, const
                                float32_t *in, int32_t *pResult)
```

SVM linear prediction.

**Return** none.

**Parameters**

- [in] S: Pointer to an instance of the linear SVM structure.
- [in] in: Pointer to input vector
- [out] pResult: Decision value

## Polynomial SVM

```
void riscv_svm_polynomial_init_f16(riscv_svm_polynomial_instance_f16 *S, uint32_t nbOfSup-
                                portVectors, uint32_t vectorDimension, float16_t inter-
                                cept, const float16_t *dualCoefficients, const float16_t
                                *supportVectors, const int32_t *classes, int32_t degree,
                                float16_t coef0, float16_t gamma)
```

```
void riscv_svm_polynomial_init_f32(riscv_svm_polynomial_instance_f32 *S, uint32_t nbOfSup-
                                portVectors, uint32_t vectorDimension, float32_t inter-
                                cept, const float32_t *dualCoefficients, const float32_t
                                *supportVectors, const int32_t *classes, int32_t degree,
                                float32_t coef0, float32_t gamma)
```

```
void riscv_svm_polynomial_predict_f16(const riscv_svm_polynomial_instance_f16 *S, const
                                float16_t *in, int32_t *pResult)
```

```
void riscv_svm_polynomial_predict_f32(const riscv_svm_polynomial_instance_f32 *S, const
                                float32_t *in, int32_t *pResult)
```

group **polysvm**

Polynomial SVM classifier

## Functions

```
void riscv_svm_polynomial_init_f16(riscv_svm_polynomial_instance_f16 *S, uint32_t nbOf-
                                SupportVectors, uint32_t vectorDimension, float16_t in-
                                tercept, const float16_t *dualCoefficients, const
                                float16_t *supportVectors, const int32_t *classes,
                                int32_t degree, float16_t coef0, float16_t gamma)
```

SVM polynomial instance init function.

Classes are integer used as output of the function (instead of having -1,1 as class values).

**Return** none.

**Parameters**

- [in] *S*: points to an instance of the polynomial SVM structure.
- [in] *nbOfSupportVectors*: Number of support vectors
- [in] *vectorDimension*: Dimension of vector space
- [in] *intercept*: Intercept
- [in] *dualCoefficients*: Array of dual coefficients
- [in] *supportVectors*: Array of support vectors
- [in] *classes*: Array of 2 classes ID
- [in] *degree*: Polynomial degree
- [in] *coef0*: *coeff0* (scikit-learn terminology)
- [in] *gamma*: *gamma* (scikit-learn terminology)

```
void riscv_svm_polynomial_init_f32(riscv_svm_polynomial_instance_f32 *S, uint32_t nbOf-
    SupportVectors, uint32_t vectorDimension, float32_t in-
    tercept, const float32_t *dualCoefficients, const
    float32_t *supportVectors, const int32_t *classes,
    int32_t degree, float32_t coef0, float32_t gamma)
```

SVM polynomial instance init function.

Classes are integer used as output of the function (instead of having -1,1 as class values).

**Return** none.

**Parameters**

- [in] *S*: points to an instance of the polynomial SVM structure.
- [in] *nbOfSupportVectors*: Number of support vectors
- [in] *vectorDimension*: Dimension of vector space
- [in] *intercept*: Intercept
- [in] *dualCoefficients*: Array of dual coefficients
- [in] *supportVectors*: Array of support vectors
- [in] *classes*: Array of 2 classes ID
- [in] *degree*: Polynomial degree
- [in] *coef0*: *coeff0* (scikit-learn terminology)
- [in] *gamma*: *gamma* (scikit-learn terminology)

```
void riscv_svm_polynomial_predict_f16(const riscv_svm_polynomial_instance_f16 *S,
    const float16_t *in, int32_t *pResult)
```

SVM polynomial prediction.

**Return** none.

**Parameters**

- [in] *S*: Pointer to an instance of the polynomial SVM structure.
- [in] *in*: Pointer to input vector
- [out] *pResult*: Decision value

void **riscv\_svm\_polynomial\_predict\_f32** (**const** riscv\_svm\_polynomial\_instance\_f32 \**S*,  
  **const** float32\_t \**in*, int32\_t \**pResult*)  
SVM polynomial prediction.

**Return** none.

**Parameters**

- [in] *S*: Pointer to an instance of the polynomial SVM structure.
- [in] *in*: Pointer to input vector
- [out] *pResult*: Decision value

## RBF SVM

void **riscv\_svm\_rbf\_init\_f16** (riscv\_svm\_rbf\_instance\_f16 \**S*, uint32\_t *nbOfSupportVectors*, uint32\_t  
                                  *vectorDimension*, float16\_t *intercept*, **const** float16\_t \**dualCoeffi-*  
                                  *cients*, **const** float16\_t \**supportVectors*, **const** int32\_t \**classes*,  
                                  float16\_t *gamma*)

void **riscv\_svm\_rbf\_init\_f32** (riscv\_svm\_rbf\_instance\_f32 \**S*, uint32\_t *nbOfSupportVectors*, uint32\_t  
                                  *vectorDimension*, float32\_t *intercept*, **const** float32\_t \**dualCoeffi-*  
                                  *cients*, **const** float32\_t \**supportVectors*, **const** int32\_t \**classes*,  
                                  float32\_t *gamma*)

void **riscv\_svm\_rbf\_predict\_f16** (**const** riscv\_svm\_rbf\_instance\_f16 \**S*, **const** float16\_t \**in*,  
                                  int32\_t \**pResult*)

void **riscv\_svm\_rbf\_predict\_f32** (**const** riscv\_svm\_rbf\_instance\_f32 \**S*, **const** float32\_t \**in*,  
                                  int32\_t \**pResult*)

group **rbfsvm**  
RBF SVM classifier

## Functions

void **riscv\_svm\_rbf\_init\_f16** (riscv\_svm\_rbf\_instance\_f16 \**S*, uint32\_t *nbOfSupportVectors*,  
                                  uint32\_t *vectorDimension*, float16\_t *intercept*, **const** float16\_t  
                                  \**dualCoefficients*, **const** float16\_t \**supportVectors*, **const**  
                                  int32\_t \**classes*, float16\_t *gamma*)

SVM radial basis function instance init function.

Classes are integer used as output of the function (instead of having -1,1 as class values).

**Return** none.

**Parameters**

- [in] *S*: points to an instance of the polynomial SVM structure.
- [in] *nbOfSupportVectors*: Number of support vectors
- [in] *vectorDimension*: Dimension of vector space
- [in] *intercept*: Intercept



- [in] `dualCoefficients`: Array of dual coefficients
- [in] `supportVectors`: Array of support vectors
- [in] `classes`: Array of 2 classes ID
- [in] `gamma`: gamma (scikit-learn terminology)

void **riscv\_svm\_rbf\_init\_f32** (riscv\_svm\_rbf\_instance\_f32 \*S, uint32\_t *nbOfSupportVectors*, uint32\_t *vectorDimension*, float32\_t *intercept*, **const** float32\_t \**dualCoefficients*, **const** float32\_t \**supportVectors*, **const** int32\_t \**classes*, float32\_t *gamma*)

SVM radial basis function instance init function.

Classes are integer used as output of the function (instead of having -1,1 as class values).

**Return** none.

#### Parameters

- [in] `S`: points to an instance of the polynomial SVM structure.
- [in] `nbOfSupportVectors`: Number of support vectors
- [in] `vectorDimension`: Dimension of vector space
- [in] `intercept`: Intercept
- [in] `dualCoefficients`: Array of dual coefficients
- [in] `supportVectors`: Array of support vectors
- [in] `classes`: Array of 2 classes ID
- [in] `gamma`: gamma (scikit-learn terminology)

void **riscv\_svm\_rbf\_predict\_f16** (**const** riscv\_svm\_rbf\_instance\_f16 \*S, **const** float16\_t \**in*, int32\_t \**pResult*)

SVM rbf prediction.

**Return** none.

#### Parameters

- [in] `S`: Pointer to an instance of the rbf SVM structure.
- [in] `in`: Pointer to input vector
- [out] `pResult`: decision value

void **riscv\_svm\_rbf\_predict\_f32** (**const** riscv\_svm\_rbf\_instance\_f32 \*S, **const** float32\_t \**in*, int32\_t \**pResult*)

SVM rbf prediction.

**Return** none.

#### Parameters

- [in] `S`: Pointer to an instance of the rbf SVM structure.
- [in] `in`: Pointer to input vector
- [out] `pResult`: decision value

## Sigmoid SVM

```
void riscv_svm_sigmoid_init_f16 (riscv_svm_sigmoid_instance_f16 *S, uint32_t nbOfSupportVec-  
                                tors, uint32_t vectorDimension, float16_t intercept, const  
                                float16_t *dualCoefficients, const float16_t *supportVectors,  
                                const int32_t *classes, float16_t coef0, float16_t gamma)  
  
void riscv_svm_sigmoid_init_f32 (riscv_svm_sigmoid_instance_f32 *S, uint32_t nbOfSupportVec-  
                                tors, uint32_t vectorDimension, float32_t intercept, const  
                                float32_t *dualCoefficients, const float32_t *supportVectors,  
                                const int32_t *classes, float32_t coef0, float32_t gamma)  
  
void riscv_svm_sigmoid_predict_f16 (const riscv_svm_sigmoid_instance_f16 *S, const  
                                     float16_t *in, int32_t *pResult)  
  
void riscv_svm_sigmoid_predict_f32 (const riscv_svm_sigmoid_instance_f32 *S, const  
                                     float32_t *in, int32_t *pResult)  
  
group sigmoidsvm  
      Sigmoid SVM classifier
```

## Functions

```
void riscv_svm_sigmoid_init_f16 (riscv_svm_sigmoid_instance_f16 *S, uint32_t nbOfSup-  
                                portVectors, uint32_t vectorDimension, float16_t inter-  
                                cept, const float16_t *dualCoefficients, const float16_t  
                                *supportVectors, const int32_t *classes, float16_t coef0,  
                                float16_t gamma)
```

SVM sigmoid instance init function.

Classes are integer used as output of the function (instead of having -1,1 as class values).

**Return** none.

### Parameters

- [in] *S*: points to an instance of the rbf SVM structure.
- [in] *nbOfSupportVectors*: Number of support vectors
- [in] *vectorDimension*: Dimension of vector space
- [in] *intercept*: Intercept
- [in] *dualCoefficients*: Array of dual coefficients
- [in] *supportVectors*: Array of support vectors
- [in] *classes*: Array of 2 classes ID
- [in] *coef0*: *coeff0* (scikit-learn terminology)
- [in] *gamma*: *gamma* (scikit-learn terminology)

```
void riscv_svm_sigmoid_init_f32 (riscv_svm_sigmoid_instance_f32 *S, uint32_t nbOfSup-  
                                portVectors, uint32_t vectorDimension, float32_t inter-  
                                cept, const float32_t *dualCoefficients, const float32_t  
                                *supportVectors, const int32_t *classes, float32_t coef0,  
                                float32_t gamma)
```

SVM sigmoid instance init function.

Classes are integer used as output of the function (instead of having -1,1 as class values).

**Return** none.

**Parameters**

- [in] *S*: points to an instance of the rbf SVM structure.
- [in] *nbOfSupportVectors*: Number of support vectors
- [in] *vectorDimension*: Dimension of vector space
- [in] *intercept*: Intercept
- [in] *dualCoefficients*: Array of dual coefficients
- [in] *supportVectors*: Array of support vectors
- [in] *classes*: Array of 2 classes ID
- [in] *coef0*: *coeff0* (scikit-learn terminology)
- [in] *gamma*: *gamma* (scikit-learn terminology)

void **riscv\_svm\_sigmoid\_predict\_f16**(const riscv\_svm\_sigmoid\_instance\_f16 \**S*, const float16\_t \**in*, int32\_t \**pResult*)

SVM sigmoid prediction.

**Return** none.

**Parameters**

- [in] *S*: Pointer to an instance of the rbf SVM structure.
- [in] *in*: Pointer to input vector
- [out] *pResult*: Decision value

void **riscv\_svm\_sigmoid\_predict\_f32**(const riscv\_svm\_sigmoid\_instance\_f32 \**S*, const float32\_t \**in*, int32\_t \**pResult*)

SVM sigmoid prediction.

**Return** none.

**Parameters**

- [in] *S*: Pointer to an instance of the rbf SVM structure.
- [in] *in*: Pointer to input vector
- [out] *pResult*: Decision value

*group* **groupSVM**

This set of functions is implementing SVM classification on 2 classes. The training must be done from scikit-learn. The parameters can be easily generated from the scikit-learn object. Some examples are given in DSP/Testing/PatternGeneration/SVM.py

If more than 2 classes are needed, the functions in this folder will have to be used, as building blocks, to do multi-class classification.

No multi-class classification is provided in this SVM folder.

### 3.3.15 Transform Functions

#### Complex FFT Functions

##### Complex FFT Tables

```
const uint16_t riscvBitRevTable[1024]
const uint64_t twiddleCoefF64_16[32]
const uint64_t twiddleCoefF64_32[64]
const uint64_t twiddleCoefF64_64[128]
const uint64_t twiddleCoefF64_128[256]
const uint64_t twiddleCoefF64_256[512]
const uint64_t twiddleCoefF64_512[1024]
const uint64_t twiddleCoefF64_1024[2048]
const uint64_t twiddleCoefF64_2048[4096]
const uint64_t twiddleCoefF64_4096[8192]
const float32_t twiddleCoef_16[32]
const float32_t twiddleCoef_32[64]
const float32_t twiddleCoef_64[128]
const float32_t twiddleCoef_128[256]
const float32_t twiddleCoef_256[512]
const float32_t twiddleCoef_512[1024]
const float32_t twiddleCoef_1024[2048]
const float32_t twiddleCoef_2048[4096]
const float32_t twiddleCoef_4096[8192]
const q31_t twiddleCoef_16_q31[24]
const q31_t twiddleCoef_32_q31[48]
const q31_t twiddleCoef_64_q31[96]
const q31_t twiddleCoef_128_q31[192]
const q31_t twiddleCoef_256_q31[384]
const q31_t twiddleCoef_512_q31[768]
const q31_t twiddleCoef_1024_q31[1536]
const q31_t twiddleCoef_2048_q31[3072]
const q31_t twiddleCoef_4096_q31[6144]
const q15_t twiddleCoef_16_q15[24]
const q15_t twiddleCoef_32_q15[48]
const q15_t twiddleCoef_64_q15[96]
const q15_t twiddleCoef_128_q15[192]
```

```

const q15_t twiddleCoef_256_q15[384]
const q15_t twiddleCoef_512_q15[768]
const q15_t twiddleCoef_1024_q15[1536]
const q15_t twiddleCoef_2048_q15[3072]
const q15_t twiddleCoef_4096_q15[6144]
const float16_t twiddleCoefF16_16[32]
const float16_t twiddleCoefF16_32[64]
const float16_t twiddleCoefF16_64[128]
const float16_t twiddleCoefF16_128[256]
const float16_t twiddleCoefF16_256[512]
const float16_t twiddleCoefF16_512[1024]
const float16_t twiddleCoefF16_1024[2048]
const float16_t twiddleCoefF16_2048[4096]
const float16_t twiddleCoefF16_4096[8192]
const float16_t twiddleCoefF16_rfft_32[32]
const float16_t twiddleCoefF16_rfft_64[64]
const float16_t twiddleCoefF16_rfft_128[128]
const float16_t twiddleCoefF16_rfft_256[256]
const float16_t twiddleCoefF16_rfft_512[512]
const float16_t twiddleCoefF16_rfft_1024[1024]
const float16_t twiddleCoefF16_rfft_2048[2048]
const float16_t twiddleCoefF16_rfft_4096[4096]
group CFFT_CIFFT

```

## Variables

```
const uint16_t riscvBitRevTable[1024]
```

Table for bit reversal process.

Pseudo code for Generation of Bit reversal Table is

where  $N = 4096$ ,  $\log N = 12$

$N$  is the maximum FFT Size supported

```
const uint64_t twiddleCoefF64_16[32]
```

Double Precision Floating-point Twiddle factors Table Generation.

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 16$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const uint64_t twiddleCoeff64_32[64]
```

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 32$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const uint64_t twiddleCoeff64_64[128]
```

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 64$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const uint64_t twiddleCoeff64_128[256]
```

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 128$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const uint64_t twiddleCoeff64_256[512]
```

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 256$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const uint64_t twiddleCoeff64_512[1024]
```

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 512$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const uint64_t twiddleCoeff64_1024[2048]
```

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 1024$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const uint64_t twiddleCoeff64_2048[4096]
```

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 2048$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const uint64_t twiddleCoefF64_4096[8192]
```

Example code for Double Precision Floating-point Twiddle factors Generation:

where  $N = 4096$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_16[32]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 16$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_32[64]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 32$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_64[128]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 64$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_128[256]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 128$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_256[512]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 256$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_512[1024]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 512$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_1024[2048]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 1024$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_2048[4096]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 2048$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float32_t twiddleCoef_4096[8192]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 4096$ ,  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const q31_t twiddleCoef_16_q31[24]
```

Q31 Twiddle factors Table.

Example code for Q31 Twiddle factors Generation::

where  $N = 16$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q31_t twiddleCoef_32_q31[48]
```

Example code for Q31 Twiddle factors Generation::

where  $N = 32$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q31_t twiddleCoef_64_q31[96]
```

Example code for Q31 Twiddle factors Generation::

where  $N = 64$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q31_t twiddleCoef_128_q31[192]
```

Example code for Q31 Twiddle factors Generation::



where  $N = 128$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q31_t twiddleCoef_256_q31[384]
```

Example code for Q31 Twiddle factors Generation::

where  $N = 256$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q31_t twiddleCoef_512_q31[768]
```

Example code for Q31 Twiddle factors Generation::

where  $N = 512$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q31_t twiddleCoef_1024_q31[1536]
```

Example code for Q31 Twiddle factors Generation::

where  $N = 1024$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q31_t twiddleCoef_2048_q31[3072]
```

Example code for Q31 Twiddle factors Generation::

where  $N = 2048$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q31_t twiddleCoef_4096_q31[6144]
```

Example code for Q31 Twiddle factors Generation::

where  $N = 4096$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to Q31(Fixed point 1.31):  $\text{round}(\text{twiddleCoefQ31}(i) * \text{pow}(2, 31))$

```
const q15_t twiddleCoef_16_q15[24]
```

q15 Twiddle factors Table

Example code for q15 Twiddle factors Generation::

where  $N = 16$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const q15\_t twiddleCoef\_32\_q15[48]**

Example code for q15 Twiddle factors Generation::

where  $N = 32$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const q15\_t twiddleCoef\_64\_q15[96]**

Example code for q15 Twiddle factors Generation::

where  $N = 64$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const q15\_t twiddleCoef\_128\_q15[192]**

Example code for q15 Twiddle factors Generation::

where  $N = 128$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const q15\_t twiddleCoef\_256\_q15[384]**

Example code for q15 Twiddle factors Generation::

where  $N = 256$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const q15\_t twiddleCoef\_512\_q15[768]**

Example code for q15 Twiddle factors Generation::

where  $N = 512$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const q15\_t twiddleCoef\_1024\_q15[1536]**

Example code for q15 Twiddle factors Generation::

where  $N = 1024$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const** q15\_t **twiddleCoef\_2048\_q15**[3072]

Example code for q15 Twiddle factors Generation::

where  $N = 2048$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const** q15\_t **twiddleCoef\_4096\_q15**[6144]

Example code for q15 Twiddle factors Generation::

where  $N = 4096$ ,  $PI = 3.14159265358979$

Cos and Sin values are interleaved fashion

Convert Floating point to q15(Fixed point 1.15):  $\text{round}(\text{twiddleCoefq15}(i) * \text{pow}(2, 15))$

**const** float16\_t **twiddleCoefF16\_16**[32]

Floating-point Twiddle factors Table Generation.

Example code for Floating-point Twiddle factors Generation:

where  $N = 16$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

**const** float16\_t **twiddleCoefF16\_32**[64]

Example code for Floating-point Twiddle factors Generation:

where  $N = 32$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

**const** float16\_t **twiddleCoefF16\_64**[128]

Example code for Floating-point Twiddle factors Generation:

where  $N = 64$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

**const** float16\_t **twiddleCoefF16\_128**[256]

Example code for Floating-point Twiddle factors Generation:

where  $N = 128$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float16_t twiddleCoefF16_256[512]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 256$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float16_t twiddleCoefF16_512[1024]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 512$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float16_t twiddleCoefF16_1024[2048]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 1024$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float16_t twiddleCoefF16_2048[4096]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 2048$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float16_t twiddleCoefF16_4096[8192]
```

Example code for Floating-point Twiddle factors Generation:

where  $N = 4096$  and  $PI = 3.14159265358979$

Cos and Sin values are in interleaved fashion

```
const float16_t twiddleCoefF16_rfft_32[32]
```

Example code for Floating-point RFFT Twiddle factors Generation:

Real and Imag values are in interleaved fashion

```
const float16_t twiddleCoefF16_rfft_64[64]
```

```
const float16_t twiddleCoefF16_rfft_128[128]
```

```
const float16_t twiddleCoefF16_rfft_256[256]
```

```
const float16_t twiddleCoefF16_rfft_512[512]
```

```
const float16_t twiddleCoefF16_rfft_1024[1024]
```

```
const float16_t twiddleCoefF16_rfft_2048[2048]
```

```

    const float16_t twiddleCoeffF16_rfft_4096[4096]

void riscv_cfft_f16 (const riscv_cfft_instance_f16 *S, float16_t *p1, uint8_t ifftFlag, uint8_t bitReverseFlag)

void riscv_cfft_f32 (const riscv_cfft_instance_f32 *S, float32_t *p1, uint8_t ifftFlag, uint8_t bitReverseFlag)

void riscv_cfft_f64 (const riscv_cfft_instance_f64 *S, float64_t *p1, uint8_t ifftFlag, uint8_t bitReverseFlag)

riscv_status riscv_cfft_init_f16 (riscv_cfft_instance_f16 *S, uint16_t fftLen)

riscv_status riscv_cfft_init_f32 (riscv_cfft_instance_f32 *S, uint16_t fftLen)

riscv_status riscv_cfft_init_f64 (riscv_cfft_instance_f64 *S, uint16_t fftLen)

riscv_status riscv_cfft_init_q15 (riscv_cfft_instance_q15 *S, uint16_t fftLen)

riscv_status riscv_cfft_init_q31 (riscv_cfft_instance_q31 *S, uint16_t fftLen)

void riscv_cfft_q15 (const riscv_cfft_instance_q15 *S, q15_t *p1, uint8_t ifftFlag, uint8_t bitReverseFlag)

void riscv_cfft_q31 (const riscv_cfft_instance_q31 *S, q31_t *p1, uint8_t ifftFlag, uint8_t bitReverseFlag)

void riscv_cfft_radix2_f16 (const riscv_cfft_radix2_instance_f16 *S, float16_t *pSrc)

void riscv_cfft_radix2_f32 (const riscv_cfft_radix2_instance_f32 *S, float32_t *pSrc)

riscv_status riscv_cfft_radix2_init_f16 (riscv_cfft_radix2_instance_f16 *S, uint16_t fftLen, uint8_t ifftFlag, uint8_t bitReverseFlag)

riscv_status riscv_cfft_radix2_init_f32 (riscv_cfft_radix2_instance_f32 *S, uint16_t fftLen, uint8_t ifftFlag, uint8_t bitReverseFlag)

riscv_status riscv_cfft_radix2_init_q15 (riscv_cfft_radix2_instance_q15 *S, uint16_t fftLen, uint8_t ifftFlag, uint8_t bitReverseFlag)

riscv_status riscv_cfft_radix2_init_q31 (riscv_cfft_radix2_instance_q31 *S, uint16_t fftLen, uint8_t ifftFlag, uint8_t bitReverseFlag)

void riscv_cfft_radix2_q15 (const riscv_cfft_radix2_instance_q15 *S, q15_t *pSrc)

void riscv_cfft_radix2_q31 (const riscv_cfft_radix2_instance_q31 *S, q31_t *pSrc)

void riscv_cfft_radix4by2_f16 (float16_t *pSrc, uint32_t fftLen, const float16_t *pCoef)

void riscv_cfft_radix4_f16 (const riscv_cfft_radix4_instance_f16 *S, float16_t *pSrc)

void riscv_cfft_radix4_f32 (const riscv_cfft_radix4_instance_f32 *S, float32_t *pSrc)

riscv_status riscv_cfft_radix4_init_f16 (riscv_cfft_radix4_instance_f16 *S, uint16_t fftLen, uint8_t ifftFlag, uint8_t bitReverseFlag)

riscv_status riscv_cfft_radix4_init_f32 (riscv_cfft_radix4_instance_f32 *S, uint16_t fftLen, uint8_t ifftFlag, uint8_t bitReverseFlag)

riscv_status riscv_cfft_radix4_init_q15 (riscv_cfft_radix4_instance_q15 *S, uint16_t fftLen, uint8_t ifftFlag, uint8_t bitReverseFlag)

riscv_status riscv_cfft_radix4_init_q31 (riscv_cfft_radix4_instance_q31 *S, uint16_t fftLen, uint8_t ifftFlag, uint8_t bitReverseFlag)

void riscv_cfft_radix4_q15 (const riscv_cfft_radix4_instance_q15 *S, q15_t *pSrc)

void riscv_cfft_radix4_q31 (const riscv_cfft_radix4_instance_q31 *S, q31_t *pSrc)

```

*group* **ComplexFFT**

The Fast Fourier Transform (FFT) is an efficient algorithm for computing the Discrete Fourier Transform (DFT). The FFT can be orders of magnitude faster than the DFT, especially for long lengths. The algorithms described in this section operate on complex data. A separate set of functions is devoted to handling of real sequences.

There are separate algorithms for handling floating-point, Q15, and Q31 data types. The algorithms available for each data type are described next.

The FFT functions operate in-place. That is, the array holding the input data will also be used to hold the corresponding result. The input data is complex and contains  $2 \times \text{fftLen}$  interleaved values as shown below. The FFT result will be contained in the same array and the frequency domain values will have the same interleaving.

**Floating-point** The floating-point complex FFT uses a mixed-radix algorithm. Multiple radix-8 stages are performed along with a single radix-2 or radix-4 stage, as needed. The algorithm supports lengths of [16, 32, 64, ..., 4096] and each length uses a different twiddle factor table.

The function uses the standard FFT definition and output values may grow by a factor of  $\text{fftLen}$  when computing the forward transform. The inverse transform includes a scale of  $1/\text{fftLen}$  as part of the calculation and this matches the textbook definition of the inverse FFT.

For the MVE version, the new `riscv_cfft_init_f32` initialization function is **mandatory**. **Compilation flags are available to include only the required tables for the needed FFTs**. Other FFT versions can continue to be initialized as explained below.

For not MVE versions, pre-initialized data structures containing twiddle factors and bit reversal tables are provided and defined in `riscv_const_structs.h`. Include this header in your function and then pass one of the constant structures as an argument to `riscv_cfft_f32`. For example:

```
riscv_cfft_f32(riscv_cfft_sR_f32_len64, pSrc, 1, 1)
```

computes a 64-point inverse complex FFT including bit reversal. The data structures are treated as constant data and not modified during the calculation. The same data structure can be reused for multiple transforms including mixing forward and inverse transforms.

Earlier releases of the library provided separate radix-2 and radix-4 algorithms that operated on floating-point data. These functions are still provided but are deprecated. The older functions are slower and less general than the new functions.

An example of initialization of the constants for the `riscv_cfft_f32` function follows:

```
const static riscv_cfft_instance_f32 *S;
...
switch (length) {
    case 16:
        S = &riscv_cfft_sR_f32_len16;
        break;
    case 32:
        S = &riscv_cfft_sR_f32_len32;
        break;
    case 64:
        S = &riscv_cfft_sR_f32_len64;
        break;
    case 128:
        S = &riscv_cfft_sR_f32_len128;
        break;
    case 256:
        S = &riscv_cfft_sR_f32_len256;
        break;
```

(continues on next page)

(continued from previous page)

```

case 512:
    S = &riscv_cfft_sR_f32_len512;
    break;
case 1024:
    S = &riscv_cfft_sR_f32_len1024;
    break;
case 2048:
    S = &riscv_cfft_sR_f32_len2048;
    break;
case 4096:
    S = &riscv_cfft_sR_f32_len4096;
    break;
}

```

The new `riscv_cfft_init_f32` can also be used.

**Q15 and Q31** The floating-point complex FFT uses a mixed-radix algorithm. Multiple radix-4 stages are performed along with a single radix-2 stage, as needed. The algorithm supports lengths of [16, 32, 64, ..., 4096] and each length uses a different twiddle factor table.

The function uses the standard FFT definition and output values may grow by a factor of `fftLen` when computing the forward transform. The inverse transform includes a scale of  $1/\text{fftLen}$  as part of the calculation and this matches the textbook definition of the inverse FFT.

Pre-initialized data structures containing twiddle factors and bit reversal tables are provided and defined in `riscv_const_structs.h`. Include this header in your function and then pass one of the constant structures as an argument to `riscv_cfft_q31`. For example:

```
riscv_cfft_q31(riscv_cfft_sR_q31_len64, pSrc, 1, 1)
```

computes a 64-point inverse complex FFT including bit reversal. The data structures are treated as constant data and not modified during the calculation. The same data structure can be reused for multiple transforms including mixing forward and inverse transforms.

Earlier releases of the library provided separate radix-2 and radix-4 algorithms that operated on floating-point data. These functions are still provided but are deprecated. The older functions are slower and less general than the new functions.

An example of initialization of the constants for the `riscv_cfft_q31` function follows:

```

const static riscv_cfft_instance_q31 *S;
...
switch (length) {
    case 16:
        S = &riscv_cfft_sR_q31_len16;
        break;
    case 32:
        S = &riscv_cfft_sR_q31_len32;
        break;
    case 64:
        S = &riscv_cfft_sR_q31_len64;
        break;
    case 128:
        S = &riscv_cfft_sR_q31_len128;
        break;
    case 256:
        S = &riscv_cfft_sR_q31_len256;
        break;
}

```

(continues on next page)

(continued from previous page)

```

case 512:
    S = &riscv_cfft_sR_q31_len512;
    break;
case 1024:
    S = &riscv_cfft_sR_q31_len1024;
    break;
case 2048:
    S = &riscv_cfft_sR_q31_len2048;
    break;
case 4096:
    S = &riscv_cfft_sR_q31_len4096;
    break;
}

```

## Functions

void **riscv\_cfft\_f16**(**const** riscv\_cfft\_instance\_f16 \*S, float16\_t \*p1, uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Processing function for the floating-point complex FFT.

**Return** none

### Parameters

- [in] S: points to an instance of the floating-point CFFT structure
- [inout] p1: points to the complex data buffer of size 2\*fftLen. Processing occurs in-place
- [in] ifftFlag: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] bitReverseFlag: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

void **riscv\_cfft\_f32**(**const** riscv\_cfft\_instance\_f32 \*S, float32\_t \*p1, uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Processing function for the floating-point complex FFT.

**Return** none

### Parameters

- [in] S: points to an instance of the floating-point CFFT structure
- [inout] p1: points to the complex data buffer of size 2\*fftLen. Processing occurs in-place
- [in] ifftFlag: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] bitReverseFlag: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output



- value = 1: enables bit reversal of output

void **riscv\_cfft\_f64** (const riscv\_cfft\_instance\_f64 \*S, float64\_t \*p1, uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Processing function for the Double Precision floating-point complex FFT.

**Return** none

**Parameters**

- [in] S: points to an instance of the Double Precision floating-point CFFT structure
- [inout] p1: points to the complex data buffer of size 2\*fftLen. Processing occurs in-place
- [in] ifftFlag: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] bitReverseFlag: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

riscv\_status **riscv\_cfft\_init\_f16** (riscv\_cfft\_instance\_f16 \*S, uint16\_t *fftLen*)

Initialization function for the cfft f16 function.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Use of this function is mandatory only for the MVE version of the FFT.** Other versions can still initialize directly the data structure using variables declared in riscv\_const\_structs.h

**Parameters**

- [inout] S: points to an instance of the floating-point CFFT structure
- [in] fftLen: fft length (number of complex samples)

riscv\_status **riscv\_cfft\_init\_f32** (riscv\_cfft\_instance\_f32 \*S, uint16\_t *fftLen*)

Initialization function for the cfft f32 function.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Use of this function is mandatory only for the MVE version of the FFT.** Other versions can still initialize directly the data structure using variables declared in riscv\_const\_structs.h

**Parameters**

- [inout] S: points to an instance of the floating-point CFFT structure
- [in] fftLen: fft length (number of complex samples)

riscv\_status **riscv\_cfft\_init\_f64** (riscv\_cfft\_instance\_f64 \*S, uint16\_t *fftLen*)

Initialization function for the cfft f64 function.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Use of this function is mandatory only for the MVE version of the FFT.** Other versions can still initialize directly the data structure using variables declared in `riscv_const_structs.h`

**Parameters**

- [inout] `S`: points to an instance of the floating-point CFFT structure
- [in] `fftLen`: fft length (number of complex samples)

riscv\_status **riscv\_cfft\_init\_q15** (riscv\_cfft\_instance\_q15 \*`S`, uint16\_t `fftLen`)  
Initialization function for the cfft q15 function.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Use of this function is mandatory only for the MVE version of the FFT.** Other versions can still initialize directly the data structure using variables declared in `riscv_const_structs.h`

**Parameters**

- [inout] `S`: points to an instance of the floating-point CFFT structure
- [in] `fftLen`: fft length (number of complex samples)

riscv\_status **riscv\_cfft\_init\_q31** (riscv\_cfft\_instance\_q31 \*`S`, uint16\_t `fftLen`)  
Initialization function for the cfft q31 function.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Use of this function is mandatory only for the MVE version of the FFT.** Other versions can still initialize directly the data structure using variables declared in `riscv_const_structs.h`

**Parameters**

- [inout] `S`: points to an instance of the floating-point CFFT structure
- [in] `fftLen`: fft length (number of complex samples)

void **riscv\_cfft\_q15** (**const** riscv\_cfft\_instance\_q15 \*`S`, q15\_t \*`p1`, uint8\_t `ifftFlag`, uint8\_t `bitReverseFlag`)  
Processing function for Q15 complex FFT.

**Return** none

**Parameters**

- [in] `S`: points to an instance of Q15 CFFT structure
- [inout] `p1`: points to the complex data buffer of size  $2 * \text{fftLen}$ . Processing occurs in-place
- [in] `ifftFlag`: flag that selects transform direction
  - value = 0: forward transform

- value = 1: inverse transform
- [in] bitReverseFlag: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

void **riscv\_cfft\_q31**(const riscv\_cfft\_instance\_q31 \*S, q31\_t \*p1, uint8\_t ifftFlag, uint8\_t bitReverseFlag)

Processing function for the Q31 complex FFT.

**Return** none

**Parameters**

- [in] S: points to an instance of the fixed-point CFFT structure
- [inout] p1: points to the complex data buffer of size 2\*fftLen. Processing occurs in-place
- [in] ifftFlag: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] bitReverseFlag: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

void **riscv\_cfft\_radix2\_f16**(const riscv\_cfft\_radix2\_instance\_f16 \*S, float16\_t \*pSrc)

Radix-2 CFFT/CIFFT.

**Return** none

**Parameters**

- [in] S: points to an instance of the floating-point Radix-2 CFFT/CIFFT structure
- [inout] pSrc: points to the complex data buffer of size 2\*fftLen. Processing occurs in-place

void **riscv\_cfft\_radix2\_f32**(const riscv\_cfft\_radix2\_instance\_f32 \*S, float32\_t \*pSrc)

Radix-2 CFFT/CIFFT.

**Return** none

**Parameters**

- [in] S: points to an instance of the floating-point Radix-2 CFFT/CIFFT structure
- [inout] pSrc: points to the complex data buffer of size 2\*fftLen. Processing occurs in-place

riscv\_status **riscv\_cfft\_radix2\_init\_f16**(riscv\_cfft\_radix2\_instance\_f16 \*S, uint16\_t fftLen, uint8\_t ifftFlag, uint8\_t bitReverseFlag)

Initialization function for the floating-point CFFT/CIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful

- RISC\_V\_MATH\_ARGUMENT\_ERROR : `fftLen` is not a supported length

**Details** The parameter `ifftFlag` controls whether a forward or inverse transform is computed. Set(=1) `ifftFlag` for calculation of CIFFT otherwise CFFT is calculated

The parameter `bitReverseFlag` controls whether output is in normal order or bit reversed order. Set(=1) `bitReverseFlag` for output to be in normal order otherwise output is in bit reversed order.

The parameter `fftLen` Specifies length of CFFT/CIFFT process. Supported FFT Lengths are 16, 64, 256, 1024.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

#### Parameters

- [inout] `S`: points to an instance of the floating-point CFFT/CIFFT structure
- [in] `fftLen`: length of the FFT
- [in] `ifftFlag`: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] `bitReverseFlag`: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

riscv\_status **riscv\_cfft\_radix2\_init\_f32** (riscv\_cfft\_radix2\_instance\_f32 \*`S`, uint16\_t `fftLen`,  
uint8\_t `ifftFlag`, uint8\_t `bitReverseFlag`)

Initialization function for the floating-point CFFT/CIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : `fftLen` is not a supported length

**Details** The parameter `ifftFlag` controls whether a forward or inverse transform is computed. Set(=1) `ifftFlag` for calculation of CIFFT otherwise CFFT is calculated

The parameter `bitReverseFlag` controls whether output is in normal order or bit reversed order. Set(=1) `bitReverseFlag` for output to be in normal order otherwise output is in bit reversed order.

The parameter `fftLen` Specifies length of CFFT/CIFFT process. Supported FFT Lengths are 16, 64, 256, 1024.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

#### Parameters

- [inout] `S`: points to an instance of the floating-point CFFT/CIFFT structure
- [in] `fftLen`: length of the FFT
- [in] `ifftFlag`: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] `bitReverseFlag`: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output

- value = 1: enables bit reversal of output

riscv\_status **riscv\_cfft\_radix2\_init\_q15** (riscv\_cfft\_radix2\_instance\_q15 \*S, uint16\_t *fftLen*,  
uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Initialization function for the Q15 CFFT/CIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLen* is not a supported length

**Details** The parameter *ifftFlag* controls whether a forward or inverse transform is computed. Set(=1) *ifftFlag* for calculation of CIFFT otherwise CFFT is calculated

The parameter *bitReverseFlag* controls whether output is in normal order or bit reversed order. Set(=1) *bitReverseFlag* for output to be in normal order otherwise output is in bit reversed order.

The parameter *fftLen* Specifies length of CFFT/CIFFT process. Supported FFT Lengths are 16, 64, 256, 1024.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

#### Parameters

- [inout] S: points to an instance of the Q15 CFFT/CIFFT structure.
- [in] *fftLen*: length of the FFT.
- [in] *ifftFlag*: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] *bitReverseFlag*: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

riscv\_status **riscv\_cfft\_radix2\_init\_q31** (riscv\_cfft\_radix2\_instance\_q31 \*S, uint16\_t *fftLen*,  
uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Initialization function for the Q31 CFFT/CIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLen* is not a supported length

**Details** The parameter *ifftFlag* controls whether a forward or inverse transform is computed. Set(=1) *ifftFlag* for calculation of CIFFT otherwise CFFT is calculated

The parameter *bitReverseFlag* controls whether output is in normal order or bit reversed order. Set(=1) *bitReverseFlag* for output to be in normal order otherwise output is in bit reversed order.

The parameter *fftLen* Specifies length of CFFT/CIFFT process. Supported FFT Lengths are 16, 64, 256, 1024.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

#### Parameters

- [inout] S: points to an instance of the Q31 CFFT/CIFFT structure

- [in] `fftLen`: length of the FFT
- [in] `ifftFlag`: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] `bitReverseFlag`: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

void **riscv\_cfft\_radix2\_q15** (**const** riscv\_cfft\_radix2\_instance\_q15 \*S, q15\_t \*pSrc)  
Processing function for the fixed-point CFFT/CIFFT.

**Return** none

**Parameters**

- [in] S: points to an instance of the fixed-point CFFT/CIFFT structure
- [inout] pSrc: points to the complex data buffer of size  $2 \times \text{fftLen}$ . Processing occurs in-place

void **riscv\_cfft\_radix2\_q31** (**const** riscv\_cfft\_radix2\_instance\_q31 \*S, q31\_t \*pSrc)  
Processing function for the fixed-point CFFT/CIFFT.

**Return** none

**Parameters**

- [in] S: points to an instance of the fixed-point CFFT/CIFFT structure
- [inout] pSrc: points to the complex data buffer of size  $2 \times \text{fftLen}$ . Processing occurs in-place

void **riscv\_cfft\_radix4by2\_f16** (float16\_t \*pSrc, uint32\_t *fftLen*, **const** float16\_t \*pCoef)

void **riscv\_cfft\_radix4\_f16** (**const** riscv\_cfft\_radix4\_instance\_f16 \*S, float16\_t \*pSrc)  
Processing function for the floating-point Radix-4 CFFT/CIFFT.

**Return** none

**Parameters**

- [in] S: points to an instance of the floating-point Radix-4 CFFT/CIFFT structure
- [inout] pSrc: points to the complex data buffer of size  $2 \times \text{fftLen}$ . Processing occurs in-place

void **riscv\_cfft\_radix4\_f32** (**const** riscv\_cfft\_radix4\_instance\_f32 \*S, float32\_t \*pSrc)  
Processing function for the floating-point Radix-4 CFFT/CIFFT.

**Return** none

**Parameters**

- [in] S: points to an instance of the floating-point Radix-4 CFFT/CIFFT structure
- [inout] pSrc: points to the complex data buffer of size  $2 \times \text{fftLen}$ . Processing occurs in-place

riscv\_status **riscv\_cfft\_radix4\_init\_f16**(riscv\_cfft\_radix4\_instance\_f16 \*S, uint16\_t *fftLen*,  
uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Initialization function for the floating-point CFFT/CIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLen* is not a supported length

**Details** The parameter *ifftFlag* controls whether a forward or inverse transform is computed. Set(=1) *ifftFlag* for calculation of CIFFT otherwise CFFT is calculated

The parameter *bitReverseFlag* controls whether output is in normal order or bit reversed order. Set(=1) *bitReverseFlag* for output to be in normal order otherwise output is in bit reversed order.

The parameter *fftLen* Specifies length of CFFT/CIFFT process. Supported FFT Lengths are 16, 64, 256, 1024.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

#### Parameters

- [inout] S: points to an instance of the floating-point CFFT/CIFFT structure
- [in] *fftLen*: length of the FFT
- [in] *ifftFlag*: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] *bitReverseFlag*: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

riscv\_status **riscv\_cfft\_radix4\_init\_f32**(riscv\_cfft\_radix4\_instance\_f32 \*S, uint16\_t *fftLen*,  
uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Initialization function for the floating-point CFFT/CIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLen* is not a supported length

**Details** The parameter *ifftFlag* controls whether a forward or inverse transform is computed. Set(=1) *ifftFlag* for calculation of CIFFT otherwise CFFT is calculated

The parameter *bitReverseFlag* controls whether output is in normal order or bit reversed order. Set(=1) *bitReverseFlag* for output to be in normal order otherwise output is in bit reversed order.

The parameter *fftLen* Specifies length of CFFT/CIFFT process. Supported FFT Lengths are 16, 64, 256, 1024.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

#### Parameters

- [inout] S: points to an instance of the floating-point CFFT/CIFFT structure
- [in] *fftLen*: length of the FFT
- [in] *ifftFlag*: flag that selects transform direction

- value = 0: forward transform
- value = 1: inverse transform
- [in] `bitReverseFlag`: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

riscv\_status **riscv\_cfft\_radix4\_init\_q15** (riscv\_cfft\_radix4\_instance\_q15 \*S, uint16\_t *fftLen*,  
uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Initialization function for the Q15 CFFT/CIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLen* is not a supported length

**Details** The parameter *ifftFlag* controls whether a forward or inverse transform is computed. Set(=1) *ifftFlag* for calculation of CIFFT otherwise CFFT is calculated

The parameter *bitReverseFlag* controls whether output is in normal order or bit reversed order. Set(=1) *bitReverseFlag* for output to be in normal order otherwise output is in bit reversed order.

The parameter *fftLen* Specifies length of CFFT/CIFFT process. Supported FFT Lengths are 16, 64, 256, 1024.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

#### Parameters

- [inout] *S*: points to an instance of the Q15 CFFT/CIFFT structure
- [in] *fftLen*: length of the FFT
- [in] *ifftFlag*: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] *bitReverseFlag*: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

riscv\_status **riscv\_cfft\_radix4\_init\_q31** (riscv\_cfft\_radix4\_instance\_q31 \*S, uint16\_t *fftLen*,  
uint8\_t *ifftFlag*, uint8\_t *bitReverseFlag*)

Initialization function for the Q31 CFFT/CIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLen* is not a supported length

**Details** The parameter *ifftFlag* controls whether a forward or inverse transform is computed. Set(=1) *ifftFlag* for calculation of CIFFT otherwise CFFT is calculated

The parameter *bitReverseFlag* controls whether output is in normal order or bit reversed order. Set(=1) *bitReverseFlag* for output to be in normal order otherwise output is in bit reversed order.



The parameter `fftLen` Specifies length of CFFT/CIFFT process. Supported FFT Lengths are 16, 64, 256, 1024.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

#### Parameters

- [inout] `S`: points to an instance of the Q31 CFFT/CIFFT structure.
- [in] `fftLen`: length of the FFT.
- [in] `ifftFlag`: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] `bitReverseFlag`: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

void **riscv\_cfft\_radix4\_q15** (const riscv\_cfft\_radix4\_instance\_q15 \*S, q15\_t \*pSrc)

Processing function for the Q15 CFFT/CIFFT.

**Return** none

**Input and output formats:** Internally input is downscaled by 2 for every stage to avoid saturations inside CFFT/CIFFT process. Hence the output format is different for different FFT sizes. The input and output formats for different FFT sizes and number of bits to upscale are mentioned in the tables below for CFFT and CIFFT:

CFFT Size	Input format	Output Format	Number of bits to upscale
16	1.15	5.11	4
64	1.15	7.9	6
256	1.15	9.7	8
1024	1.15	11.5	10

CIFFT Size
16
64
256
1024

#### Parameters

- [in] `S`: points to an instance of the Q15 CFFT/CIFFT structure.
- [inout] `pSrc`: points to the complex data buffer. Processing occurs in-place.

void **riscv\_cfft\_radix4\_q31** (const riscv\_cfft\_radix4\_instance\_q31 \*S, q31\_t \*pSrc)

Processing function for the Q31 CFFT/CIFFT.

**Return** none

**Input and output formats:** Internally input is downscaled by 2 for every stage to avoid saturations inside CFFT/CIFFT process. Hence the output format is different for different FFT sizes. The input and output formats for different FFT sizes and number of bits to upscale are mentioned in the tables below for CFFT and CIFFT:

CFFT Size	Input format	Output Format	Number of bits to upscale	CIFFT Size
16	1.31	5.27	4	16
64	1.31	7.25	6	64
256	1.31	9.23	8	256
1024	1.31	11.21	10	1024

#### Parameters

- [in] S: points to an instance of the Q31 CFFT/CIFFT structure
- [inout] pSrc: points to the complex data buffer of size 2\*fftLen. Processing occurs in-place

### DCT Type IV Functions

#### DCT Type IV Tables

```
const float32_t Weights_128[256]
const float32_t cos_factors_128[128]
const float32_t Weights_512[1024]
const float32_t cos_factors_512[512]
const float32_t Weights_2048[4096]
const float32_t cos_factors_2048[2048]
const float32_t Weights_8192[16384]
const float32_t cos_factors_8192[8192]
const q31_t WeightsQ31_128[256]
const q31_t cos_factorsQ31_128[128]
const q31_t WeightsQ31_512[1024]
const q31_t cos_factorsQ31_512[512]
const q31_t WeightsQ31_2048[4096]
const q31_t cos_factorsQ31_2048[2048]
const q31_t WeightsQ31_8192[16384]
const q31_t cos_factorsQ31_8192[8192]
group DCT4_IDCT4_Table
end of RealFFT_Table group
```

#### Variables

```
const float32_t Weights_128[256]
Weights Table.
```

Weights tables are generated using the formula :

C command to generate the table

where  $N$  is the Number of weights to be calculated and  $c$  is  $\pi / (2 * N)$

In the tables below the real and imaginary values are placed alternatively, hence the array length is  $2 * N$ .  
cosFactor tables are generated using the formula :

C command to generate the table

where  $N$  is the number of factors to generate and  $c$  is  $\pi / (2 * N)$

```
const float32_t cos_factors_128[128]
const float32_t Weights_512[1024]
const float32_t cos_factors_512[512]
const float32_t Weights_2048[4096]
const float32_t cos_factors_2048[2048]
const float32_t Weights_8192[16384]
const float32_t cos_factors_8192[8192]
const q31_t WeightsQ31_128[256]
```

Weights tables are generated using the formula :

C command to generate the table

where  $N$  is the Number of weights to be calculated and  $c$  is  $\pi / (2 * N)$

Convert the output to q31 format by multiplying with  $2^{31}$  and saturated if required.

In the tables below the real and imaginary values are placed alternatively, hence the array length is  $2 * N$ .  
cosFactor tables are generated using the formula :

C command to generate the table

where  $N$  is the number of factors to generate and  $c$  is  $\pi / (2 * N)$

Then converted to q31 format by multiplying with  $2^{31}$  and saturated if required.

```
const q31_t cos_factorsQ31_128[128]
const q31_t WeightsQ31_512[1024]
const q31_t cos_factorsQ31_512[512]
const q31_t WeightsQ31_2048[4096]
const q31_t cos_factorsQ31_2048[2048]
const q31_t WeightsQ31_8192[16384]
const q31_t cos_factorsQ31_8192[8192]
```

```
void riscv_dct4_f32 (const riscv_dct4_instance_f32 *S, float32_t *pState, float32_t *pInlineBuffer)
riscv_status riscv_dct4_init_f32 (riscv_dct4_instance_f32 *S, riscv_rfft_instance_f32 *S_RFFT,
                                riscv_cfft_radix4_instance_f32 *S_CFFT, uint16_t N, uint16_t
                                Nby2, float32_t normalize)
riscv_status riscv_dct4_init_q15 (riscv_dct4_instance_q15 *S, riscv_rfft_instance_q15 *S_RFFT,
                                riscv_cfft_radix4_instance_q15 *S_CFFT, uint16_t N, uint16_t
                                Nby2, q15_t normalize)
```

```
riscv_status riscv_dct4_init_q31 (riscv_dct4_instance_q31 *S, riscv_rfft_instance_q31 *S_RFFT,
                                riscv_cfft_radix4_instance_q31 *S_CFFT, uint16_t N, uint16_t
                                Nby2, q31_t normalize)
```

```
void riscv_dct4_q15 (const riscv_dct4_instance_q15 *S, q15_t *pState, q15_t *pInlineBuffer)
```

```
void riscv_dct4_q31 (const riscv_dct4_instance_q31 *S, q31_t *pState, q31_t *pInlineBuffer)
```

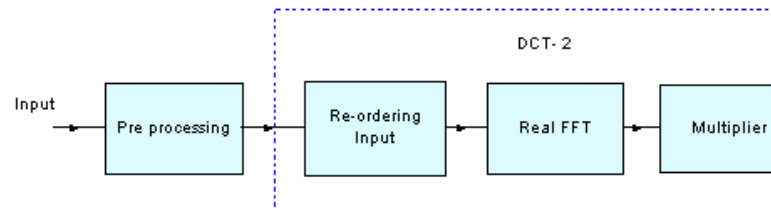
#### group **DCT4\_IDCT4**

Representation of signals by minimum number of values is important for storage and transmission. The possibility of large discontinuity between the beginning and end of a period of a signal in DFT can be avoided by extending the signal so that it is even-symmetric. Discrete Cosine Transform (DCT) is constructed such that its energy is heavily concentrated in the lower part of the spectrum and is very widely used in signal and image coding applications. The family of DCTs (DCT type- 1,2,3,4) is the outcome of different combinations of homogeneous boundary conditions. DCT has an excellent energy-packing capability, hence has many applications and in data compression in particular.

DCT is essentially the Discrete Fourier Transform(DFT) of an even-extended real signal. Reordering of the input data makes the computation of DCT just a problem of computing the DFT of a real signal with a few additional operations. This approach provides regular, simple, and very efficient DCT algorithms for practical hardware and software implementations.

DCT type-II can be implemented using Fast fourier transform (FFT) internally, as the transform is applied on real values, Real FFT can be used. DCT4 is implemented using DCT2 as their implementations are similar except with some added pre-processing and post-processing. DCT2 implementation can be described in the following steps:

- Re-ordering input
- Calculating Real FFT
- Multiplication of weights and Real FFT output and getting real part from the product.



This process is explained by the block diagram below:

**Algorithm** The N-point type-IV DCT is defined as a real, linear transformation by the formula: where  $k = 0, 1, 2, \dots, N-1$

$$X_c(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos \left[ \left( n + \frac{1}{2} \right) \left( k + \frac{1}{2} \right) \frac{\pi}{N} \right]$$

Its inverse is defined as follows: where  $n = 0, 1, 2, \dots, N-1$

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} X_c(k) \cos \left[ \left( n + \frac{1}{2} \right) \left( k + \frac{1}{2} \right) \frac{\pi}{N} \right]$$

The DCT4 matrices become involutory (i.e. they are self-inverse) by multiplying with an overall scale factor of  $\sqrt{2/N}$ . The symmetry of the transform matrix indicates that the fast algorithms for the forward and inverse transform computation are identical. Note that the implementation of Inverse DCT4 and DCT4 is same, hence same process function can be used for both.

**Lengths supported by the transform:** As DCT4 internally uses Real FFT, it supports all the lengths 128, 512, 2048 and 8192. The library provides separate functions for Q15, Q31, and floating-point data types.

**Instance Structure** The instances for Real FFT and FFT, cosine values table and twiddle factor table are stored in an instance data structure. A separate instance structure must be defined for each transform. There are separate instance structure declarations for each of the 3 supported data types.

**Initialization Functions** There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Initializes Real FFT as its process function is used internally in DCT4, by calling `riscv_rfft_init_f32()`.

Use of the initialization function is optional. However, if the initialization function is used, then the instance structure cannot be placed into a const data section. To place an instance structure into a const data section, the instance structure must be manually initialized. Manually initialize the instance structure as follows: where `N` is the length of the DCT4; `Nby2` is half of the length of the DCT4; `normalize` is normalizing factor used and is equal to  $\sqrt{2/N}$ ; `pTwiddle` points to the twiddle factor table; `pCosFactor` points to the cosFactor table; `pRfft` points to the real FFT instance; `pCfft` points to the complex FFT instance; The CFFT and RFFT structures also needs to be initialized, refer to `riscv_cfft_radix4_f32()` and `riscv_rfft_f32()` respectively for details regarding static initialization.

**Fixed-Point Behavior** Care must be taken when using the fixed-point versions of the DCT4 transform functions. In particular, the overflow and saturation behavior of the accumulator used in each function must be considered. Refer to the function specific documentation below for usage guidelines.

## Functions

void **riscv\_dct4\_f32** (**const** riscv\_dct4\_instance\_f32 \*S, float32\_t \*pState, float32\_t \*pInlineBuffer)  
*uffer*

Processing function for the floating-point DCT4/IDCT4.

**Return** none

**Parameters**

- [in] S: points to an instance of the floating-point DCT4/IDCT4 structure
- [in] pState: points to state buffer
- [inout] pInlineBuffer: points to the in-place input and output buffer

riscv\_status **riscv\_dct4\_init\_f32** (riscv\_dct4\_instance\_f32 \*S, riscv\_rfft\_instance\_f32 \*S\_RFFT, riscv\_cfft\_radix4\_instance\_f32 \*S\_CFFT, uint16\_t N, uint16\_t Nby2, float32\_t normalize)

Initialization function for the floating-point DCT4/IDCT4.

DCT Size	Normalizing factor value
2048	0.03125
512	0.0625
128	0.125

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : N is not a supported transform length

**Normalizing factor** The normalizing factor is  $\sqrt{2/N}$ , which depends on the size of transform N. Floating-point normalizing factors are mentioned in the table below for different DCT sizes:

**Parameters**

- [inout] S: points to an instance of floating-point DCT4/IDCT4 structure
- [in] S\_RFFT: points to an instance of floating-point RFFT/RIFFT structure
- [in] S\_CFFT: points to an instance of floating-point CFFT/CIFFT structure
- [in] N: length of the DCT4
- [in] Nby2: half of the length of the DCT4
- [in] normalize: normalizing factor.

riscv\_status **riscv\_dct4\_init\_q15** (riscv\_dct4\_instance\_q15 \*S, riscv\_rfft\_instance\_q15 \*S\_RFFT, riscv\_cfft\_radix4\_instance\_q15 \*S\_CFFT, uint16\_t N, uint16\_t Nby2, q15\_t normalize)

Initialization function for the Q15 DCT4/IDCT4.

DCT Size	Normalizing factor value (hexadecimal)
2048	0x400
512	0x800
128	0x1000

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : N is not a supported transform length

**Normalizing factor** The normalizing factor is  $\sqrt{2/N}$ , which depends on the size of transform N. Normalizing factors in 1.15 format are mentioned in the table below for different DCT sizes:

**Parameters**

- [inout] S: points to an instance of Q15 DCT4/IDCT4 structure
- [in] S\_RFFT: points to an instance of Q15 RFFT/RIFFT structure
- [in] S\_CFFT: points to an instance of Q15 CFFT/CIFFT structure
- [in] N: length of the DCT4
- [in] Nby2: half of the length of the DCT4
- [in] normalize: normalizing factor

```
riscv_status riscv_dct4_init_q31 (riscv_dct4_instance_q31 *S, riscv_rfft_instance_q31
                                *S_RFFT, riscv_cfft_radix4_instance_q31 *S_CFFT, uint16_t
                                N, uint16_t Nby2, q31_t normalize)
```

Initialization function for the Q31 DCT4/IDCT4.

DCT Size	Normalizing factor value (hexadecimal)
2048	0x40000000
512	0x80000000
128	0x100000000

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : N is not a supported transform length

**Normalizing factor:** The normalizing factor is  $\sqrt{2/N}$ , which depends on the size of transform N. Normalizing factors in 1.31 format are mentioned in the table below for different DCT sizes:

**Parameters**

- [inout] S: points to an instance of Q31 DCT4/IDCT4 structure.
- [in] S\_RFFT: points to an instance of Q31 RFFT/RIFFT structure
- [in] S\_CFFT: points to an instance of Q31 CFFT/CIFFT structure
- [in] N: length of the DCT4.
- [in] Nby2: half of the length of the DCT4.
- [in] normalize: normalizing factor.

```
void riscv_dct4_q15 (const riscv_dct4_instance_q15 *S, q15_t *pState, q15_t *pInlineBuffer)
```

Processing function for the Q15 DCT4/IDCT4.

DCT Size	Input format	Output format	Number of bits to upscale
2048	1.15	11.5	10
512	1.15	9.7	8
128	1.15	7.9	6

**Return** none

**Input an output formats** Internally inputs are downsampled in the RFFT process function to avoid overflows. Number of bits downsampled, depends on the size of the transform. The input and output formats for different DCT sizes and number of bits to upscale are mentioned in the table below:

**Parameters**

- [in] S: points to an instance of the Q15 DCT4 structure.
- [in] pState: points to state buffer.

- [inout] pInlineBuffer: points to the in-place input and output buffer.

void **riscv\_dct4\_q31** (const riscv\_dct4\_instance\_q31 \*S, q31\_t \*pState, q31\_t \*pInlineBuffer)  
Processing function for the Q31 DCT4/IDCT4.

DCT Size	Input format	Output format	Number of bits to upscale
2048	2.30	12.20	11
512	2.30	10.22	9
128	2.30	8.24	7

**Return** none

**Input an output formats** Input samples need to be downscaled by 1 bit to avoid saturations in the Q31 DCT process, as the conversion from DCT2 to DCT4 involves one subtraction. Internally inputs are downscaled in the RFFT process function to avoid overflows. Number of bits downscaled, depends on the size of the transform. The input and output formats for different DCT sizes and number of bits to upscale are mentioned in the table below:

**Parameters**

- [in] S: points to an instance of the Q31 DCT4 structure.
- [in] pState: points to state buffer.
- [inout] pInlineBuffer: points to the in-place input and output buffer.

## Real FFT Functions

### Real FFT Tables

```
const float32_t realCoefA[8192]
const float32_t realCoefB[8192]
const q31_t realCoefAQ31[8192]
const q31_t realCoefBQ31[8192]
const q15_t __ALIGNED(4)
group RealFFT_Table
```

### Functions

```
const q15_t __ALIGNED(4)
Weights Table.
Q15 table for reciprocal.
end of DCT4_IDCT4_Table group
Generation fixed-point realCoefAQ15 array in Q15 format:
n = 4096
Convert to fixed point Q15 format round(pATable[i] * pow(2, 15))
Generation of real_CoefB array:
```



$n = 4096$

Convert to fixed point Q15 format  $\text{round}(pBTable[i] * \text{pow}(2, 15))$

Weights tables are generated using the formula :

C command to generate the table

where  $N$  is the Number of weights to be calculated and  $c$  is  $\pi / (2 * N)$

Converted the output to q15 format by multiplying with  $2^{31}$  and saturated if required.

In the tables below the real and imaginary values are placed alternatively, hence the array length is  $2 * N$ .

cosFactor tables are generated using the formula :

C command to generate the table

where  $N$  is the number of factors to generate and  $c$  is  $\pi / (2 * N)$

Then converted to q15 format by multiplying with  $2^{31}$  and saturated if required.

### Variables

**const** float32\_t **realCoefA**[8192]

Generation of realCoefA array:

$n = 4096$

**const** float32\_t **realCoefB**[8192]

Generation of realCoefB array:

$n = 4096$

**const** q31\_t **realCoefAQ31**[8192]

Generation fixed-point realCoefAQ31 array in Q31 format:

$n = 4096$

Convert to fixed point Q31 format  $\text{round}(pATable[i] * \text{pow}(2, 31))$

**const** q31\_t **realCoefBQ31**[8192]

Generation of realCoefBQ31 array:

$n = 4096$

Convert to fixed point Q31 format  $\text{round}(pBTable[i] * \text{pow}(2, 31))$

void **riscv\_rfft\_f32** (**const** riscv\_rfft\_instance\_f32 \*S, float32\_t \*pSrc, float32\_t \*pDst)

void **riscv\_rfft\_fast\_f16** (**const** riscv\_rfft\_fast\_instance\_f16 \*S, float16\_t \*p, float16\_t \*pOut, uint8\_t *fftFlag*)

void **riscv\_rfft\_fast\_f32** (**const** riscv\_rfft\_fast\_instance\_f32 \*S, float32\_t \*p, float32\_t \*pOut, uint8\_t *fftFlag*)

void **riscv\_rfft\_fast\_f64** (riscv\_rfft\_fast\_instance\_f64 \*S, float64\_t \*p, float64\_t \*pOut, uint8\_t *fftFlag*)

riscv\_status **riscv\_rfft\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S, uint16\_t *fftLen*)

riscv\_status **riscv\_rfft\_fast\_init\_f32** (riscv\_rfft\_fast\_instance\_f32 \*S, uint16\_t *fftLen*)

**static** riscv\_status **riscv\_rfft\_32\_fast\_init\_f64** (riscv\_rfft\_fast\_instance\_f64 \*S)

```

static riscv_status riscv_rfft_64_fast_init_f64 (riscv_rfft_fast_instance_f64 *S)
static riscv_status riscv_rfft_128_fast_init_f64 (riscv_rfft_fast_instance_f64 *S)
static riscv_status riscv_rfft_256_fast_init_f64 (riscv_rfft_fast_instance_f64 *S)
static riscv_status riscv_rfft_512_fast_init_f64 (riscv_rfft_fast_instance_f64 *S)
static riscv_status riscv_rfft_1024_fast_init_f64 (riscv_rfft_fast_instance_f64 *S)
static riscv_status riscv_rfft_2048_fast_init_f64 (riscv_rfft_fast_instance_f64 *S)
static riscv_status riscv_rfft_4096_fast_init_f64 (riscv_rfft_fast_instance_f64 *S)
riscv_status riscv_rfft_fast_init_f64 (riscv_rfft_fast_instance_f64 *S, uint16_t fftLen)
riscv_status riscv_rfft_init_f32 (riscv_rfft_instance_f32 *S, riscv_cfft_radix4_instance_f32
                                *S_CFFT, uint32_t fftLenReal, uint32_t ifftFlagR, uint32_t bitReverseFlag)
riscv_status riscv_rfft_init_q15 (riscv_rfft_instance_q15 *S, uint32_t fftLenReal, uint32_t ifftFlagR,
                                uint32_t bitReverseFlag)
riscv_status riscv_rfft_init_q31 (riscv_rfft_instance_q31 *S, uint32_t fftLenReal, uint32_t ifftFlagR,
                                uint32_t bitReverseFlag)
void riscv_rfft_q15 (const riscv_rfft_instance_q15 *S, q15_t *pSrc, q15_t *pDst)
void riscv_rfft_q31 (const riscv_rfft_instance_q31 *S, q31_t *pSrc, q31_t *pDst)

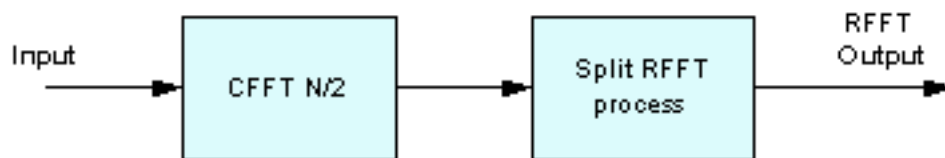
```

#### group **RealFFT**

The NMSIS DSP library includes specialized algorithms for computing the FFT of real data sequences. The FFT is defined over complex data but in many applications the input is real. Real FFT algorithms take advantage of the symmetry properties of the FFT and have a speed advantage over complex algorithms of the same length.

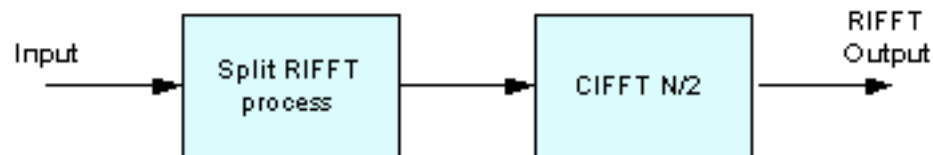
The Fast RFFT algorithm relays on the mixed radix CFFT that save processor usage.

The real length N forward FFT of a sequence is computed using the steps shown below.



The real sequence is initially treated as if it were complex to perform a CFFT. Later, a processing stage reshapes the data to obtain half of the frequency spectrum in complex format. Except the first complex number that contains the two real numbers  $X[0]$  and  $X[N/2]$  all the data is complex. In other words, the first complex sample contains two real values packed.

The input for the inverse RFFT should keep the same format as the output of the forward RFFT. A first processing stage pre-process the data to later perform an inverse CFFT.



The algorithms for floating-point, Q15, and Q31 data are slightly different and we describe each algorithm in turn.

**Floating-point** The main functions are `riscv_rfft_fast_f16()` and `riscv_rfft_fast_init_f16()`.

The FFT of a real N-point sequence has even symmetry in the frequency domain. The second half of the data equals the conjugate of the first half flipped in frequency. Looking at the data, we see that we can uniquely represent the FFT using only N/2 complex numbers. These are packed into the output array in alternating real and imaginary components:

$$X = \{ \text{real}[0], \text{imag}[0], \text{real}[1], \text{imag}[1], \text{real}[2], \text{imag}[2] \dots \text{real}[(N/2)-1], \text{imag}[(N/2)-1] \}$$

It happens that the first complex number (real[0], imag[0]) is actually all real. real[0] represents the DC offset, and imag[0] should be 0. (real[1], imag[1]) is the fundamental frequency, (real[2], imag[2]) is the first harmonic and so on.

The real FFT functions pack the frequency domain data in this fashion. The forward transform outputs the data in this form and the inverse transform expects input data in this form. The function always performs the needed bitreversal so that the input and output data is always in normal order. The functions support lengths of [32, 64, 128, ..., 4096] samples.

**Q15 and Q31** The real algorithms are defined in a similar manner and utilize N/2 complex transforms behind the scenes.

The complex transforms used internally include scaling to prevent fixed-point overflows. The overall scaling equals  $1/(\text{fftLen}/2)$ . Due to the use of complex transform internally, the source buffer is modified by the rfft.

A separate instance structure must be defined for each transform used but twiddle factor and bit reversal tables can be reused.

There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Initializes twiddle factor table and bit reversal table pointers.
- Initializes the internal complex FFT data structure.

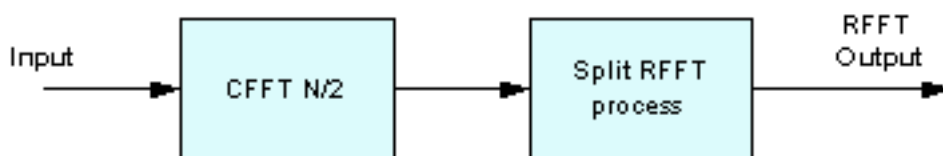
Use of the initialization function is optional **except for MVE versions where it is mandatory**. If you don't use the initialization functions, then the structures should be initialized with code similar to the one below: where `fftLenReal` is the length of the real transform; `fftLenBy2` length of the internal complex transform (`fftLenReal/2`). `ifftFlagR` Selects forward (=0) or inverse (=1) transform. `bitReverseFlagR` Selects bit reversed output (=0) or normal order output (=1). `twidCoefRModifier` stride modifier for the twiddle factor table. The value is based on the FFT length; `pTwiddleAReal` points to the A array of twiddle coefficients; `pTwiddleBReal` points to the B array of twiddle coefficients; `pCfft` points to the CFFT Instance structure. The CFFT structure must also be initialized.

Note that with MVE versions you can't initialize instance structures directly and **must use the initialization function**.

The NMSIS DSP library includes specialized algorithms for computing the FFT of real data sequences. The FFT is defined over complex data but in many applications the input is real. Real FFT algorithms take advantage of the symmetry properties of the FFT and have a speed advantage over complex algorithms of the same length.

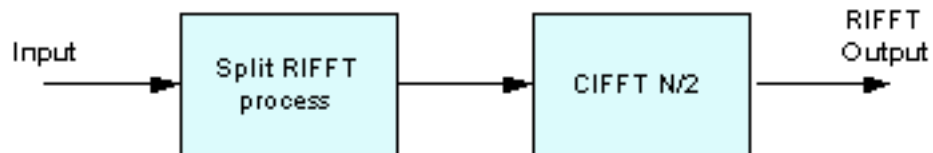
The Fast RFFT algorithm relays on the mixed radix CFFT that save processor usage.

The real length N forward FFT of a sequence is computed using the steps shown below.



The real sequence is initially treated as if it were complex to perform a CFFT. Later, a processing stage reshapes the data to obtain half of the frequency spectrum in complex format. Except the first complex number that contains the two real numbers  $X[0]$  and  $X[N/2]$  all the data is complex. In other words, the first complex sample contains two real values packed.

The input for the inverse RFFT should keep the same format as the output of the forward RFFT. A first processing stage pre-process the data to later perform an inverse CFFT.



The algorithms for floating-point, Q15, and Q31 data are slightly different and we describe each algorithm in turn.

**Floating-point** The main functions are `riscv_rfft_fast_f32()` and `riscv_rfft_fast_init_f32()`. The older functions `riscv_rfft_f32()` and `riscv_rfft_init_f32()` have been deprecated but are still documented.

The FFT of a real  $N$ -point sequence has even symmetry in the frequency domain. The second half of the data equals the conjugate of the first half flipped in frequency. Looking at the data, we see that we can uniquely represent the FFT using only  $N/2$  complex numbers. These are packed into the output array in alternating real and imaginary components:

$$X = \{ \text{real}[0], \text{imag}[0], \text{real}[1], \text{imag}[1], \text{real}[2], \text{imag}[2] \dots \text{real}[(N/2)-1], \text{imag}[(N/2)-1] \}$$

It happens that the first complex number ( $\text{real}[0], \text{imag}[0]$ ) is actually all real.  $\text{real}[0]$  represents the DC offset, and  $\text{imag}[0]$  should be 0. ( $\text{real}[1], \text{imag}[1]$ ) is the fundamental frequency, ( $\text{real}[2], \text{imag}[2]$ ) is the first harmonic and so on.

The real FFT functions pack the frequency domain data in this fashion. The forward transform outputs the data in this form and the inverse transform expects input data in this form. The function always performs the needed bitreversal so that the input and output data is always in normal order. The functions support lengths of [32, 64, 128, ..., 4096] samples.

**Q15 and Q31** The real algorithms are defined in a similar manner and utilize  $N/2$  complex transforms behind the scenes.

The complex transforms used internally include scaling to prevent fixed-point overflows. The overall scaling equals  $1/(\text{fftLen}/2)$ . Due to the use of complex transform internally, the source buffer is modified by the rfft.

A separate instance structure must be defined for each transform used but twiddle factor and bit reversal tables can be reused.

There is also an associated initialization function for each data type. The initialization function performs the following operations:

- Sets the values of the internal structure fields.
- Initializes twiddle factor table and bit reversal table pointers.
- Initializes the internal complex FFT data structure.

Use of the initialization function is optional **except for MVE versions where it is mandatory**. If you don't use the initialization functions, then the structures should be initialized with code similar to the one below: where `fftLenReal` is the length of the real transform; `fftLenBy2` length of the internal complex transform (`fftLenReal/2`). `ifftFlagR` Selects forward (`=0`) or inverse (`=1`) transform. `bitReverseFlagR` Selects bit reversed output (`=0`) or normal order output (`=1`). `twidCoefRModifier` stride modifier for the twiddle factor table. The value is based on the FFT length; `pTwiddleAReal` points to the A array of

twiddle coefficients; `pTwiddleBReal` points to the B array of twiddle coefficients; `pCfft` points to the CFFT Instance structure. The CFFT structure must also be initialized.

Note that with MVE versions you can't initialize instance structures directly and **must use the initialization function**.

## Functions

void **riscv\_rfft\_f32** (**const** riscv\_rfft\_instance\_f32 \*S, float32\_t \*pSrc, float32\_t \*pDst)

Processing function for the floating-point RFFT/RIFFT. Source buffer is modified by this function.

**Return** none

For the RIFFT, the source buffer must at least have length `fftLenReal + 2`. The last two elements must be equal to what would be generated by the RFFT: (`pSrc[0] - pSrc[1]`) and 0.0f

### Parameters

- [in] S: points to an instance of the floating-point RFFT/RIFFT structure
- [in] pSrc: points to the input buffer
- [out] pDst: points to the output buffer

void **riscv\_rfft\_fast\_f16** (**const** riscv\_rfft\_fast\_instance\_f16 \*S, float16\_t \*p, float16\_t \*pOut, uint8\_t *ifftFlag*)

Processing function for the floating-point real FFT.

**Return** none

### Parameters

- [in] S: points to an `riscv_rfft_fast_instance_f16` structure
- [in] p: points to input buffer (Source buffer is modified by this function.)
- [in] pOut: points to output buffer
- [in] *ifftFlag*:
  - value = 0: RFFT
  - value = 1: RIFFT

void **riscv\_rfft\_fast\_f32** (**const** riscv\_rfft\_fast\_instance\_f32 \*S, float32\_t \*p, float32\_t \*pOut, uint8\_t *ifftFlag*)

Processing function for the floating-point real FFT.

**Return** none

### Parameters

- [in] S: points to an `riscv_rfft_fast_instance_f32` structure
- [in] p: points to input buffer (Source buffer is modified by this function.)
- [in] pOut: points to output buffer
- [in] *ifftFlag*:
  - value = 0: RFFT
  - value = 1: RIFFT

void **riscv\_rfft\_fast\_f64** (riscv\_rfft\_fast\_instance\_f64 \*S, float64\_t \*p, float64\_t \*pOut, uint8\_t *ifftFlag*)

Processing function for the Double Precision floating-point real FFT.

**Return** none

**Parameters**

- [in] S: points to an riscv\_rfft\_fast\_instance\_f64 structure
- [in] p: points to input buffer (Source buffer is modified by this function.)
- [in] pOut: points to output buffer
- [in] ifftFlag:
  - value = 0: RFFT
  - value = 1: RIFFT

**static** riscv\_status **riscv\_rfft\_32\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S)

Initialization function for the 32pt floating-point real FFT.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an riscv\_rfft\_fast\_instance\_f16 structure

**static** riscv\_status **riscv\_rfft\_64\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S)

Initialization function for the 64pt floating-point real FFT.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an riscv\_rfft\_fast\_instance\_f16 structure

**static** riscv\_status **riscv\_rfft\_128\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S)

Initialization function for the 128pt floating-point real FFT.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an riscv\_rfft\_fast\_instance\_f16 structure

**static** riscv\_status **riscv\_rfft\_256\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S)

Initialization function for the 256pt floating-point real FFT.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f16` structure

**static** riscv\_status **riscv\_rfft\_512\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S)  
Initialization function for the 512pt floating-point real FFT.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f16` structure

**static** riscv\_status **riscv\_rfft\_1024\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S)  
Initialization function for the 1024pt floating-point real FFT.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f16` structure

**static** riscv\_status **riscv\_rfft\_2048\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S)  
Initialization function for the 2048pt floating-point real FFT.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f16` structure

**static** riscv\_status **riscv\_rfft\_4096\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S)  
Initialization function for the 4096pt floating-point real FFT.

**Return** execution status

- RISCV\_MATH\_SUCCESS : Operation successful
- RISCV\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f16` structure

riscv\_status **riscv\_rfft\_fast\_init\_f16** (riscv\_rfft\_fast\_instance\_f16 \*S, uint16\_t *fftLen*)  
Initialization function for the floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : `fftLen` is not a supported length

**Description** The parameter `fftLen` specifies the length of RFFT/CIFFT process. Supported FFT Lengths are 32, 64, 128, 256, 512, 1024, 2048, 4096.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f16` structure
- `[in]` `fftLen`: length of the Real Sequence

**static** `riscv_status` **`riscv_rfft_32_fast_init_f32`** (`riscv_rfft_fast_instance_f32 *S`)  
Initialization function for the 32pt floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f32` structure

**static** `riscv_status` **`riscv_rfft_64_fast_init_f32`** (`riscv_rfft_fast_instance_f32 *S`)  
Initialization function for the 64pt floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f32` structure

**static** `riscv_status` **`riscv_rfft_128_fast_init_f32`** (`riscv_rfft_fast_instance_f32 *S`)  
Initialization function for the 128pt floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f32` structure

**static** `riscv_status` **`riscv_rfft_256_fast_init_f32`** (`riscv_rfft_fast_instance_f32 *S`)  
Initialization function for the 256pt floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected



**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f32` structure

**static** riscv\_status **riscv\_rfft\_512\_fast\_init\_f32** (riscv\_rfft\_fast\_instance\_f32 \*S)  
 Initialization function for the 512pt floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f32` structure

**static** riscv\_status **riscv\_rfft\_1024\_fast\_init\_f32** (riscv\_rfft\_fast\_instance\_f32 \*S)  
 Initialization function for the 1024pt floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f32` structure

**static** riscv\_status **riscv\_rfft\_2048\_fast\_init\_f32** (riscv\_rfft\_fast\_instance\_f32 \*S)  
 Initialization function for the 2048pt floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f32` structure

**static** riscv\_status **riscv\_rfft\_4096\_fast\_init\_f32** (riscv\_rfft\_fast\_instance\_f32 \*S)  
 Initialization function for the 4096pt floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an `riscv_rfft_fast_instance_f32` structure

riscv\_status **riscv\_rfft\_fast\_init\_f32** (riscv\_rfft\_fast\_instance\_f32 \*S, uint16\_t *fftLen*)  
 Initialization function for the floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful

- **RISCV\_MATH\_ARGUMENT\_ERROR** : `fftLen` is not a supported length

**Description** The parameter `fftLen` specifies the length of RFFT/CIFFT process. Supported FFT Lengths are 32, 64, 128, 256, 512, 1024, 2048, 4096.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f32` structure
- `[in]` `fftLen`: length of the Real Sequence

**static** `riscv_status` **`riscv_rfft_32_fast_init_f64`** (`riscv_rfft_fast_instance_f64 *S`)  
Initialization function for the 32pt double precision floating-point real FFT.

**Return** execution status

- **RISCV\_MATH\_SUCCESS** : Operation successful
- **RISCV\_MATH\_ARGUMENT\_ERROR** : an error is detected

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f64` structure

**static** `riscv_status` **`riscv_rfft_64_fast_init_f64`** (`riscv_rfft_fast_instance_f64 *S`)  
Initialization function for the 64pt Double Precision floating-point real FFT.

**Return** execution status

- **RISCV\_MATH\_SUCCESS** : Operation successful
- **RISCV\_MATH\_ARGUMENT\_ERROR** : an error is detected

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f64` structure

**static** `riscv_status` **`riscv_rfft_128_fast_init_f64`** (`riscv_rfft_fast_instance_f64 *S`)  
Initialization function for the 128pt Double Precision floating-point real FFT.

**Return** execution status

- **RISCV\_MATH\_SUCCESS** : Operation successful
- **RISCV\_MATH\_ARGUMENT\_ERROR** : an error is detected

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f64` structure

**static** `riscv_status` **`riscv_rfft_256_fast_init_f64`** (`riscv_rfft_fast_instance_f64 *S`)  
Initialization function for the 256pt Double Precision floating-point real FFT.

**Return** execution status

- **RISCV\_MATH\_SUCCESS** : Operation successful
- **RISCV\_MATH\_ARGUMENT\_ERROR** : an error is detected

**Parameters**

- `[inout]` `S`: points to an `riscv_rfft_fast_instance_f64` structure

**static** riscv\_status **riscv\_rfft\_512\_fast\_init\_f64** (riscv\_rfft\_fast\_instance\_f64 \*S)

Initialization function for the 512pt Double Precision floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an riscv\_rfft\_fast\_instance\_f64 structure

**static** riscv\_status **riscv\_rfft\_1024\_fast\_init\_f64** (riscv\_rfft\_fast\_instance\_f64 \*S)

Initialization function for the 1024pt Double Precision floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an riscv\_rfft\_fast\_instance\_f64 structure

**static** riscv\_status **riscv\_rfft\_2048\_fast\_init\_f64** (riscv\_rfft\_fast\_instance\_f64 \*S)

Initialization function for the 2048pt Double Precision floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an riscv\_rfft\_fast\_instance\_f64 structure

**static** riscv\_status **riscv\_rfft\_4096\_fast\_init\_f64** (riscv\_rfft\_fast\_instance\_f64 \*S)

Initialization function for the 4096pt Double Precision floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : an error is detected

**Parameters**

- [inout] S: points to an riscv\_rfft\_fast\_instance\_f64 structure

riscv\_status **riscv\_rfft\_fast\_init\_f64** (riscv\_rfft\_fast\_instance\_f64 \*S, uint16\_t *fftLen*)

Initialization function for the Double Precision floating-point real FFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLen* is not a supported length

**Description** The parameter *fftLen* specifies the length of RFFT/CIFFT process. Supported FFT Lengths are 32, 64, 128, 256, 512, 1024, 2048, 4096.

This Function also initializes Twiddle factor table pointer and Bit reversal table pointer.

**Parameters**

- [inout] *S*: points to an `riscv_rfft_fast_instance_f64` structure
- [in] *fftLen*: length of the Real Sequence

riscv\_status **riscv\_rfft\_init\_f32** (riscv\_rfft\_instance\_f32 \**S*, riscv\_cfft\_radix4\_instance\_f32 \**S\_CFFT*, uint32\_t *fftLenReal*, uint32\_t *ifftFlagR*, uint32\_t *bitReverseFlag*)

Initialization function for the floating-point RFFT/RIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLenReal* is not a supported length

**Description** The parameter *fftLenReal* specifies length of RFFT/RIFFT Process. Supported FFT Lengths are 128, 512, 2048.

The parameter *ifftFlagR* controls whether a forward or inverse transform is computed. Set(=1) *ifftFlagR* to calculate RIFFT, otherwise RFFT is calculated.

The parameter *bitReverseFlag* controls whether output is in normal order or bit reversed order. Set(=1) *bitReverseFlag* for output to be in normal order otherwise output is in bit reversed order.

This function also initializes Twiddle factor table.

**Parameters**

- [inout] *S*: points to an instance of the floating-point RFFT/RIFFT structure
- [inout] *S\_CFFT*: points to an instance of the floating-point CFFT/CIFFT structure
- [in] *fftLenReal*: length of the FFT.
- [in] *ifftFlagR*: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] *bitReverseFlag*: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

riscv\_status **riscv\_rfft\_init\_q15** (riscv\_rfft\_instance\_q15 \**S*, uint32\_t *fftLenReal*, uint32\_t *ifftFlagR*, uint32\_t *bitReverseFlag*)

Initialization function for the Q15 RFFT/RIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : *fftLenReal* is not a supported length

**Details** The parameter *fftLenReal* specifies length of RFFT/RIFFT Process. Supported FFT Lengths are 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192.

The parameter *ifftFlagR* controls whether a forward or inverse transform is computed. Set(=1) *ifftFlagR* to calculate RIFFT, otherwise RFFT is calculated.

The parameter `bitReverseFlag` controls whether output is in normal order or bit reversed order. Set(=1) `bitReverseFlag` for output to be in normal order otherwise output is in bit reversed order.

This function also initializes Twiddle factor table.

#### Parameters

- [inout] `S`: points to an instance of the Q15 RFFT/RIFFT structure
- [in] `fftLenReal`: length of the FFT
- [in] `ifftFlagR`: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] `bitReverseFlag`: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

riscv\_status **riscv\_rfft\_init\_q31** (riscv\_rfft\_instance\_q31 \**S*, uint32\_t *fftLenReal*, uint32\_t *ifftFlagR*, uint32\_t *bitReverseFlag*)

Initialization function for the Q31 RFFT/RIFFT.

**Return** execution status

- RISC\_V\_MATH\_SUCCESS : Operation successful
- RISC\_V\_MATH\_ARGUMENT\_ERROR : `fftLenReal` is not a supported length

**Details** The parameter `fftLenReal` specifies length of RFFT/RIFFT Process. Supported FFT Lengths are 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192.

The parameter `ifftFlagR` controls whether a forward or inverse transform is computed. Set(=1) `ifftFlagR` to calculate RIFFT, otherwise RFFT is calculated.

The parameter `bitReverseFlag` controls whether output is in normal order or bit reversed order. Set(=1) `bitReverseFlag` for output to be in normal order otherwise output is in bit reversed order.

This function also initializes Twiddle factor table.

#### Parameters

- [inout] `S`: points to an instance of the Q31 RFFT/RIFFT structure
- [in] `fftLenReal`: length of the FFT
- [in] `ifftFlagR`: flag that selects transform direction
  - value = 0: forward transform
  - value = 1: inverse transform
- [in] `bitReverseFlag`: flag that enables / disables bit reversal of output
  - value = 0: disables bit reversal of output
  - value = 1: enables bit reversal of output

void **riscv\_rfft\_q15** (const riscv\_rfft\_instance\_q15 \**S*, q15\_t \**pSrc*, q15\_t \**pDst*)

Processing function for the Q15 RFFT/RIFFT.

**Return** none

**Input an output formats** Internally input is downscaled by 2 for every stage to avoid saturations inside CFFT/CIFFT process. Hence the output format is different for different RFFT sizes. The input and output formats for different RFFT sizes and number of bits to upscale are mentioned in the tables below for RFFT and RIFFT:

RFFT Size	Input Format	Output Format	Number of bits to upscale
32	1.15	5.11	4
64	1.15	6.10	5
128	1.15	7.9	6
256	1.15	8.8	7
512	1.15	9.7	8
1024	1.15	10.6	9
2048	1.15	11.5	10
4096	1.15	12.4	11
8192	1.15	13.3	12

RFFT Size	Input Format	Output Format	Number of bits to upscale
32	1.15	5.11	0
64	1.15	6.10	0
128	1.15	7.9	0
256	1.15	8.8	0
512	1.15	9.7	0
1024	1.15	10.6	0
2048	1.15	11.5	0
4096	1.15	12.4	0
8192	1.15	13.3	0

If the input buffer is of length N, the output buffer must have length 2\*N. The input buffer is modified by this function.

For the RIFFT, the source buffer must at least have length  $\text{fftLenReal} + 2$ . The last two elements must be equal to what would be generated by the RFFT:  $(\text{pSrc}[0] - \text{pSrc}[1]) \gg 1$  and 0

#### Parameters

- [in] S: points to an instance of the Q15 RFFT/RIFFT structure
- [in] pSrc: points to input buffer (Source buffer is modified by this function.)
- [out] pDst: points to output buffer

void **riscv\_rfft\_q31** (const riscv\_rfft\_instance\_q31 \*S, q31\_t \*pSrc, q31\_t \*pDst)

Processing function for the Q31 RFFT/RIFFT.

**Return** none

**Input an output formats** Internally input is downscaled by 2 for every stage to avoid saturations inside CFFT/CIFFT process. Hence the output format is different for different RFFT sizes. The input and

output formats for different RFFT sizes and number of bits to upscale are mentioned in the tables below for RFFT and RIFFT:

RFFT Size	Input Format	Output Format	Number of bits to upscale
32	1.31	5.27	4
64	1.31	6.26	5
128	1.31	7.25	6
256	1.31	8.24	7
512	1.31	9.23	8
1024	1.31	10.22	9
2048	1.31	11.21	10
4096	1.31	21.20	11
8192	1.31	13.19	12

RFFT Size	Input Format	Output Format	Number of bits to upscale
32	1.31	5.27	0
64	1.31	6.26	0
128	1.31	7.25	0
256	1.31	8.24	0
512	1.31	9.23	0
1024	1.31	10.22	0
2048	1.31	11.21	0
4096	1.31	12.20	0
8192	1.31	13.19	0

If the input buffer is of length  $N$ , the output buffer must have length  $2*N$ . The input buffer is modified by this function.

For the RIFFT, the source buffer must at least have length  $\text{fftLenReal} + 2$ . The last two elements must be equal to what would be generated by the RFFT:  $(\text{pSrc}[0] - \text{pSrc}[1]) \gg 1$  and 0

#### Parameters

- [in] *S*: points to an instance of the Q31 RFFT/RIFFT structure
- [in] *pSrc*: points to input buffer (Source buffer is modified by this function)
- [out] *pDst*: points to output buffer

*group* **groupTransforms**

## 3.4 Changelog

### 3.4.1 V1.0.2

This is release 1.0.2 version of NMSIS-DSP library.

- Sync up to CMSIS DSP library 1.9.0
- Adding initial support for RISC-V vector extension support
- **Caution:** `riscv_math.h` is separated into several header files. `Extra PrivateInclude` folder is included as header folder.

### 3.4.2 V1.0.1

This is release V1.0.1 version of NMSIS-DSP library.

- Both Nuclei RISC-V 32 and 64 bit cores are supported now.
- Libraries are optimized for RISC-V 32 and 64 bit DSP instructions.
- The NN examples are now using Nuclei SDK as running environment.

### 3.4.3 V1.0.0

This is the first version of NMSIS-DSP library.

We adapt the CMSIS-DSP v1.6.0 library to use RISC-V DSP instructions, all the API names now are renamed from `arm_XXX` to `riscv_XXX`.



## 4.1 Overview

### 4.1.1 Introduction

This user manual describes the NMSIS NN software library, a collection of efficient neural network kernels developed to maximize the performance and minimize the memory footprint of neural networks on Nuclei N/NX Class Processors cores.

The library is divided into a number of functions each covering a specific category:

- Neural Network Convolution Functions
- Neural Network Activation Functions
- Fully-connected Layer Functions
- Neural Network Pooling Functions
- Softmax Functions
- Neural Network Support Functions

The library has separate functions for operating on different weight and activation data types including 8-bit integers (q7\_t) and 16-bit integers (q15\_t). The description of the kernels are included in the function description.

The implementation details are also described in this paper [CMSIS-NN: Efficient Neural Network Kernels for Arm Cortex-M CPUs<sup>23</sup>](#).

### 4.1.2 Block Diagram

### 4.1.3 Examples

The library ships with a number of examples which demonstrate how to use the library functions.

- *Convolutional Neural Network Example* (page 733)
- *Gated Recurrent Unit Example* (page 734)

---

<sup>23</sup> <https://arxiv.org/abs/1801.06601>

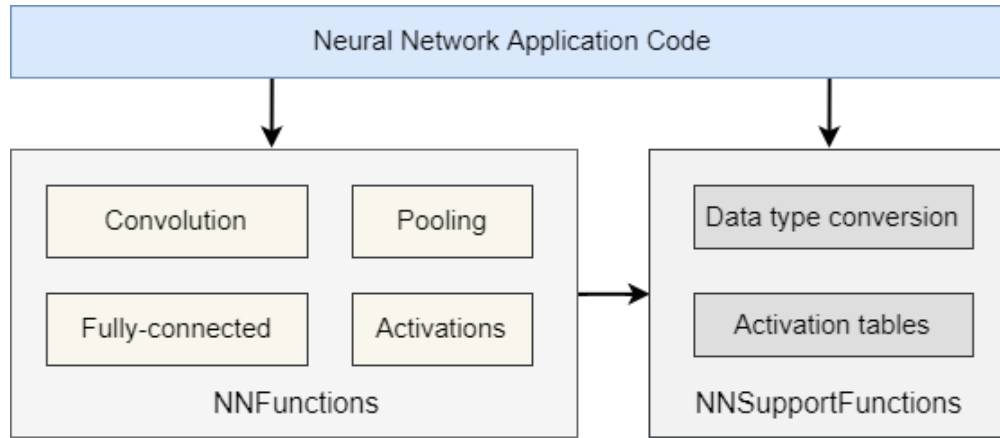


Fig. 1: NMSIS NN Block Diagram

#### 4.1.4 Pre-processor Macros

Each library project have different pre-processor macros.

This library is only built for little endian targets.

**RISCV\_MATH\_DSP:** Define macro `RISCV_MATH_DSP`, If the silicon supports DSP instructions.

**RISCV\_NN\_TRUNCATE:** Define macro `RISCV_NN_TRUNCATE` to use floor instead of round-to-the-nearest-int for the computation.

## 4.2 Using NMSIS-NN

Here we will describe how to run the nmsis nn examples in Nuclei Spike.

### 4.2.1 Preparation

- Nuclei Modified Spike - `xl_spike`
- Nuclei SDK modified for `xl_spike` branch `dev_xlspike`
- Nuclei RISC-V GNU Toolchain, `riscv gcc 10` with vector extension is required.
- CMake `>= 3.5`

### 4.2.2 Tool Setup

1. Export **PATH** correctly for `xl_spike` and `riscv-nuclei-elf-gcc`

```
export PATH=/path/to/xl_spike/bin:/path/to/riscv-nuclei-elf-gcc/bin/:$PATH
```

### 4.2.3 Build NMSIS NN Library

1. Download or clone NMSIS source code into **NMSIS** directory.
2. cd to `NMSIS/NMSIS/` directory

3. Build NMSIS NN library using `make gen_nn_lib`
4. Strip debug informations using `make strip_nn_lib` to make the generated library smaller
5. The nn library will be generated into `./Library/NN/GCC` folder
6. The nn libraries will be look like this:

```
$ ls -lh Library/NN/GCC/
total 13M
-rw-rw-r-- 1 hqfang nucleisys 425K Aug  6 16:10 libnmsis_nn_rv32imac.a
-rw-rw-r-- 1 hqfang nucleisys 743K Aug  6 16:10 libnmsis_nn_rv32imacp.a
-rw-rw-r-- 1 hqfang nucleisys 425K Aug  6 16:10 libnmsis_nn_rv32imafc.a
-rw-rw-r-- 1 hqfang nucleisys 743K Aug  6 16:10 libnmsis_nn_rv32imafcp.a
-rw-rw-r-- 1 hqfang nucleisys 425K Aug  6 16:10 libnmsis_nn_rv32imafdc.a
-rw-rw-r-- 1 hqfang nucleisys 743K Aug  6 16:10 libnmsis_nn_rv32imafdcv.a
-rw-rw-r-- 1 hqfang nucleisys 605K Aug  6 16:10 libnmsis_nn_rv64imac.a
-rw-rw-r-- 1 hqfang nucleisys 1.1M Aug  6 16:10 libnmsis_nn_rv64imacp.a
-rw-rw-r-- 1 hqfang nucleisys 606K Aug  6 16:10 libnmsis_nn_rv64imafc.a
-rw-rw-r-- 1 hqfang nucleisys 1.1M Aug  6 16:10 libnmsis_nn_rv64imafcp.a
-rw-rw-r-- 1 hqfang nucleisys 1.1M Aug  6 16:10 libnmsis_nn_rv64imafcpv.a
-rw-rw-r-- 1 hqfang nucleisys 931K Aug  6 16:10 libnmsis_nn_rv64imafcv.a
-rw-rw-r-- 1 hqfang nucleisys 606K Aug  6 16:10 libnmsis_nn_rv64imafdc.a
-rw-rw-r-- 1 hqfang nucleisys 1.1M Aug  6 16:10 libnmsis_nn_rv64imafdcv.a
-rw-rw-r-- 1 hqfang nucleisys 1.1M Aug  6 16:10 libnmsis_nn_rv64imafdcvp.a
-rw-rw-r-- 1 hqfang nucleisys 932K Aug  6 16:10 libnmsis_nn_rv64imafdcvp.a
```

7. library name with extra `p` is build with RISC-V DSP enabled.
  - `libnmsis_nn_rv32imac.a`: Build for **RISCV\_ARCH=rv32imac** without DSP enabled.
  - `libnmsis_nn_rv32imacp.a`: Build for **RISCV\_ARCH=rv32imac** with DSP enabled.
8. library name with extra `v` is build with RISC-V Vector enabled, only valid for RISC-V 64bit processor.
  - `libnmsis_nn_rv64imac.a`: Build for **RISCV\_ARCH=rv64imac** without Vector.
  - `libnmsis_nn_rv64imacv.a`: Build for **RISCV\_ARCH=rv64imac** with Vector enabled.

---

#### Note:

- You can also directly build both DSP and NN library using `make gen`
  - You can strip the generated DSP and NN library using `make strip`
  - DSP and Vector extension can be combined, such as `p`, `v` and `pv`
  - Vector extension currently is only available with RISC-V 64 bit processor
- 

## 4.2.4 How to run

1. Set environment variables `NUCLEI_SDK_ROOT` and `NUCLEI_SDK_NMSIS`, and set Nuclei SDK SoC to `xl-spike`

```
export NUCLEI_SDK_ROOT=/path/to/nuclei_sdk
export NUCLEI_SDK_NMSIS=/path/to/NMSIS/NMSIS
export SOC=xlspike
```

2. Let us take `./cifar10/` for example

2. `cd ./cifar10/`

3. Run with RISCv DSP enabled and Vector enabled NMSIS-NN library for CORE ux600

```
# Clean project
make DSP_ENABLE=ON VECTOR_ENABLE=ON CORE=ux600 clean
# Build project
make DSP_ENABLE=ON VECTOR_ENABLE=ON CORE=ux600 all
# Run application using xl_spike
make DSP_ENABLE=ON VECTOR_ENABLE=ON CORE=ux600 run
```

4. Run with RISCv DSP disabled and Vector disabled NMSIS-NN library for CORE ux600

```
make DSP_ENABLE=OFF VECTOR_ENABLE=OFF CORE=ux600 clean
make DSP_ENABLE=OFF VECTOR_ENABLE=OFF CORE=ux600 all
make DSP_ENABLE=OFF VECTOR_ENABLE=OFF CORE=ux600 run
```

---

**Note:**

- You can easily run this example in your hardware, if you have enough memory to run it, just modify the SOC to the one you are using in step 1.
- 

## 4.3 NMSIS NN API

If you want to access doxygen generated NMSIS NN API, please click [NMSIS NN API Doxygen Documentation](#).

### 4.3.1 Neural Network Functions

#### Activation Functions

```
void riscv_nn_activations_direct_q15 (q15_t *data, uint16_t size, uint16_t int_width,
                                     riscv_nn_activation_type type)
```

```
void riscv_nn_activations_direct_q7 (q7_t *data, uint16_t size, uint16_t int_width,
                                     riscv_nn_activation_type type)
```

```
void riscv_relu6_s8 (q7_t *data, uint16_t size)
```

```
void riscv_relu_q15 (q15_t *data, uint16_t size)
```

```
void riscv_relu_q7 (q7_t *data, uint16_t size)
```

*group* **Acti**

Perform activation layers, including ReLU (Rectified Linear Unit), sigmoid and tanh

#### Functions

```
void riscv_nn_activations_direct_q15 (q15_t *data, uint16_t size, uint16_t int_width,
                                     riscv_nn_activation_type type)
```

neural network activation function using direct table look-up

Q15 neural network activation function using direct table look-up.

**Note** Refer header file for details.

void **riscv\_nn\_activations\_direct\_q7** (q7\_t \*data, uint16\_t size, uint16\_t int\_width,  
riscv\_nn\_activation\_type type)  
Q7 neural network activation function using direct table look-up.

This is the direct table look-up approach.

#### Parameters

- [inout] data: pointer to input
- [in] size: number of elements
- [in] int\_width: bit-width of the integer part, assume to be smaller than 3
- [in] type: type of activation functions

Assume here the integer part of the fixed-point is  $\leq 3$ . More than 3 just not making much sense, makes no difference with saturation followed by any of these activation functions.

void **riscv\_relu6\_s8** (q7\_t \*data, uint16\_t size)  
s8 ReLU6 function

#### Parameters

- [inout] data: pointer to input
- [in] size: number of elements

void **riscv\_relu\_q15** (q15\_t \*data, uint16\_t size)  
Q15 RELU function.

Optimized relu with QSUB instructions.

#### Parameters

- [inout] data: pointer to input
- [in] size: number of elements

void **riscv\_relu\_q7** (q7\_t \*data, uint16\_t size)  
Q7 RELU function.

Optimized relu with QSUB instructions.

#### Parameters

- [inout] data: pointer to input
- [in] size: number of elements

### Basic math functions

riscv\_status **riscv\_elementwise\_add\_s8** (const int8\_t \*input\_1\_vect, const int8\_t \*input\_2\_vect,  
const int32\_t input\_1\_offset, const int32\_t input\_1\_mult,  
const int32\_t input\_1\_shift, const int32\_t input\_2\_offset,  
const int32\_t input\_2\_mult, const int32\_t input\_2\_shift,  
const int32\_t left\_shift, int8\_t \*output, const int32\_t  
out\_offset, const int32\_t out\_mult, const int32\_t  
out\_shift, const int32\_t out\_activation\_min, const  
int32\_t out\_activation\_max, const uint32\_t block\_size)

```
riscv_status riscv_elementwise_mul_s8(const int8_t *input_1_vect, const int8_t *input_2_vect,  
                                     const int32_t input_1_offset, const int32_t input_2_offset,  
                                     int8_t *output, const int32_t out_offset,  
                                     const int32_t out_mult, const int32_t out_shift,  
                                     const int32_t out_activation_min, const int32_t  
                                     out_activation_max, const uint32_t block_size)
```

group **BasicMath**

Element wise add and multiplication functions.

## Functions

```
riscv_status riscv_elementwise_add_s8(const int8_t *input_1_vect, const int8_t *input_2_vect,  
                                     const int32_t input_1_offset, const int32_t input_1_mult,  
                                     const int32_t input_1_shift, const int32_t input_2_offset,  
                                     const int32_t input_2_mult, const int32_t input_2_shift, const  
                                     int32_t left_shift, int8_t *output, const int32_t  
                                     out_offset, const int32_t out_mult, const int32_t  
                                     out_shift, const int32_t out_activation_min,  
                                     const int32_t out_activation_max, const uint32_t  
                                     block_size)
```

s8 element wise add of two vectors

**Return** The function returns RISCVMATH\_SUCCESS

### Parameters

- [in] input\_1\_vect: pointer to input vector 1
- [in] input\_2\_vect: pointer to input vector 2
- [in] input\_1\_offset: offset for input 1. Range: Range: -127 to 128
- [in] input\_1\_mult: multiplier for input 1
- [in] input\_1\_shift: shift for input 1
- [in] input\_2\_offset: offset for input 2. Range: Range: -127 to 128
- [in] input\_2\_mult: multiplier for input 2
- [in] input\_2\_shift: shift for input 2
- [in] left\_shift: input left shift
- [inout] output: pointer to output vector
- [in] out\_offset: output offset
- [in] out\_mult: output multiplier
- [in] out\_shift: output shift
- [in] out\_activation\_min: minimum value to clamp output to
- [in] out\_activation\_max: maximum value to clamp output to
- [in] block\_size: number of samples

```
riscv_status riscv_elementwise_mul_s8(const int8_t *input_1_vect, const int8_t *input_2_vect, const int32_t input_1_offset, const int32_t input_2_offset, int8_t *output, const int32_t out_offset, const int32_t out_mult, const int32_t out_shift, const int32_t out_activation_min, const int32_t out_activation_max, const uint32_t block_size)
```

s8 element wise multiplication of two vectors

s8 element wise multiplication

**Note** Refer header file for details.

## Concatenation Functions

```
void riscv_concatenation_s8_w(const int8_t *input, const uint16_t input_x, const uint16_t input_y, const uint16_t input_z, const uint16_t input_w, int8_t *output, const uint32_t offset_w)
```

```
void riscv_concatenation_s8_x(const int8_t *input, const uint16_t input_x, const uint16_t input_y, const uint16_t input_z, const uint16_t input_w, int8_t *output, const uint16_t output_x, const uint32_t offset_x)
```

```
void riscv_concatenation_s8_y(const int8_t *input, const uint16_t input_x, const uint16_t input_y, const uint16_t input_z, const uint16_t input_w, int8_t *output, const uint16_t output_y, const uint32_t offset_y)
```

```
void riscv_concatenation_s8_z(const int8_t *input, const uint16_t input_x, const uint16_t input_y, const uint16_t input_z, const uint16_t input_w, int8_t *output, const uint16_t output_z, const uint32_t offset_z)
```

group **Concatenation**

## Functions

```
void riscv_concatenation_s8_w(const int8_t *input, const uint16_t input_x, const uint16_t input_y, const uint16_t input_z, const uint16_t input_w, int8_t *output, const uint32_t offset_w)
```

int8/uint8 concatenation function to be used for concatenating N-tensors along the W axis (Batch size)  
This function should be called for each input tensor to concatenate. The argument offset\_w will be used to store the input tensor in the correct position in the output tensor

i.e. offset\_w = 0 for(i = 0 i < num\_input\_tensors; ++i) { riscv\_concatenation\_s8\_w(&input[i], ..., &output, ..., ..., offset\_w) offset\_w += input\_w[i] }

This function assumes that the output tensor has:

1. The same width of the input tensor
2. The same height of the input tensor
3. The same number o channels of the input tensor

Unless specified otherwise, arguments are mandatory.

**Note** This function, data layout independent, can be used to concatenate either int8 or uint8 tensors because it does not involve any arithmetic operation

**Parameters**

- [in] `input`: Pointer to input tensor
- [in] `input_x`: Width of input tensor
- [in] `input_y`: Height of input tensor
- [in] `input_z`: Channels in input tensor
- [in] `input_w`: Batch size in input tensor
- [out] `output`: Pointer to output tensor
- [in] `offset_w`: The offset on the W axis to start concatenating the input tensor It is user responsibility to provide the correct value

```
void riscv_concatenation_s8_x(const int8_t *input, const uint16_t input_x, const
                             uint16_t input_y, const uint16_t input_z, const uint16_t
                             input_w, int8_t *output, const uint16_t output_x, const
                             uint32_t offset_x)
```

int8/uint8 concatenation function to be used for concatenating N-tensors along the X axis This function should be called for each input tensor to concatenate. The argument `offset_x` will be used to store the input tensor in the correct position in the output tensor

i.e. `offset_x = 0` for `(i = 0 i < num_input_tensors; ++i) { riscv_concatenation_s8_x(&input[i], ..., &output, ..., ..., offset_x) offset_x += input_x[i] }`

This function assumes that the output tensor has:

1. The same height of the input tensor
2. The same number of channels of the input tensor
3. The same batch size of the input tensor

Unless specified otherwise, arguments are mandatory.

**Input constraints** `offset_x` is less than `output_x`

**Note** This function, data layout independent, can be used to concatenate either int8 or uint8 tensors because it does not involve any arithmetic operation

#### Parameters

- [in] `input`: Pointer to input tensor
- [in] `input_x`: Width of input tensor
- [in] `input_y`: Height of input tensor
- [in] `input_z`: Channels in input tensor
- [in] `input_w`: Batch size in input tensor
- [out] `output`: Pointer to output tensor
- [in] `output_x`: Width of output tensor
- [in] `offset_x`: The offset (in number of elements) on the X axis to start concatenating the input tensor It is user responsibility to provide the correct value

```
void riscv_concatenation_s8_y(const int8_t *input, const uint16_t input_x, const
                             uint16_t input_y, const uint16_t input_z, const uint16_t
                             input_w, int8_t *output, const uint16_t output_y, const
                             uint32_t offset_y)
```

int8/uint8 concatenation function to be used for concatenating N-tensors along the Y axis This function



should be called for each input tensor to concatenate. The argument `offset_y` will be used to store the input tensor in the correct position in the output tensor

i.e. `offset_y = 0` for `(i = 0 i < num_input_tensors; ++i) { riscv_concatenation_s8_y(&input[i], ..., &output, ..., ..., offset_y) offset_y += input_y[i] }`

This function assumes that the output tensor has:

1. The same width of the input tensor
2. The same number of channels of the input tensor
3. The same batch size of the input tensor

Unless specified otherwise, arguments are mandatory.

**Input constraints** `offset_y` is less than `output_y`

**Note** This function, data layout independent, can be used to concatenate either `int8` or `uint8` tensors because it does not involve any arithmetic operation

#### Parameters

- `[in] input`: Pointer to input tensor
- `[in] input_x`: Width of input tensor
- `[in] input_y`: Height of input tensor
- `[in] input_z`: Channels in input tensor
- `[in] input_w`: Batch size in input tensor
- `[out] output`: Pointer to output tensor
- `[in] output_y`: Height of output tensor
- `[in] offset_y`: The offset on the Y axis to start concatenating the input tensor It is user responsibility to provide the correct value

```
void riscv_concatenation_s8_z(const int8_t *input, const uint16_t input_x, const
                             uint16_t input_y, const uint16_t input_z, const uint16_t
                             input_w, int8_t *output, const uint16_t output_z, const
                             uint32_t offset_z)
```

`int8/uint8` concatenation function to be used for concatenating N-tensors along the Z axis This function should be called for each input tensor to concatenate. The argument `offset_z` will be used to store the input tensor in the correct position in the output tensor

i.e. `offset_z = 0` for `(i = 0 i < num_input_tensors; ++i) { riscv_concatenation_s8_z(&input[i], ..., &output, ..., ..., offset_z) offset_z += input_z[i] }`

This function assumes that the output tensor has:

1. The same width of the input tensor
2. The same height of the input tensor
3. The same batch size of the input tensor

Unless specified otherwise, arguments are mandatory.

**Input constraints** `offset_z` is less than `output_z`

**Note** This function, data layout independent, can be used to concatenate either `int8` or `uint8` tensors because it does not involve any arithmetic operation

#### Parameters

- [in] `input`: Pointer to input tensor
- [in] `input_x`: Width of input tensor
- [in] `input_y`: Height of input tensor
- [in] `input_z`: Channels in input tensor
- [in] `input_w`: Batch size in input tensor
- [out] `output`: Pointer to output tensor
- [in] `output_z`: Channels in output tensor
- [in] `offset_z`: The offset on the Z axis to start concatenating the input tensor It is user responsibility to provide the correct value

## Convolution Functions

```
riscv_status riscv_convolve_1_x_n_s8 (const nmsis_nn_context *ctx, const nmsis_nn_conv_params *conv_params, const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

```
int32_t riscv_convolve_1_x_n_s8_get_buffer_size (const nmsis_nn_dims *input_dims, const nmsis_nn_dims *filter_dims)
```

```
riscv_status riscv_convolve_1x1_HWC_q7_fast_nonsquare (const q7_t *Im_in, const uint16_t dim_im_in_x, const uint16_t dim_im_in_y, const uint16_t ch_im_in, const q7_t *wt, const uint16_t ch_im_out, const uint16_t dim_kernel_x, const uint16_t dim_kernel_y, const uint16_t padding_x, const uint16_t padding_y, const uint16_t stride_x, const uint16_t stride_y, const q7_t *bias, const uint16_t bias_shift, const uint16_t out_shift, q7_t *Im_out, const uint16_t dim_im_out_x, const uint16_t dim_im_out_y, q15_t *bufferA, q7_t *bufferB)
```

```
riscv_status riscv_convolve_1x1_s8_fast (const nmsis_nn_context *ctx, const nmsis_nn_conv_params *conv_params, const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

```
int32_t riscv_convolve_1x1_s8_fast_get_buffer_size (const nmsis_nn_dims *input_dims)
```

```

riscv_status riscv_convolve_HWC_q15_basic(const q15_t *Im_in, const uint16_t dim_im_in,
                                          const uint16_t ch_im_in, const q15_t *wt, const
                                          uint16_t ch_im_out, const uint16_t dim_kernel,
                                          const uint16_t padding, const uint16_t stride,
                                          const q15_t *bias, const uint16_t bias_shift,
                                          const uint16_t out_shift, q15_t *Im_out, const
                                          uint16_t dim_im_out, q15_t *bufferA, q7_t *bufferB)

riscv_status riscv_convolve_HWC_q15_fast(const q15_t *Im_in, const uint16_t dim_im_in,
                                          const uint16_t ch_im_in, const q15_t *wt, const
                                          uint16_t ch_im_out, const uint16_t dim_kernel,
                                          const uint16_t padding, const uint16_t stride,
                                          const q15_t *bias, const uint16_t bias_shift, const
                                          uint16_t out_shift, q15_t *Im_out, const uint16_t
                                          dim_im_out, q15_t *bufferA, q7_t *bufferB)

riscv_status riscv_convolve_HWC_q15_fast_nonsquare(const q15_t *Im_in, const
                                                    uint16_t dim_im_in_x, const
                                                    uint16_t dim_im_in_y, const
                                                    uint16_t ch_im_in, const q15_t
                                                    *wt, const uint16_t ch_im_out,
                                                    const uint16_t dim_kernel_x, const
                                                    uint16_t dim_kernel_y, const uint16_t
                                                    padding_x, const uint16_t padding_y,
                                                    const uint16_t stride_x, const
                                                    uint16_t stride_y, const q15_t *bias,
                                                    const uint16_t bias_shift, const
                                                    uint16_t out_shift, q15_t *Im_out,
                                                    const uint16_t dim_im_out_x, const
                                                    uint16_t dim_im_out_y, q15_t *bufferA,
                                                    q7_t *bufferB)

riscv_status riscv_convolve_HWC_q7_basic(const q7_t *Im_in, const uint16_t dim_im_in,
                                          const uint16_t ch_im_in, const q7_t *wt, const
                                          uint16_t ch_im_out, const uint16_t dim_kernel,
                                          const uint16_t padding, const uint16_t stride,
                                          const q7_t *bias, const uint16_t bias_shift, const
                                          uint16_t out_shift, q7_t *Im_out, const uint16_t
                                          dim_im_out, q15_t *bufferA, q7_t *bufferB)

riscv_status riscv_convolve_HWC_q7_basic_nonsquare(const q7_t *Im_in, const uint16_t
                                                    dim_im_in_x, const uint16_t
                                                    dim_im_in_y, const uint16_t
                                                    ch_im_in, const q7_t *wt, const
                                                    uint16_t ch_im_out, const uint16_t
                                                    dim_kernel_x, const uint16_t
                                                    dim_kernel_y, const uint16_t
                                                    padding_x, const uint16_t padding_y,
                                                    const uint16_t stride_x, const
                                                    uint16_t stride_y, const q7_t *bias,
                                                    const uint16_t bias_shift, const
                                                    uint16_t out_shift, q7_t *Im_out, const
                                                    uint16_t dim_im_out_x, const uint16_t
                                                    dim_im_out_y, q15_t *bufferA, q7_t
                                                    *bufferB)

```

```
riscv_status riscv_convolve_HWC_q7_fast (const q7_t *Im_in, const uint16_t dim_im_in, const
    uint16_t ch_im_in, const q7_t *wt, const uint16_t
    ch_im_out, const uint16_t dim_kernel, const uint16_t
    padding, const uint16_t stride, const q7_t *bias,
    const uint16_t bias_shift, const uint16_t out_shift,
    q7_t *Im_out, const uint16_t dim_im_out, q15_t
    *bufferA, q7_t *bufferB)

riscv_status riscv_convolve_HWC_q7_fast_nonsquare (const q7_t *Im_in, const uint16_t
    dim_im_in_x, const uint16_t
    dim_im_in_y, const uint16_t ch_im_in,
    const q7_t *wt, const uint16_t
    ch_im_out, const uint16_t dim_kernel_x,
    const uint16_t dim_kernel_y, const
    uint16_t padding_x, const uint16_t
    padding_y, const uint16_t stride_x,
    const uint16_t stride_y, const q7_t
    *bias, const uint16_t bias_shift, const
    uint16_t out_shift, q7_t *Im_out, const
    uint16_t dim_im_out_x, const uint16_t
    dim_im_out_y, q15_t *bufferA, q7_t
    *bufferB)

riscv_status riscv_convolve_HWC_q7_RGB (const q7_t *Im_in, const uint16_t dim_im_in, const
    uint16_t ch_im_in, const q7_t *wt, const uint16_t
    ch_im_out, const uint16_t dim_kernel, const uint16_t
    padding, const uint16_t stride, const q7_t *bias,
    const uint16_t bias_shift, const uint16_t out_shift, q7_t
    *Im_out, const uint16_t dim_im_out, q15_t *bufferA, q7_t
    *bufferB)

riscv_status riscv_convolve_s8 (const nmsis_nn_context *ctx, const nmsis_nn_conv_params
    *conv_params, const nmsis_nn_per_channel_quant_params
    *quant_params, const nmsis_nn_dims *input_dims, const q7_t
    *input_data, const nmsis_nn_dims *filter_dims, const q7_t
    *filter_data, const nmsis_nn_dims *bias_dims, const int32_t
    *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)

int32_t riscv_convolve_s8_get_buffer_size (const nmsis_nn_dims *input_dims, const nm-
    sis_nn_dims *filter_dims)

riscv_status riscv_convolve_wrapper_s8 (const nmsis_nn_context *ctx, const nm-
    sis_nn_conv_params *conv_params, const nm-
    sis_nn_per_channel_quant_params *quant_params,
    const nmsis_nn_dims *input_dims, const q7_t *in-
    put_data, const nmsis_nn_dims *filter_dims, const q7_t
    *filter_data, const nmsis_nn_dims *bias_dims, const
    int32_t *bias_data, const nmsis_nn_dims *output_dims,
    q7_t *output_data)

int32_t riscv_convolve_wrapper_s8_get_buffer_size (const nmsis_nn_conv_params
    *conv_params, const nmsis_nn_dims
    *input_dims, const nmsis_nn_dims
    *filter_dims, const nmsis_nn_dims
    *output_dims)
```

```

riscv_status riscv_depthwise_conv_3x3_s8(const nmsis_nn_context *ctx, const nmsis_nn_dw_conv_params *dw_conv_params, const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)

static void depthwise_conv_s8_mult_4(const int8_t *input, const int32_t input_x, const int32_t input_y, const int32_t input_ch, const int8_t *kernel, const int32_t output_ch, const int32_t ch_mult, const int32_t kernel_x, const int32_t kernel_y, const int32_t pad_x, const int32_t pad_y, const int32_t stride_x, const int32_t stride_y, const int32_t *bias, int8_t *output, const int32_t *output_shift, const int32_t *output_mult, const int32_t output_x, const int32_t output_y, const int32_t output_offset, const int32_t input_offset, const int32_t output_activation_min, const int32_t output_activation_max)

static void depthwise_conv_s8_generic(const q7_t *input, const uint16_t input_batches, const uint16_t input_x, const uint16_t input_y, const uint16_t input_ch, const q7_t *kernel, const uint16_t output_ch, const uint16_t ch_mult, const uint16_t kernel_x, const uint16_t kernel_y, const uint16_t pad_x, const uint16_t pad_y, const uint16_t stride_x, const uint16_t stride_y, const int32_t *bias, q7_t *output, const int32_t *output_shift, const int32_t *output_mult, const uint16_t output_x, const uint16_t output_y, const int32_t output_offset, const int32_t input_offset, const int32_t output_activation_min, const int32_t output_activation_max)

riscv_status riscv_depthwise_conv_s8(const nmsis_nn_context *ctx, const nmsis_nn_dw_conv_params *dw_conv_params, const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)

riscv_status riscv_depthwise_conv_s8_opt(const nmsis_nn_context *ctx, const nmsis_nn_dw_conv_params *dw_conv_params, const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)

int32_t riscv_depthwise_conv_s8_opt_get_buffer_size(const nmsis_nn_dims *input_dims, const nmsis_nn_dims *filter_dims)

```

```
static void depthwise_conv_u8_mult_4(const uint8_t *input, const int32_t input_x, const
int32_t input_y, const int32_t input_ch, const uint8_t
*kernel, const int32_t output_ch, const int32_t
ch_mult, const int32_t kernel_x, const int32_t ker-
nel_y, const int32_t pad_x, const int32_t pad_y,
const int32_t stride_x, const int32_t stride_y, const
int32_t *bias, uint8_t *output, const int32_t output_shift,
const int32_t output_mult, const int32_t output_x,
const int32_t output_y, const int32_t output_offset,
const int32_t input_offset, const int32_t filter_offset,
const int32_t output_activation_min, const int32_t out-
put_activation_max)

static void depthwise_conv_u8_generic(const uint8_t *input, const int32_t input_x, const
int32_t input_y, const int32_t input_ch, const
uint8_t *kernel, const int32_t output_ch, const
int32_t ch_mult, const int32_t kernel_x, const
int32_t kernel_y, const int32_t pad_x, const int32_t
pad_y, const int32_t stride_x, const int32_t stride_y,
const int32_t *bias, uint8_t *output, const int32_t
output_shift, const int32_t output_mult, const int32_t
output_x, const int32_t output_y, const int32_t out-
put_offset, const int32_t input_offset, const int32_t
filter_offset, const int32_t output_activation_min,
const int32_t output_activation_max)

riscv_status riscv_depthwise_conv_u8_basic_ver1(const uint8_t *input, const uint16_t
input_x, const uint16_t input_y, const
uint16_t input_ch, const uint8_t *kernel,
const uint16_t kernel_x, const uint16_t
kernel_y, const int16_t ch_mult, const
int16_t pad_x, const int16_t pad_y, const
int16_t stride_x, const int16_t stride_y,
const int16_t dilation_x, const int16_t
dilation_y, const int32_t *bias, const
int32_t input_offset, const int32_t fil-
ter_offset, const int32_t output_offset,
uint8_t *output, const uint16_t output_x,
const uint16_t output_y, const int32_t
output_activation_min, const int32_t
output_activation_max, const int32_t
output_shift, const int32_t output_mult)

riscv_status riscv_depthwise_conv_wrapper_s8(const nmsis_nn_context *ctx, const nm-
sis_nn_dw_conv_params *dw_conv_params,
const nmsis_nn_per_channel_quant_params
*quant_params, const nmsis_nn_dims *in-
put_dims, const q7_t *input_data, const
nmsis_nn_dims *filter_dims, const q7_t *fil-
ter_data, const nmsis_nn_dims *bias_dims,
const int32_t *bias_data, const nm-
sis_nn_dims *output_dims, q7_t *output_data)
```

```

int32_t riscv_depthwise_conv_wrapper_s8_get_buffer_size(const          nmsis_nn_dw_conv_params
                                                         *dw_conv_params,  const
nmsis_nn_dims  *input_dims,
const          nmsis_nn_dims
*filter_dims,  const  nmsis_nn_dims *output_dims)

riscv_status riscv_depthwise_separable_conv_HWC_q7(const q7_t *Im_in, const uint16_t
dim_im_in, const uint16_t ch_im_in,
const q7_t *wt, const uint16_t
ch_im_out, const uint16_t dim_kernel,
const uint16_t padding, const
uint16_t stride, const q7_t *bias,
const uint16_t bias_shift, const
uint16_t out_shift, q7_t *Im_out, const
uint16_t dim_im_out, q15_t *bufferA,
q7_t *bufferB)

riscv_status riscv_depthwise_separable_conv_HWC_q7_nonsquare(const q7_t *Im_in,
const          uint16_t
dim_im_in_x,  const
uint16_t      dim_im_in_y,
const uint16_t ch_im_in,
const q7_t *wt, const
uint16_t      ch_im_out,
const          uint16_t
dim_kernel_x,  const
uint16_t      dim_kernel_y,
const          uint16_t
padding_x,     const
uint16_t      padding_y,
const uint16_t stride_x,
const uint16_t stride_y,
const q7_t *bias,
const          uint16_t
bias_shift,    const
uint16_t out_shift, q7_t
*Im_out, const uint16_t
dim_im_out_x,  const
uint16_t      dim_im_out_y,
q15_t *bufferA, q7_t
*bufferB)

```

#### group **NNConv**

Collection of convolution, depthwise convolution functions and their variants.

The convolution is implemented in 2 steps: im2col and GEMM

im2col is a process of converting each patch of image data into a column. After im2col, the convolution is computed as matrix-matrix multiplication.

To reduce the memory footprint, the im2col is performed partially. Each iteration, only a few column (i.e., patches) are generated and computed with GEMM kernels similar to NMSIS-DSP riscv\_mat\_mult functions.

## Functions

```
riscv_status riscv_convolve_1_x_n_s8(const nmsis_nn_context *ctx, const nmsis_nn_conv_params *conv_params, const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

1xn convolution

- Supported framework : TensorFlow Lite Micro
- The following constraints on the arguments apply
  1. input\_dims->n equals 1
  2. output\_dims->w is a multiple of 4
  3. Explicit constraints(since it is for 1xN convolution) -## input\_dims->h equals 1 -## output\_dims->h equals 1 -## filter\_dims->h equals 1

**Return** The function returns either RISCVC\_MATH\_SIZE\_MISMATCH if argument constraints fail. or, RISCVC\_MATH\_SUCCESS on successful completion.

### Parameters

- [inout] ctx: Function context that contains the additional buffer if required by the function. riscv\_convolve\_1\_x\_n\_s8\_get\_buffer\_size will return the buffer\_size if required
- [in] conv\_params: Convolution parameters (e.g. strides, dilations, pads,...). Range of conv\_params->input\_offset : [-127, 128] Range of conv\_params->output\_offset : [-128, 127]
- [in] quant\_params: Per-channel quantization info. It contains the multiplier and shift values to be applied to each output channel
- [in] input\_dims: Input (activation) tensor dimensions. Format: [N, H, W, C\_IN]
- [in] input\_data: Input (activation) data pointer. Data type: int8
- [in] filter\_dims: Filter tensor dimensions. Format: [C\_OUT, 1, WK, C\_IN] where WK is the horizontal spatial filter dimension
- [in] filter\_data: Filter data pointer. Data type: int8
- [in] bias\_dims: Bias tensor dimensions. Format: [C\_OUT]
- [in] bias\_data: Optional bias data pointer. Data type: int32
- [in] output\_dims: Output tensor dimensions. Format: [N, H, W, C\_OUT]
- [out] output\_data: Output data pointer. Data type: int8

```
int32_t riscv_convolve_1_x_n_s8_get_buffer_size(const nmsis_nn_dims *input_dims, const nmsis_nn_dims *filter_dims)
```

Get the required additional buffer size for 1xn convolution.

**Return** The function returns required buffer size(bytes)

### Parameters

- [in] input\_dims: Input (activation) tensor dimensions. Format: [N, H, W, C\_IN]



- [in] `filter_dims`: Filter tensor dimensions. Format: [C\_OUT, 1, WK, C\_IN] where WK is the horizontal spatial filter dimension

```
riscv_status riscv_convolve_1x1_HWC_q7_fast_nonsquare (const q7_t *Im_in, const
                                                         uint16_t dim_im_in_x, const
                                                         uint16_t dim_im_in_y, const
                                                         uint16_t ch_im_in, const
                                                         q7_t *wt, const uint16_t
                                                         ch_im_out, const uint16_t
                                                         dim_kernel_x, const uint16_t
                                                         dim_kernel_y, const uint16_t
                                                         padding_x, const uint16_t
                                                         padding_y, const uint16_t
                                                         stride_x, const uint16_t
                                                         stride_y, const q7_t *bias,
                                                         const uint16_t bias_shift,
                                                         const uint16_t out_shift,
                                                         q7_t *Im_out, const uint16_t
                                                         dim_im_out_x, const uint16_t
                                                         dim_im_out_y, q15_t *bufferA,
                                                         q7_t *bufferB)
```

Fast Q7 version of 1x1 convolution (non-square shape)

This function is optimized for convolution with 1x1 kernel size (i.e., `dim_kernel_x=1` and `dim_kernel_y=1`). It can be used for the second half of MobileNets [1] after depthwise separable convolution.

**Return** The function returns either `RISCV_MATH_SIZE_MISMATCH` or `RISCV_MATH_SUCCESS` based on the outcome of size checking.

#### Parameters

- [in] `Im_in`: pointer to input tensor
- [in] `dim_im_in_x`: input tensor dimension x
- [in] `dim_im_in_y`: input tensor dimension y
- [in] `ch_im_in`: number of input tensor channels
- [in] `wt`: pointer to kernel weights
- [in] `ch_im_out`: number of filters, i.e., output tensor channels
- [in] `dim_kernel_x`: filter kernel size x
- [in] `dim_kernel_y`: filter kernel size y
- [in] `padding_x`: padding size x
- [in] `padding_y`: padding size y
- [in] `stride_x`: convolution stride x
- [in] `stride_y`: convolution stride y
- [in] `bias`: pointer to bias
- [in] `bias_shift`: amount of left-shift for bias
- [in] `out_shift`: amount of right-shift for output
- [inout] `Im_out`: pointer to output tensor

- [in] dim\_im\_out\_x: output tensor dimension x
- [in] dim\_im\_out\_y: output tensor dimension y
- [inout] bufferA: pointer to buffer space for input
- [inout] bufferB: pointer to buffer space for output

This function is the version with full list of optimization tricks, but with some constraints: ch\_im\_in is multiple of 4 ch\_im\_out is multiple of 2

[1] MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications <https://arxiv.org/abs/1704.04861>

```
riscv_status riscv_convolve_1x1_s8_fast (const nmsis_nn_context *ctx, const nmsis_nn_conv_params *conv_params, const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

Fast s8 version for 1x1 convolution (non-square shape)

- Supported framework : TensorFlow Lite Micro
- The following constraints on the arguments apply
  1. input\_dims->c is a multiple of 4
  2. conv\_params->padding.w = conv\_params->padding.h = 0
  3. conv\_params->stride.w = conv\_params->stride.h = 1

**Return** The function returns either RISCVC\_MATH\_SIZE\_MISMATCH if argument constraints fail. or, RISCVC\_MATH\_SUCCESS on successful completion.

#### Parameters

- [inout] ctx: Function context that contains the additional buffer if required by the function. riscv\_convolve\_1x1\_s8\_fast\_get\_buffer\_size will return the buffer\_size if required
- [in] conv\_params: Convolution parameters (e.g. strides, dilations, pads,...). Range of conv\_params->input\_offset : [-127, 128] Range of conv\_params->output\_offset : [-128, 127]
- [in] quant\_params: Per-channel quantization info. It contains the multiplier and shift values to be applied to each output channel
- [in] input\_dims: Input (activation) tensor dimensions. Format: [N, H, W, C\_IN]
- [in] input\_data: Input (activation) data pointer. Data type: int8
- [in] filter\_dims: Filter tensor dimensions. Format: [C\_OUT, 1, 1, C\_IN]
- [in] filter\_data: Filter data pointer. Data type: int8
- [in] bias\_dims: Bias tensor dimensions. Format: [C\_OUT]
- [in] bias\_data: Optional bias data pointer. Data type: int32
- [in] output\_dims: Output tensor dimensions. Format: [N, H, W, C\_OUT]
- [out] output\_data: Output data pointer. Data type: int8

```
int32_t riscv_convolve_1x1_s8_fast_get_buffer_size(const nmsis_nn_dims *input_dims)
```

Get the required buffer size for riscv\_convolve\_1x1\_s8\_fast.

**Return** The function returns the required buffer size in bytes

**Parameters**

- [in] input\_dims: Input (activation) dimensions

```
riscv_status riscv_convolve_HWC_q15_basic(const q15_t *Im_in, const uint16_t dim_im_in, const uint16_t ch_im_in, const q15_t *wt, const uint16_t ch_im_out, const uint16_t dim_kernel, const uint16_t padding, const uint16_t stride, const q15_t *bias, const uint16_t bias_shift, const uint16_t out_shift, q15_t *Im_out, const uint16_t dim_im_out, q15_t *bufferA, q7_t *bufferB)
```

Basic Q15 convolution function.

**Buffer size:**

**Return** The function returns RISC\_V\_MATH\_SUCCESS

**Parameters**

- [in] Im\_in: pointer to input tensor
- [in] dim\_im\_in: input tensor dimension
- [in] ch\_im\_in: number of input tensor channels
- [in] wt: pointer to kernel weights
- [in] ch\_im\_out: number of filters, i.e., output tensor channels
- [in] dim\_kernel: filter kernel size
- [in] padding: padding sizes
- [in] stride: convolution stride
- [in] bias: pointer to bias
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [inout] Im\_out: pointer to output tensor
- [in] dim\_im\_out: output tensor dimension
- [inout] bufferA: pointer to buffer space for input
- [inout] bufferB: pointer to buffer space for output

bufferA size: ch\_im\_in\*dim\_kernel\*dim\_kernel

bufferB size: 0

This basic version is designed to work for any input tensor and weight dimension.

```
riscv_status riscv_convolve_HWC_q15_fast (const q15_t *Im_in, const uint16_t dim_im_in,  
                                           const uint16_t ch_im_in, const q15_t *wt,  
                                           const uint16_t ch_im_out, const uint16_t  
dim_kernel, const uint16_t padding, const  
uint16_t stride, const q15_t *bias, const  
uint16_t bias_shift, const uint16_t out_shift,  
q15_t *Im_out, const uint16_t dim_im_out, q15_t  
*bufferA, q7_t *bufferB)
```

Fast Q15 convolution function.

**Buffer size:**

**Return** The function returns either RISCVMATH\_SIZE\_MISMATCH or RISCVMATH\_SUCCESS based on the outcome of size checking.

**Parameters**

- [in] Im\_in: pointer to input tensor
- [in] dim\_im\_in: input tensor dimension
- [in] ch\_im\_in: number of input tensor channels
- [in] wt: pointer to kernel weights
- [in] ch\_im\_out: number of filters, i.e., output tensor channels
- [in] dim\_kernel: filter kernel size
- [in] padding: padding sizes
- [in] stride: convolution stride
- [in] bias: pointer to bias
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [inout] Im\_out: pointer to output tensor
- [in] dim\_im\_out: output tensor dimension
- [inout] bufferA: pointer to buffer space for input
- [inout] bufferB: pointer to buffer space for output

bufferA size: 2\*ch\_im\_in\*dim\_kernel\*dim\_kernel

bufferB size: 0

**Input dimension constraints:**

ch\_im\_in is multiple of 2

ch\_im\_out is multiple of 2

```

riscv_status riscv_convolve_HWC_q15_fast_nonsquare (const q15_t *Im_in, const
                                                    uint16_t dim_im_in_x, const
                                                    uint16_t dim_im_in_y, const
                                                    uint16_t ch_im_in, const q15_t
                                                    *wt, const uint16_t ch_im_out,
                                                    const uint16_t dim_kernel_x,
                                                    const uint16_t dim_kernel_y,
                                                    const uint16_t padding_x, const
                                                    uint16_t padding_y, const
                                                    uint16_t stride_x, const uint16_t
                                                    stride_y, const q15_t *bias,
                                                    const uint16_t bias_shift, const
                                                    uint16_t out_shift, q15_t *Im_out,
                                                    const uint16_t dim_im_out_x,
                                                    const uint16_t dim_im_out_y,
                                                    q15_t *bufferA, q7_t *bufferB)

```

Fast Q15 convolution function (non-square shape)

**Buffer size:**

**Return** The function returns either `RISCV_MATH_SIZE_MISMATCH` or `RISCV_MATH_SUCCESS` based on the outcome of size checking.

**Parameters**

- [in] `Im_in`: pointer to input tensor
- [in] `dim_im_in_x`: input tensor dimension x
- [in] `dim_im_in_y`: input tensor dimension y
- [in] `ch_im_in`: number of input tensor channels
- [in] `wt`: pointer to kernel weights
- [in] `ch_im_out`: number of filters, i.e., output tensor channels
- [in] `dim_kernel_x`: filter kernel size x
- [in] `dim_kernel_y`: filter kernel size y
- [in] `padding_x`: padding size x
- [in] `padding_y`: padding size y
- [in] `stride_x`: convolution stride x
- [in] `stride_y`: convolution stride y
- [in] `bias`: pointer to bias
- [in] `bias_shift`: amount of left-shift for bias
- [in] `out_shift`: amount of right-shift for output
- [inout] `Im_out`: pointer to output tensor
- [in] `dim_im_out_x`: output tensor dimension x
- [in] `dim_im_out_y`: output tensor dimension y
- [inout] `bufferA`: pointer to buffer space for input
- [inout] `bufferB`: pointer to buffer space for output

bufferA size:  $2 \times \text{ch\_im\_in} \times \text{dim\_kernel} \times \text{dim\_kernel}$

bufferB size: 0

**Input dimension constraints:**

ch\_im\_in is multiple of 2

ch\_im\_out is multiple of 2

```
riscv_status riscv_convolve_HWC_q7_basic (const q7_t *Im_in, const uint16_t dim_im_in,
                                           const uint16_t ch_im_in, const q7_t *wt,
                                           const uint16_t ch_im_out, const uint16_t
                                           dim_kernel, const uint16_t padding, const
                                           uint16_t stride, const q7_t *bias, const
                                           uint16_t bias_shift, const uint16_t out_shift,
                                           q7_t *Im_out, const uint16_t dim_im_out, q15_t
                                           *bufferA, q7_t *bufferB)
```

Basic Q7 convolution function.

**Buffer size:**

**Return** The function returns RISC\_V\_MATH\_SUCCESS

**Parameters**

- [in] Im\_in: pointer to input tensor
- [in] dim\_im\_in: input tensor dimension
- [in] ch\_im\_in: number of input tensor channels
- [in] wt: pointer to kernel weights
- [in] ch\_im\_out: number of filters, i.e., output tensor channels
- [in] dim\_kernel: filter kernel size
- [in] padding: padding sizes
- [in] stride: convolution stride
- [in] bias: pointer to bias
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [inout] Im\_out: pointer to output tensor
- [in] dim\_im\_out: output tensor dimension
- [inout] bufferA: pointer to buffer space for input
- [inout] bufferB: pointer to buffer space for output

bufferA size:  $2 \times \text{ch\_im\_in} \times \text{dim\_kernel} \times \text{dim\_kernel}$

bufferB size: 0

This basic version is designed to work for any input tensor and weight dimension.

```

riscv_status riscv_convolve_HWC_q7_basic_nonsquare (const q7_t *Im_in, const
                                                    uint16_t dim_im_in_x, const
                                                    uint16_t dim_im_in_y, const
                                                    uint16_t ch_im_in, const q7_t
                                                    *wt, const uint16_t ch_im_out,
                                                    const uint16_t dim_kernel_x,
                                                    const uint16_t dim_kernel_y,
                                                    const uint16_t padding_x, const
                                                    uint16_t padding_y, const
                                                    uint16_t stride_x, const uint16_t
                                                    stride_y, const q7_t *bias, const
                                                    uint16_t bias_shift, const uint16_t
                                                    out_shift, q7_t *Im_out, const
                                                    uint16_t dim_im_out_x, const
                                                    uint16_t dim_im_out_y, q15_t
                                                    *bufferA, q7_t *bufferB)

```

Basic Q7 convolution function (non-square shape)

Basic Q7 convolution function (non-square shape)

**Return** The function returns RISC\_V\_MATH\_SUCCESS

#### Parameters

- [in] Im\_in: pointer to input tensor
- [in] dim\_im\_in\_x: input tensor dimension x
- [in] dim\_im\_in\_y: input tensor dimension y
- [in] ch\_im\_in: number of input tensor channels
- [in] wt: pointer to kernel weights
- [in] ch\_im\_out: number of filters, i.e., output tensor channels
- [in] dim\_kernel\_x: filter kernel size x
- [in] dim\_kernel\_y: filter kernel size y
- [in] padding\_x: padding size x
- [in] padding\_y: padding size y
- [in] stride\_x: convolution stride x
- [in] stride\_y: convolution stride y
- [in] bias: pointer to bias
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [inout] Im\_out: pointer to output tensor
- [in] dim\_im\_out\_x: output tensor dimension x
- [in] dim\_im\_out\_y: output tensor dimension y
- [inout] bufferA: pointer to buffer space for input
- [inout] bufferB: pointer to buffer space for output

```
riscv_status riscv_convolve_HWC_q7_fast (const q7_t *Im_in, const uint16_t dim_im_in,  
                                         const uint16_t ch_im_in, const q7_t *wt,  
                                         const uint16_t ch_im_out, const uint16_t  
                                         dim_kernel, const uint16_t padding, const  
                                         uint16_t stride, const q7_t *bias, const uint16_t  
                                         bias_shift, const uint16_t out_shift, q7_t *Im_out,  
                                         const uint16_t dim_im_out, q15_t *bufferA, q7_t  
                                         *bufferB)
```

Fast Q7 convolution function.

**Buffer size:**

**Return** The function returns either RISCVMATH\_SIZE\_MISMATCH or RISCVMATH\_SUCCESS based on the outcome of size checking.

**Parameters**

- [in] Im\_in: pointer to input tensor
- [in] dim\_im\_in: input tensor dimension
- [in] ch\_im\_in: number of input tensor channels
- [in] wt: pointer to kernel weights
- [in] ch\_im\_out: number of filters, i.e., output tensor channels
- [in] dim\_kernel: filter kernel size
- [in] padding: padding sizes
- [in] stride: convolution stride
- [in] bias: pointer to bias
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [inout] Im\_out: pointer to output tensor
- [in] dim\_im\_out: output tensor dimension
- [inout] bufferA: pointer to buffer space for input
- [inout] bufferB: pointer to buffer space for output

bufferA size: 2\*ch\_im\_in\*dim\_kernel\*dim\_kernel

bufferB size: 0

**Input dimension constraints:**

ch\_im\_in is multiple of 4 ( because of the SIMD32 read and swap )

ch\_im\_out is multiple of 2 ( because 2x2 mat\_mult kernel )

The im2col converts the Q7 tensor input into Q15 column, which is stored in bufferA. There is reordering happening during this im2col process with riscv\_q7\_to\_q15\_reordered\_no\_shift. For every four elements, the second and third elements are swapped.

The computation kernel riscv\_nn\_mat\_mult\_kernel\_q7\_q15\_reordered does the GEMM computation with the reordered columns.

To speed-up the determination of the padding condition, we split the computation into 3x3 parts, i.e., {top, mid, bottom} X {left, mid, right}. This reduces the total number of boundary condition checks and improves the data copying performance.



```

riscv_status riscv_convolve_HWC_q7_fast_nonsquare (const q7_t *Im_in, const
                                                    uint16_t dim_im_in_x, const
                                                    uint16_t dim_im_in_y, const
                                                    uint16_t ch_im_in, const q7_t
                                                    *wt, const uint16_t ch_im_out,
                                                    const uint16_t dim_kernel_x,
                                                    const uint16_t dim_kernel_y,
                                                    const uint16_t padding_x, const
                                                    uint16_t padding_y, const uint16_t
                                                    stride_x, const uint16_t stride_y,
                                                    const q7_t *bias, const uint16_t
                                                    bias_shift, const uint16_t out_shift,
                                                    q7_t *Im_out, const uint16_t
                                                    dim_im_out_x, const uint16_t
                                                    dim_im_out_y, q15_t *bufferA, q7_t
                                                    *bufferB)

```

Fast Q7 convolution function (non-square shape)

This function is the version with full list of optimization tricks, but with some constraints: `ch_im_in` is multiple of 4 `ch_im_out` is multiple of 2

**Return** The function returns either `RISCV_MATH_SIZE_MISMATCH` or `RISCV_MATH_SUCCESS` based on the outcome of size checking.

#### Parameters

- [in] `Im_in`: pointer to input tensor
- [in] `dim_im_in_x`: input tensor dimension x
- [in] `dim_im_in_y`: input tensor dimension y
- [in] `ch_im_in`: number of input tensor channels
- [in] `wt`: pointer to kernel weights
- [in] `ch_im_out`: number of filters, i.e., output tensor channels
- [in] `dim_kernel_x`: filter kernel size x
- [in] `dim_kernel_y`: filter kernel size y
- [in] `padding_x`: padding size x
- [in] `padding_y`: padding size y
- [in] `stride_x`: convolution stride x
- [in] `stride_y`: convolution stride y
- [in] `bias`: pointer to bias
- [in] `bias_shift`: amount of left-shift for bias
- [in] `out_shift`: amount of right-shift for output
- [inout] `Im_out`: pointer to output tensor
- [in] `dim_im_out_x`: output tensor dimension x
- [in] `dim_im_out_y`: output tensor dimension y
- [inout] `bufferA`: pointer to buffer space for input
- [inout] `bufferB`: pointer to buffer space for output

```
riscv_status riscv_convolve_HWC_q7_RGB(const q7_t *Im_in, const uint16_t dim_im_in,  
                                         const uint16_t ch_im_in, const q7_t *wt, const  
                                         uint16_t ch_im_out, const uint16_t dim_kernel,  
                                         const uint16_t padding, const uint16_t stride,  
                                         const q7_t *bias, const uint16_t bias_shift,  
                                         const uint16_t out_shift, q7_t *Im_out, const  
                                         uint16_t dim_im_out, q15_t *bufferA, q7_t *bufferB)
```

Q7 convolution function for RGB image.

Q7 version of convolution for RGB image.

**Buffer size:**

**Return** The function returns either `RISCV_MATH_SIZE_MISMATCH` or `RISCV_MATH_SUCCESS` based on the outcome of size checking.

**Parameters**

- [in] `Im_in`: pointer to input tensor
- [in] `dim_im_in`: input tensor dimension
- [in] `ch_im_in`: number of input tensor channels
- [in] `wt`: pointer to kernel weights
- [in] `ch_im_out`: number of filters, i.e., output tensor channels
- [in] `dim_kernel`: filter kernel size
- [in] `padding`: padding sizes
- [in] `stride`: convolution stride
- [in] `bias`: pointer to bias
- [in] `bias_shift`: amount of left-shift for bias
- [in] `out_shift`: amount of right-shift for output
- [inout] `Im_out`: pointer to output tensor
- [in] `dim_im_out`: output tensor dimension
- [inout] `bufferA`: pointer to buffer space for input
- [inout] `bufferB`: pointer to buffer space for output

bufferA size:  $2 \times \text{ch\_im\_in} \times \text{dim\_kernel} \times \text{dim\_kernel}$

bufferB size: 0

**Input dimension constraints:**

`ch_im_in` equals 3

This kernel is written exclusively for convolution with `ch_im_in` equals 3. This applies on the first layer of CNNs which has input image with RGB format.

```
riscv_status riscv_convolve_s8(const nmsis_nn_context *ctx, const nmsis_nn_conv_params  
                               *conv_params, const nmsis_nn_per_channel_quant_params  
                               *quant_params, const nmsis_nn_dims *input_dims, const  
                               q7_t *input_data, const nmsis_nn_dims *filter_dims, const  
                               q7_t *filter_data, const nmsis_nn_dims *bias_dims, const  
                               int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t  
                               *output_data)
```

Basic s8 convolution function.

1. Supported framework: TensorFlow Lite micro
2. q7 is used as data type eventhough it is s8 data. It is done so to be consistent with existing APIs.
3. Additional memory is required for optimization. Refer to argument 'ctx' for details.

**Return** The function returns `RISCV_MATH_SUCCESS`

#### Parameters

- [inout] `ctx`: Function context that contains the additional buffer if required by the function. `riscv_convolve_s8_get_buffer_size` will return the `buffer_size` if required
- [in] `conv_params`: Convolution parameters (e.g. strides, dilations, pads,...). Range of `conv_params->input_offset` : [-127, 128] Range of `conv_params->output_offset` : [-128, 127]
- [in] `quant_params`: Per-channel quantization info. It contains the multiplier and shift values to be applied to each output channel
- [in] `input_dims`: Input (activation) tensor dimensions. Format: [N, H, W, C\_IN]
- [in] `input_data`: Input (activation) data pointer. Data type: int8
- [in] `filter_dims`: Filter tensor dimensions. Format: [C\_OUT, HK, WK, C\_IN] where HK and WK are the spatial filter dimensions
- [in] `filter_data`: Filter data pointer. Data type: int8
- [in] `bias_dims`: Bias tensor dimensions. Format: [C\_OUT]
- [in] `bias_data`: Optional bias data pointer. Data type: int32
- [in] `output_dims`: Output tensor dimensions. Format: [N, H, W, C\_OUT]
- [out] `output_data`: Output data pointer. Data type: int8

```
int32_t riscv_convolve_s8_get_buffer_size (const nmsis_nn_dims *input_dims, const
                                         nmsis_nn_dims *filter_dims)
```

Get the required buffer size for s8 convolution function.

**Return** The function returns required buffer size(bytes)

#### Parameters

- [in] `input_dims`: Input (activation) tensor dimensions. Format: [N, H, W, C\_IN]
- [in] `filter_dims`: Filter tensor dimensions. Format: [C\_OUT, HK, WK, C\_IN] where HK and WK are the spatial filter dimensions

```
riscv_status riscv_convolve_wrapper_s8 (const nmsis_nn_context *ctx, const nmsis_nn_conv_params *conv_params, const nmsis_nn_per_channel_quant_params *quant_params,
const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims,
const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const
nmsis_nn_dims *output_dims, q7_t *output_data)
```

s8 convolution layer wrapper function with the main purpose to call the optimal kernel available in nmsis-nn to perform the convolution.

**Return** The function returns either `RISCV_MATH_SIZE_MISMATCH` if argument constraints fail. or, `RISCV_MATH_SUCCESS` on successful completion.

**Parameters**

- [inout] `ctx`: Function context that contains the additional buffer if required by the function. `riscv_convolve_wrapper_s8_get_buffer_size` will return the `buffer_size` if required
- [in] `conv_params`: Convolution parameters (e.g. strides, dilations, pads,...). Range of `conv_params->input_offset` : [-127, 128] Range of `conv_params->output_offset` : [-128, 127]
- [in] `quant_params`: Per-channel quantization info. It contains the multiplier and shift values to be applied to each output channel
- [in] `input_dims`: Input (activation) tensor dimensions. Format: [N, H, W, C\_IN]
- [in] `input_data`: Input (activation) data pointer. Data type: int8
- [in] `filter_dims`: Filter tensor dimensions. Format: [C\_OUT, HK, WK, C\_IN] where HK and WK are the spatial filter dimensions
- [in] `filter_data`: Filter data pointer. Data type: int8
- [in] `bias_dims`: Bias tensor dimensions. Format: [C\_OUT]
- [in] `bias_data`: Bias data pointer. Data type: int32
- [in] `output_dims`: Output tensor dimensions. Format: [N, H, W, C\_OUT]
- [out] `output_data`: Output data pointer. Data type: int8

```
int32_t riscv_convolve_wrapper_s8_get_buffer_size (const nmsis_nn_conv_params
                                                    *conv_params, const nmsis_nn_dims *input_dims, const nmsis_nn_dims *filter_dims,
                                                    const nmsis_nn_dims *output_dims)
```

Get the required buffer size for `riscv_convolve_wrapper_s8`.

**Return** The function returns required buffer size(bytes)

**Parameters**

- [in] `conv_params`: Convolution parameters (e.g. strides, dilations, pads,...). Range of `conv_params->input_offset` : [-127, 128] Range of `conv_params->output_offset` : [-128, 127]
- [in] `input_dims`: Input (activation) dimensions. Format: [N, H, W, C\_IN]
- [in] `filter_dims`: Filter dimensions. Format: [C\_OUT, HK, WK, C\_IN] where HK and WK are the spatial filter dimensions
- [in] `output_dims`: Output tensor dimensions. Format: [N, H, W, C\_OUT]

```
riscv_status riscv_depthwise_conv_3x3_s8 (const nmsis_nn_context *ctx, const nmsis_nn_dw_conv_params *dw_conv_params,
const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data,
const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

Optimized s8 depthwise convolution function for 3x3 kernel size with some constraints on the input arguments(documented below). Refer `riscv_depthwise_conv_s8()` for function argument details.

- Supported framework : TensorFlow Lite Micro
- The following constrains on the arguments apply
  1. Number of input channel equals number of output channels
  2. Filter height and width equals 3
  3. Padding along x is either 0 or 1.

**Return** The function returns one of the following `RISCV_MATH_SIZE_MISMATCH` - Unsupported dimension of tensors `RISCV_MATH_ARGUMENT_ERROR` - Unsupported pad size along the x axis `RISCV_MATH_SUCCESS` - Successful operation

```
static void depthwise_conv_s8_mult_4(const int8_t *input, const int32_t input_x,
                                     const int32_t input_y, const int32_t input_ch,
                                     const int8_t *kernel, const int32_t output_ch,
                                     const int32_t ch_mult, const int32_t kernel_x,
                                     const int32_t kernel_y, const int32_t pad_x,
                                     const int32_t pad_y, const int32_t stride_x,
                                     const int32_t stride_y, const int32_t *bias,
                                     int8_t *output, const int32_t *output_shift, const
                                     int32_t *output_mult, const int32_t output_x,
                                     const int32_t output_y, const int32_t out-
                                     put_offset, const int32_t input_offset, const
                                     int32_t output_activation_min, const int32_t out-
                                     put_activation_max)

static void depthwise_conv_s8_generic(const q7_t *input, const uint16_t input_batches,
                                     const uint16_t input_x, const uint16_t input_y,
                                     const uint16_t input_ch, const q7_t *kernel,
                                     const uint16_t output_ch, const uint16_t
                                     ch_mult, const uint16_t kernel_x, const
                                     uint16_t kernel_y, const uint16_t pad_x, const
                                     uint16_t pad_y, const uint16_t stride_x, const
                                     uint16_t stride_y, const int32_t *bias, q7_t
                                     *output, const int32_t *output_shift, const
                                     int32_t *output_mult, const uint16_t output_x,
                                     const uint16_t output_y, const int32_t out-
                                     put_offset, const int32_t input_offset, const
                                     int32_t output_activation_min, const int32_t
                                     output_activation_max)

riscv_status riscv_depthwise_conv_s8(const nmsis_nn_context *ctx, const nm-
s_
sis_nn_dw_conv_params *dw_conv_params, const
nmsis_nn_per_channel_quant_params *quant_params,
const nmsis_nn_dims *input_dims, const q7_t *in-
put_data, const nmsis_nn_dims *filter_dims, const
q7_t *filter_data, const nmsis_nn_dims *bias_dims,
const int32_t *bias_data, const nmsis_nn_dims
*output_dims, q7_t *output_data)
```

Basic s8 depthwise convolution function that doesn't have any constraints on the input dimensions.

- Supported framework: TensorFlow Lite
- q7 is used as data type eventhough it is s8 data. It is done so to be consistent with existing APIs.

**Return** The function returns `RISCV_MATH_SUCCESS`

### Parameters

- [inout] `ctx`: Function context (e.g. temporary buffer). Check the function definition file to see if an additional buffer is required. Optional function `{API}_get_buffer_size()` provides the buffer size if an additional buffer is required. exists if additional memory is.
- [in] `dw_conv_params`: Depthwise convolution parameters (e.g. strides, dilations, pads,...) `dw_conv_params->dilation` is not used. Range of `dw_conv_params->input_offset` : [-127, 128] Range of `dw_conv_params->input_offset` : [-128, 127]
- [in] `quant_params`: Per-channel quantization info. It contains the multiplier and shift values to be applied to each output channel
- [in] `input_dims`: Input (activation) tensor dimensions. Format: [1, H, W, C\_IN] Batch argument N is not used.
- [in] `input_data`: Input (activation) data pointer. Data type: int8
- [in] `filter_dims`: Filter tensor dimensions. Format: [1, H, W, C\_OUT]
- [in] `filter_data`: Filter data pointer. Data type: int8
- [in] `bias_dims`: Bias tensor dimensions. Format: [C\_OUT]
- [in] `bias_data`: Bias data pointer. Data type: int32
- [in] `output_dims`: Output tensor dimensions. Format: [1, H, W, C\_OUT]
- [inout] `output_data`: Output data pointer. Data type: int8

```
riscv_status riscv_depthwise_conv_s8_opt (const nmsis_nn_context *ctx, const nmsis_nn_dw_conv_params *dw_conv_params,
const nmsis_nn_per_channel_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data,
const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

Optimized s8 depthwise convolution function with constraint that in\_channel equals out\_channel. Refer `riscv_depthwise_conv_s8()` for function argument details.

- Supported framework: TensorFlow Lite
- The following constraints on the arguments apply
  1. Number of input channel equals number of output channels or `ch_mult` equals 1
- q7 is used as data type even though it is s8 data. It is done so to be consistent with existing APIs.
- Recommended when number of channels is 4 or greater.

**Return** The function returns one of the following `RISCV_MATH_SIZE_MISMATCH` - input channel != output channel or `ch_mult != 1` `RISCV_MATH_SUCCESS` - Successful operation

**Note** If number of channels is not a multiple of 4, upto 3 elements outside the boundary will be read out for the following if MVE optimizations(Arm Helium Technology) are used.

- Output shift
- Output multiplier
- Output bias

- kernel

```
int32_t riscv_depthwise_conv_s8_opt_get_buffer_size(const nmsis_nn_dims
                                                    *input_dims, const nmsis_nn_dims *filter_dims)
```

Get the required buffer size for optimized s8 depthwise convolution function with constraint that in\_channel equals out\_channel.

**Return** The function returns required buffer size in bytes

#### Parameters

- [in] input\_dims: Input (activation) tensor dimensions. Format: [1, H, W, C\_IN] Batch argument N is not used.
- [in] filter\_dims: Filter tensor dimensions. Format: [1, H, W, C\_OUT]

```
static void depthwise_conv_u8_mult_4(const uint8_t *input, const int32_t input_x,
                                     const int32_t input_y, const int32_t input_ch,
                                     const uint8_t *kernel, const int32_t output_ch,
                                     const int32_t ch_mult, const int32_t kernel_x,
                                     const int32_t kernel_y, const int32_t pad_x,
                                     const int32_t pad_y, const int32_t stride_x,
                                     const int32_t stride_y, const int32_t *bias,
                                     uint8_t *output, const int32_t output_shift,
                                     const int32_t output_mult, const int32_t output_x,
                                     const int32_t output_y, const int32_t output_offset,
                                     const int32_t input_offset, const int32_t filter_offset,
                                     const int32_t output_activation_min, const int32_t
                                     output_activation_max)
```

```
static void depthwise_conv_u8_generic(const uint8_t *input, const int32_t input_x,
                                      const int32_t input_y, const int32_t input_ch,
                                      const uint8_t *kernel, const int32_t output_ch,
                                      const int32_t ch_mult, const int32_t kernel_x,
                                      const int32_t kernel_y, const int32_t pad_x,
                                      const int32_t pad_y, const int32_t stride_x,
                                      const int32_t stride_y, const int32_t *bias,
                                      uint8_t *output, const int32_t output_shift,
                                      const int32_t output_mult, const int32_t output_x,
                                      const int32_t output_y, const int32_t output_offset,
                                      const int32_t input_offset, const int32_t filter_offset,
                                      const int32_t output_activation_min, const int32_t
                                      output_activation_max)
```

```
riscv_status riscv_depthwise_conv_u8_basic_ver1 (const uint8_t *input, const uint16_t
input_x, const uint16_t input_y,
const uint16_t input_ch, const
uint8_t *kernel, const uint16_t
kernel_x, const uint16_t kernel_y,
const int16_t ch_mult, const
int16_t pad_x, const int16_t pad_y,
const int16_t stride_x, const int16_t
stride_y, const int16_t dilation_x,
const int16_t dilation_y, const
int32_t *bias, const int32_t in-
put_offset, const int32_t filter_offset,
const int32_t output_offset, uint8_t
*output, const uint16_t output_x,
const uint16_t output_y, const
int32_t output_activation_min, const
int32_t output_activation_max, const
int32_t output_shift, const int32_t
output_mult)
```

uint8 depthwise convolution function with asymmetric quantization

uint8 depthwise convolution function with asymmetric quantization Unless specified otherwise, arguments are mandatory.

**Return** The function returns one of the following RISC\_V\_MATH\_SIZE\_MISMATCH - Not supported dimension of tensors RISC\_V\_MATH\_SUCCESS - Successful operation RISC\_V\_MATH\_ARGUMENT\_ERROR - Implementation not available

#### Parameters

- [in] input: Pointer to input tensor
- [in] input\_x: Width of input tensor
- [in] input\_y: Height of input tensor
- [in] input\_ch: Channels in input tensor
- [in] kernel: Pointer to kernel weights
- [in] kernel\_x: Width of kernel
- [in] kernel\_y: Height of kernel
- [in] ch\_mult: Number of channel multiplier
- [in] pad\_x: Padding sizes x
- [in] pad\_y: Padding sizes y
- [in] stride\_x: Convolution stride along the width
- [in] stride\_y: Convolution stride along the height
- [in] dilation\_x: Dilation along width. Not used and intended for future enhancement.
- [in] dilation\_y: Dilation along height. Not used and intended for future enhancement.
- [in] bias: Pointer to optional bias values. If no bias is available, NULL is expected
- [in] input\_offset: Input tensor zero offset
- [in] filter\_offset: Kernel tensor zero offset



- [in] output\_offset: Output tensor zero offset
- [inout] output: Pointer to output tensor
- [in] output\_x: Width of output tensor
- [in] output\_y: Height of output tensor
- [in] output\_activation\_min: Minimum value to clamp the output to. Range : {0, 255}
- [in] output\_activation\_max: Minimum value to clamp the output to. Range : {0, 255}
- [in] output\_shift: Amount of right-shift for output
- [in] output\_mult: Output multiplier for requantization

```
riscv_status riscv_depthwise_conv_wrapper_s8 (const nmsis_nn_context *ctx,
                                              const nmsis_nn_dw_conv_params
                                              *dw_conv_params, const nmsis_nn_per_channel_quant_params
                                              *quant_params, const nmsis_nn_dims
                                              *input_dims, const q7_t *input_data,
                                              const nmsis_nn_dims *filter_dims, const
                                              q7_t *filter_data, const nmsis_nn_dims
                                              *bias_dims, const int32_t *bias_data,
                                              const nmsis_nn_dims *output_dims, q7_t
                                              *output_data)
```

Wrapper function to pick the right optimized s8 depthwise convolution function.

- Supported framework: TensorFlow Lite
- Picks one of the the following functions
  1. riscv\_depthwise\_conv\_s8()
  2. riscv\_depthwise\_conv\_3x3\_s8() - RISC-V CPUs with DSP extension only
  3. riscv\_depthwise\_conv\_s8\_opt()
- q7 is used as data type eventhough it is s8 data. It is done so to be consistent with existing APIs.
- Check details of riscv\_depthwise\_conv\_s8\_opt() for potential data that can be accessed outside of the boundary.

**Return** The function returns RISC\_V\_MATH\_SUCCESS - Successful completion.

#### Parameters

- [inout] ctx: Function context (e.g. temporary buffer). Check the function definition file to see if an additional buffer is required. Optional function {API}\_get\_buffer\_size() provides the buffer size if required.
- [in] dw\_conv\_params: Depthwise convolution parameters (e.g. strides, dilations, pads,...) dw\_conv\_params->dilation is not used. Range of dw\_conv\_params->input\_offset : [-127, 128] Range of dw\_conv\_params->output\_offset : [-128, 127]
- [in] quant\_params: Per-channel quantization info. It contains the multiplier and shift values to be applied to each output channel
- [in] input\_dims: Input (activation) tensor dimensions. Format: [H, W, C\_IN] Batch argument N is not used and assumed to be 1.
- [in] input\_data: Input (activation) data pointer. Data type: int8

- [in] `filter_dims`: Filter tensor dimensions. Format: [1, H, W, C\_OUT]
- [in] `filter_data`: Filter data pointer. Data type: int8
- [in] `bias_dims`: Bias tensor dimensions. Format: [C\_OUT]
- [in] `bias_data`: Bias data pointer. Data type: int32
- [in] `output_dims`: Output tensor dimensions. Format: [1, H, W, C\_OUT]
- [inout] `output_data`: Output data pointer. Data type: int8

```
int32_t riscv_depthwise_conv_wrapper_s8_get_buffer_size (const nm-
                                                         sis_nn_dw_conv_params
                                                         *dw_conv_params,
                                                         const nmsis_nn_dims
                                                         *input_dims, const nm-
                                                         sis_nn_dims *filter_dims,
                                                         const nmsis_nn_dims
                                                         *output_dims)
```

Get size of additional buffer required by `riscv_depthwise_conv_wrapper_s8()`

**Return** Size of additional memory required for optimizations in bytes.

**Parameters**

- [in] `dw_conv_params`: Depthwise convolution parameters (e.g. strides, dilations, pads,...) `dw_conv_params->dilation` is not used. Range of `dw_conv_params->input_offset` : [-127, 128] Range of `dw_conv_params->input_offset` : [-128, 127]
- [in] `input_dims`: Input (activation) tensor dimensions. Format: [H, W, C\_IN] Batch argument N is not used and assumed to be 1.
- [in] `filter_dims`: Filter tensor dimensions. Format: [1, H, W, C\_OUT]
- [in] `output_dims`: Output tensor dimensions. Format: [1, H, W, C\_OUT]

```
riscv_status riscv_depthwise_separable_conv_HWC_q7 (const q7_t *Im_in, const
uint16_t dim_im_in, const
uint16_t ch_im_in, const q7_t
*wt, const uint16_t ch_im_out,
const uint16_t dim_kernel,
const uint16_t padding, const
uint16_t stride, const q7_t *bias,
const uint16_t bias_shift, const
uint16_t out_shift, q7_t *Im_out,
const uint16_t dim_im_out, q15_t
*bufferA, q7_t *bufferB)
```

Q7 depthwise separable convolution function.

**Buffer size:**

**Return** The function returns either `RISCV_MATH_SIZE_MISMATCH` or `RISCV_MATH_SUCCESS` based on the outcome of size checking.

**Parameters**

- [in] `Im_in`: pointer to input tensor
- [in] `dim_im_in`: input tensor dimension
- [in] `ch_im_in`: number of input tensor channels

- [in] wt: pointer to kernel weights
- [in] ch\_im\_out: number of filters, i.e., output tensor channels
- [in] dim\_kernel: filter kernel size
- [in] padding: padding sizes
- [in] stride: convolution stride
- [in] bias: pointer to bias
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [inout] Im\_out: pointer to output tensor
- [in] dim\_im\_out: output tensor dimension
- [inout] bufferA: pointer to buffer space for input
- [inout] bufferB: pointer to buffer space for output

bufferA size:  $2 \times \text{ch\_im\_in} \times \text{dim\_kernel} \times \text{dim\_kernel}$

bufferB size: 0

**Input dimension constraints:**

ch\_im\_in equals ch\_im\_out

Implementation: There are 3 nested loop here: Inner loop: calculate each output value with MAC instruction over an accumulator Mid loop: loop over different output channel Outer loop: loop over different output (x, y)

```
riscv_status riscv_depthwise_separable_conv_HWC_q7_nonsquare (const q7_t *Im_in,
                                                             const uint16_t
                                                             dim_im_in_x,
                                                             const uint16_t
                                                             dim_im_in_y,
                                                             const uint16_t
                                                             ch_im_in, const
                                                             q7_t *wt, const
                                                             uint16_t ch_im_out,
                                                             const uint16_t
                                                             dim_kernel_x,
                                                             const uint16_t
                                                             dim_kernel_y,
                                                             const uint16_t
                                                             padding_x, const
                                                             uint16_t padding_y,
                                                             const uint16_t
                                                             stride_x, const
                                                             uint16_t stride_y,
                                                             const q7_t *bias,
                                                             const uint16_t
                                                             bias_shift, const
                                                             uint16_t out_shift,
                                                             q7_t *Im_out,
                                                             const uint16_t
                                                             dim_im_out_x,
                                                             const uint16_t
                                                             dim_im_out_y,
                                                             q15_t *bufferA, q7_t
                                                             *bufferB)
```

Q7 depthwise separable convolution function (non-square shape)

This function is the version with full list of optimization tricks, but with some constraints: `ch_im_in` is equal to `ch_im_out`

**Return** The function returns either `RISCV_MATH_SIZE_MISMATCH` or `RISCV_MATH_SUCCESS` based on the outcome of size checking.

#### Parameters

- [in] `Im_in`: pointer to input tensor
- [in] `dim_im_in_x`: input tensor dimension x
- [in] `dim_im_in_y`: input tensor dimension y
- [in] `ch_im_in`: number of input tensor channels
- [in] `wt`: pointer to kernel weights
- [in] `ch_im_out`: number of filters, i.e., output tensor channels
- [in] `dim_kernel_x`: filter kernel size x
- [in] `dim_kernel_y`: filter kernel size y
- [in] `padding_x`: padding sizes x
- [in] `padding_y`: padding sizes y
- [in] `stride_x`: convolution stride x

- [in] `stride_y`: convolution stride y
- [in] `bias`: pointer to bias
- [in] `bias_shift`: amount of left-shift for bias
- [in] `out_shift`: amount of right-shift for output
- [inout] `Im_out`: pointer to output tensor
- [in] `dim_im_out_x`: output tensor dimension x
- [in] `dim_im_out_y`: output tensor dimension y
- [inout] `bufferA`: pointer to buffer space for input
- [inout] `bufferB`: pointer to buffer space for output

### Fully-connected Layer Functions

```
riscv_status riscv_fully_connected_mat_q7_vec_q15(const q15_t *pV, const q7_t *pM,
                                                    const uint16_t dim_vec, const uint16_t
                                                    num_of_rows, const uint16_t bias_shift,
                                                    const uint16_t out_shift, const q7_t
                                                    *bias, q15_t *pOut, q15_t *vec_buffer)

riscv_status riscv_fully_connected_mat_q7_vec_q15_opt(const q15_t *pV, const q7_t
                                                         *pM, const uint16_t dim_vec,
                                                         const uint16_t num_of_rows,
                                                         const uint16_t bias_shift,
                                                         const uint16_t out_shift, const
                                                         q7_t *bias, q15_t *pOut, q15_t
                                                         *vec_buffer)

riscv_status riscv_fully_connected_q15(const q15_t *pV, const q15_t *pM, const uint16_t
                                         dim_vec, const uint16_t num_of_rows, const uint16_t
                                         bias_shift, const uint16_t out_shift, const q15_t *bias,
                                         q15_t *pOut, q15_t *vec_buffer)

riscv_status riscv_fully_connected_q15_opt(const q15_t *pV, const q15_t *pM, const
                                             uint16_t dim_vec, const uint16_t num_of_rows,
                                             const uint16_t bias_shift, const uint16_t
                                             out_shift, const q15_t *bias, q15_t *pOut, q15_t
                                             *vec_buffer)

riscv_status riscv_fully_connected_q7(const q7_t *pV, const q7_t *pM, const uint16_t
                                        dim_vec, const uint16_t num_of_rows, const uint16_t
                                        bias_shift, const uint16_t out_shift, const q7_t *bias,
                                        q7_t *pOut, q15_t *vec_buffer)

riscv_status riscv_fully_connected_q7_opt(const q7_t *pV, const q7_t *pM, const uint16_t
                                            dim_vec, const uint16_t num_of_rows, const
                                            uint16_t bias_shift, const uint16_t out_shift, const
                                            q7_t *bias, q7_t *pOut, q15_t *vec_buffer)

riscv_status riscv_fully_connected_s8(const nmsis_nn_context *ctx, const nmsis_nn_fc_params
                                       *fc_params, const nmsis_nn_per_tensor_quant_params
                                       *quant_params, const nmsis_nn_dims *input_dims,
                                       const q7_t *input_data, const nmsis_nn_dims *fil-
                                       ter_dims, const q7_t *filter_data, const nmsis_nn_dims
                                       *bias_dims, const int32_t *bias_data, const nm-
                                       sis_nn_dims *output_dims, q7_t *output_data)
```

```
int32_t riscv_fully_connected_s8_get_buffer_size (const nmsis_nn_dims *filter_dims)
```

**USE\_INTRINSIC***group* **FC**

Collection of fully-connected and matrix multiplication functions.

Fully-connected layer is basically a matrix-vector multiplication with bias. The matrix is the weights and the input/output vectors are the activation values. Supported {weight, activation} precisions include {8-bit, 8-bit}, {16-bit, 16-bit}, and {8-bit, 16-bit}.

Here we have two types of kernel functions. The basic function implements the function using regular GEMV approach. The opt functions operates with weights in interleaved formats.

**Defines****USE\_INTRINSIC**

Mixed Q15-Q7 opt fully-connected layer function.

**Buffer size:**

**Return** The function returns RISCV\_MATH\_SUCCESS

**Parameters**

- [in] pV: pointer to input vector
- [in] pM: pointer to matrix weights
- [in] dim\_vec: length of the vector
- [in] num\_of\_rows: number of rows in weight matrix
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [in] bias: pointer to bias
- [inout] pOut: pointer to output vector
- [inout] vec\_buffer: pointer to buffer space for input

vec\_buffer size: 0

Q7\_Q15 version of the fully connected layer

Weights are in q7\_t and Activations are in q15\_t

Limitation: x4 version requires weight reordering to work

Here we use only one pointer to read 4 rows in the weight matrix. So if the original q7\_t matrix looks like this:

| a11 | a12 | a13 | a14 | a15 | a16 | a17 |

| a21 | a22 | a23 | a24 | a25 | a26 | a27 |

| a31 | a32 | a33 | a34 | a35 | a36 | a37 |

| a41 | a42 | a43 | a44 | a45 | a46 | a47 |

| a51 | a52 | a53 | a54 | a55 | a56 | a57 |

| a61 | a62 | a63 | a64 | a65 | a66 | a67 |

We operates on multiple-of-4 rows, so the first four rows becomes

a11	a21	a12	a22	a31	a41	a32	a42
a13	a23	a14	a24	a33	a43	a34	a44
a15	a25	a16	a26	a35	a45	a36	a46

The column left over will be in-order. which is: | a17 | a27 | a37 | a47 |

For the left-over rows, we do 1x1 computation, so the data remains as its original order.

So the stored weight matrix looks like this:

a11	a21	a12	a22	a31	a41
a32	a42	a13	a23	a14	a24
a33	a43	a34	a44	a15	a25
a16	a26	a35	a45	a36	a46
a17	a27	a37	a47	a51	a52
a53	a54	a55	a56	a57	a61
a62	a63	a64	a65	a66	a67

## Functions

```
riscv_status riscv_fully_connected_mat_q7_vec_q15 (const q15_t *pV, const q7_t *pM,
                                                    const uint16_t dim_vec, const
                                                    uint16_t num_of_rows, const
                                                    uint16_t bias_shift, const uint16_t
                                                    out_shift, const q7_t *bias, q15_t
                                                    *pOut, q15_t *vec_buffer)
```

Mixed Q15-Q7 fully-connected layer function.

**Buffer size:**

**Return** The function returns RISCVMATH\_SUCCESS

**Parameters**

- [in] pV: pointer to input vector
- [in] pM: pointer to matrix weights
- [in] dim\_vec: length of the vector
- [in] num\_of\_rows: number of rows in weight matrix
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [in] bias: pointer to bias
- [inout] pOut: pointer to output vector
- [inout] vec\_buffer: pointer to buffer space for input

vec\_buffer size: 0

Q7\_Q15 version of the fully connected layer

Weights are in q7\_t and Activations are in q15\_t

```
riscv_status riscv_fully_connected_mat_q7_vec_q15_opt (const q15_t *pV, const q7_t  
*pM, const uint16_t dim_vec,  
*pM, const uint16_t num_of_rows,  
const uint16_t bias_shift,  
const uint16_t out_shift,  
const q7_t *bias, q15_t  
*pOut, q15_t *vec_buffer)
```

Mixed Q15-Q7 opt fully-connected layer function.

**Return** The function returns RISCVMATH\_SUCCESS

**Parameters**

- [in] pV: pointer to input vector
- [in] pM: pointer to matrix weights
- [in] dim\_vec: length of the vector
- [in] num\_of\_rows: number of rows in weight matrix
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [in] bias: pointer to bias
- [inout] pOut: pointer to output vector
- [inout] vec\_buffer: pointer to buffer space for input

```
riscv_status riscv_fully_connected_q15 (const q15_t *pV, const q15_t *pM, const  
uint16_t dim_vec, const uint16_t num_of_rows,  
const uint16_t bias_shift, const uint16_t out_shift,  
const q15_t *bias, q15_t *pOut, q15_t *vec_buffer)
```

Q15 opt fully-connected layer function.

Q15 basic fully-connected layer function.

**Buffer size:**

**Return** The function returns RISCVMATH\_SUCCESS

**Parameters**

- [in] pV: pointer to input vector
- [in] pM: pointer to matrix weights
- [in] dim\_vec: length of the vector
- [in] num\_of\_rows: number of rows in weight matrix
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [in] bias: pointer to bias
- [inout] pOut: pointer to output vector
- [inout] vec\_buffer: pointer to buffer space for input

vec\_buffer size: 0



```
riscv_status riscv_fully_connected_q15_opt (const q15_t *pV, const q15_t *pM,
                                             const uint16_t dim_vec, const uint16_t
                                             num_of_rows, const uint16_t bias_shift,
                                             const uint16_t out_shift, const q15_t *bias,
                                             q15_t *pOut, q15_t *vec_buffer)
```

Q15 opt fully-connected layer function.

**Buffer size:**

**Return** The function returns RISC\_V\_MATH\_SUCCESS

**Parameters**

- [in] pV: pointer to input vector
- [in] pM: pointer to matrix weights
- [in] dim\_vec: length of the vector
- [in] num\_of\_rows: number of rows in weight matrix
- [in] bias\_shift: amount of left-shift for bias
- [in] out\_shift: amount of right-shift for output
- [in] bias: pointer to bias
- [inout] pOut: pointer to output vector
- [inout] vec\_buffer: pointer to buffer space for input

vec\_buffer size: 0

Here we use only one pointer to read 4 rows in the weight matrix. So if the original matrix looks like this:

```
| a11 | a12 | a13 |
| a21 | a22 | a23 |
| a31 | a32 | a33 |
| a41 | a42 | a43 |
| a51 | a52 | a53 |
| a61 | a62 | a63 |
```

We operate on multiple-of-4 rows, so the first four rows become

```
| a11 | a12 | a21 | a22 | a31 | a32 | a41 | a42 |
| a13 | a23 | a33 | a43 |
```

Remaining rows are kept the same original order.

So the stored weight matrix looks like this:

```
| a11 | a12 | a21 | a22 | a31 | a32 | a41 | a42 |
| a13 | a23 | a33 | a43 | a51 | a52 | a53 | a61 |
| a62 | a63 |
```

```
riscv_status riscv_fully_connected_q7 (const q7_t *pV, const q7_t *pM, const uint16_t
                                          dim_vec, const uint16_t num_of_rows, const
                                          uint16_t bias_shift, const uint16_t out_shift, const
                                          q7_t *bias, q7_t *pOut, q15_t *vec_buffer)
```

Q7 basic fully-connected layer function.

**Buffer size:****Return** The function returns `RISCV_MATH_SUCCESS`**Parameters**

- [in] `pV`: pointer to input vector
- [in] `pM`: pointer to matrix weights
- [in] `dim_vec`: length of the vector
- [in] `num_of_rows`: number of rows in weight matrix
- [in] `bias_shift`: amount of left-shift for bias
- [in] `out_shift`: amount of right-shift for output
- [in] `bias`: pointer to bias
- [inout] `pOut`: pointer to output vector
- [inout] `vec_buffer`: pointer to buffer space for input

`vec_buffer` size: `dim_vec`

This basic function is designed to work with regular weight matrix without interleaving.

```
riscv_status riscv_fully_connected_q7_opt (const q7_t *pV, const q7_t *pM, const
                                         uint16_t dim_vec, const uint16_t num_of_rows,
                                         const uint16_t bias_shift, const uint16_t
                                         out_shift, const q7_t *bias, q7_t *pOut, q15_t
                                         *vec_buffer)
```

Q7 opt fully-connected layer function.

**Buffer size:****Return** The function returns `RISCV_MATH_SUCCESS`**Parameters**

- [in] `pV`: pointer to input vector
- [in] `pM`: pointer to matrix weights
- [in] `dim_vec`: length of the vector
- [in] `num_of_rows`: number of rows in weight matrix
- [in] `bias_shift`: amount of left-shift for bias
- [in] `out_shift`: amount of right-shift for output
- [in] `bias`: pointer to bias
- [inout] `pOut`: pointer to output vector
- [inout] `vec_buffer`: pointer to buffer space for input

`vec_buffer` size: `dim_vec`

This opt function is designed to work with interleaved weight matrix. The vector input is assumed in `q7_t` format, we call `riscv_q7_to_q15_no_shift_shuffle` function to expand into `q15_t` format with certain weight re-ordering, refer to the function comments for more details. Here we use only one pointer to read 4 rows in the weight matrix. So if the original `q7_t` matrix looks like this:

```
| a11 | a12 | a13 | a14 | a15 | a16 | a17 |
| a21 | a22 | a23 | a24 | a25 | a26 | a27 |
```

a31	a32	a33	a34	a35	a36	a37
a41	a42	a43	a44	a45	a46	a47
a51	a52	a53	a54	a55	a56	a57
a61	a62	a63	a64	a65	a66	a67

We operates on multiple-of-4 rows, so the first four rows becomes

a11	a21	a13	a23	a31	a41	a33	a43
a12	a22	a14	a24	a32	a42	a34	a44
a15	a25	a35	a45	a16	a26	a36	a46

So within the kernel, we first read the re-ordered vector in as:

| b1 | b3 | and | b2 | b4 |

the four q31\_t weights will look like

| a11 | a13 |, | a21 | a23 |, | a31 | a33 |, | a41 | a43 |  
 | a12 | a14 |, | a22 | a24 |, | a32 | a34 |, | a42 | a44 |

The column left over will be in-order. which is:

| a17 | a27 | a37 | a47 |

For the left-over rows, we do 1x1 computation, so the data remains as its original order.

So the stored weight matrix looks like this:

a11	a21	a13	a23	a31	a41
a33	a43	a12	a22	a14	a24
a32	a42	a34	a44	a15	a25
a35	a45	a16	a26	a36	a46
a17	a27	a37	a47	a51	a52
a53	a54	a55	a56	a57	a61
a62	a63	a64	a65	a66	a67

```
riscv_status riscv_fully_connected_s8(const nmsis_nn_context *ctx, const nmsis_nn_fc_params *fc_params, const nmsis_nn_per_tensor_quant_params *quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *filter_dims, const q7_t *filter_data, const nmsis_nn_dims *bias_dims, const int32_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

Basic s8 Fully Connected function.

- Supported framework: TensorFlow Lite
- q7 is used as data type eventhough it is s8 data. It is done so to be consistent with existing APIs.

**Return** The function returns RISC\_V\_MATH\_SUCCESS

**Parameters**

- [inout] `ctx`: Function context (e.g. temporary buffer). Check the function definition file to see if an additional buffer is required. Optional function `{API}_get_buffer_size()` provides the buffer size if an additional buffer is required.
- [in] `fc_params`: Fully Connected layer parameters (e.g. strides, dilations, pads,...) Range of `fc_params->input_offset` : [-127, 128] `fc_params->filter_offset` : 0 Range of `fc_params->output_offset` : [-128, 127]
- [in] `quant_params`: Per-tensor quantization info. It contains the multiplier and shift values to be applied to the output tensor.
- [in] `input_dims`: Input (activation) tensor dimensions. Format: [N, H, W, C\_IN] Input dimension is taken as  $N \times (H * W * C_{IN})$
- [in] `input_data`: Input (activation) data pointer. Data type: int8
- [in] `filter_dims`: Two dimensional filter dimensions. Format: [N, C] N : accumulation depth and equals  $(H * W * C_{IN})$  from `input_dims` C : output depth and equals `C_OUT` in `output_dims` H & W : Not used
- [in] `filter_data`: Filter data pointer. Data type: int8
- [in] `bias_dims`: Bias tensor dimensions. Format: [C\_OUT] N, H, W : Not used
- [in] `bias_data`: Bias data pointer. Data type: int32
- [in] `output_dims`: Output tensor dimensions. Format: [N, C\_OUT] N : Batches C\_OUT : Output depth H & W : Not used.
- [inout] `output_data`: Output data pointer. Data type: int8

```
int32_t riscv_fully_connected_s8_get_buffer_size(const nmsis_nn_dims *filter_dims)
```

Get the required buffer size for S8 basic fully-connected and matrix multiplication layer function for TF Lite.

**Return** The function returns required buffer size in bytes

**Parameters**

- [in] `filter_dims`: dimension of filter

## Pooling Functions

```
riscv_status riscv_avgpool_s8(const nmsis_nn_context *ctx, const nmsis_nn_pool_params
                             *pool_params, const nmsis_nn_dims *input_dims, const q7_t *input_data,
                             const nmsis_nn_dims *filter_dims, const nmsis_nn_dims
                             *output_dims, q7_t *output_data)
```

```
int32_t riscv_avgpool_s8_get_buffer_size(const int dim_dst_width, const int ch_src)
```

```
riscv_status riscv_maxpool_s8(const nmsis_nn_context *ctx, const nmsis_nn_pool_params
                              *pool_params, const nmsis_nn_dims *input_dims, const q7_t *input_data,
                              const nmsis_nn_dims *filter_dims, const nmsis_nn_dims
                              *output_dims, q7_t *output_data)
```

```
void riscv_maxpool_q7_HWC(q7_t *Im_in, const uint16_t dim_im_in, const uint16_t ch_im_in,
                          const uint16_t dim_kernel, const uint16_t padding, const uint16_t
                          stride, const uint16_t dim_im_out, q7_t *bufferA, q7_t *Im_out)
```

```
void riscv_avepool_q7_HWC (q7_t *Im_in, const uint16_t dim_im_in, const uint16_t ch_im_in,
                          const uint16_t dim_kernel, const uint16_t padding, const uint16_t
                          stride, const uint16_t dim_im_out, q7_t *bufferA, q7_t *Im_out)
```

*group* **Pooling**

Perform pooling functions, including max pooling and average pooling

## Functions

```
riscv_status riscv_avgpool_s8 (const nmsis_nn_context *ctx, const nmsis_nn_pool_params
                              *pool_params, const nmsis_nn_dims *input_dims, const q7_t
                              *input_data, const nmsis_nn_dims *filter_dims, const nm-
                              sis_nn_dims *output_dims, q7_t *output_data)
```

s8 average pooling function.

- Supported Framework: TensorFlow Lite

**Return** The function returns RISC\_V\_MATH\_SUCCESS - Successful operation

### Parameters

- [inout] ctx: Function context (e.g. temporary buffer). Check the function definition file to see if an additional buffer is required. Optional function {API}\_get\_buffer\_size() provides the buffer size if an additional buffer is required.
- [in] pool\_params: Pooling parameters
- [in] input\_dims: Input (activation) tensor dimensions. Format: [H, W, C\_IN] Argument 'N' is not used.
- [in] input\_data: Input (activation) data pointer. Data type: int8
- [in] filter\_dims: Filter tensor dimensions. Format: [H, W] Argument N and C are not used.
- [in] output\_dims: Output tensor dimensions. Format: [H, W, C\_OUT] Argument N is not used. C\_OUT equals C\_IN.
- [inout] output\_data: Output data pointer. Data type: int8

```
int32_t riscv_avgpool_s8_get_buffer_size (const int dim_dst_width, const int ch_src)
```

Get the required buffer size for S8 average pooling function.

**Return** The function returns required buffer size in bytes

### Parameters

- [in] dim\_dst\_width: output tensor dimension
- [in] ch\_src: number of input tensor channels

```
riscv_status riscv_max_pool_s8 (const nmsis_nn_context *ctx, const nmsis_nn_pool_params
                              *pool_params, const nmsis_nn_dims *input_dims, const q7_t
                              *input_data, const nmsis_nn_dims *filter_dims, const nm-
                              sis_nn_dims *output_dims, q7_t *output_data)
```

s8 max pooling function.

- Supported Framework: TensorFlow Lite

**Return** The function returns RISC\_V\_MATH\_SUCCESS - Successful operation

**Parameters**

- [inout] `ctx`: Function context (e.g. temporary buffer). Check the function definition file to see if an additional buffer is required. Optional function `{API}_get_buffer_size()` provides the buffer size if an additional buffer is required.
- [in] `pool_params`: Pooling parameters
- [in] `input_dims`: Input (activation) tensor dimensions. Format: [H, W, C\_IN] Argument 'N' is not used.
- [in] `input_data`: Input (activation) data pointer. Data type: int8
- [in] `filter_dims`: Filter tensor dimensions. Format: [H, W] Argument N and C are not used.
- [in] `output_dims`: Output tensor dimensions. Format: [H, W, C\_OUT] Argument N is not used. C\_OUT equals C\_IN.
- [inout] `output_data`: Output data pointer. Data type: int8

```
void riscv_maxpool_q7_HWC (q7_t *Im_in, const uint16_t dim_im_in, const uint16_t ch_im_in,  
                           const uint16_t dim_kernel, const uint16_t padding, const  
                           uint16_t stride, const uint16_t dim_im_out, q7_t *bufferA, q7_t  
                           *Im_out)
```

Q7 max pooling function.

The pooling function is implemented as split x-pooling then y-pooling.

**Parameters**

- [inout] `Im_in`: pointer to input tensor
- [in] `dim_im_in`: input tensor dimension
- [in] `ch_im_in`: number of input tensor channels
- [in] `dim_kernel`: filter kernel size
- [in] `padding`: padding sizes
- [in] `stride`: convolution stride
- [in] `dim_im_out`: output tensor dimension
- [inout] `bufferA`: Not used
- [inout] `Im_out`: pointer to output tensor

This pooling function is input-destructive. Input data is undefined after calling this function.

```
void riscv_avepool_q7_HWC (q7_t *Im_in, const uint16_t dim_im_in, const uint16_t ch_im_in,  
                           const uint16_t dim_kernel, const uint16_t padding, const  
                           uint16_t stride, const uint16_t dim_im_out, q7_t *bufferA, q7_t  
                           *Im_out)
```

Q7 average pooling function.

**Buffer size:****Parameters**

- [inout] `Im_in`: pointer to input tensor
- [in] `dim_im_in`: input tensor dimension
- [in] `ch_im_in`: number of input tensor channels

- [in] `dim_kernel`: filter kernel size
- [in] `padding`: padding sizes
- [in] `stride`: convolution stride
- [in] `dim_im_out`: output tensor dimension
- [inout] `bufferA`: pointer to buffer space for input
- [inout] `Im_out`: pointer to output tensor

`bufferA` size:  $2 * \text{dim\_im\_out} * \text{ch\_im\_in}$

The pooling function is implemented as split x-pooling then y-pooling.

This pooling function is input-destructive. Input data is undefined after calling this function.

## Reshape Functions

void **riscv\_reshape\_s8** (const int8\_t \*input, int8\_t \*output, const uint32\_t total\_size)

*group* **Reshape**

### Functions

void **riscv\_reshape\_s8** (const int8\_t \*input, int8\_t \*output, const uint32\_t total\_size)

Reshape a s8 vector into another with different shape.

Basic s8 reshape function.

Refer header file for details.

## Softmax Functions

void **riscv\_softmax\_q15** (const q15\_t \*vec\_in, const uint16\_t dim\_vec, q15\_t \*p\_out)

void **riscv\_softmax\_q7** (const q7\_t \*vec\_in, const uint16\_t dim\_vec, q7\_t \*p\_out)

void **riscv\_softmax\_s8** (const int8\_t \*input, const int32\_t num\_rows, const int32\_t row\_size, const int32\_t mult, const int32\_t shift, const int32\_t diff\_min, int8\_t \*output)

void **riscv\_softmax\_u8** (const uint8\_t \*input, const int32\_t num\_rows, const int32\_t row\_size, const int32\_t mult, const int32\_t shift, const int32\_t diff\_min, uint8\_t \*output)

void **riscv\_softmax\_with\_batch\_q7** (const q7\_t \*vec\_in, const uint16\_t nb\_batches, const uint16\_t dim\_vec, q7\_t \*p\_out)

*group* **Softmax**

EXP(2) based softmax functions.

### Functions

void **riscv\_softmax\_q15** (const q15\_t \*vec\_in, const uint16\_t dim\_vec, q15\_t \*p\_out)

Q15 softmax function.

Here, instead of typical e based softmax, we use 2-based softmax, i.e.,:

**Parameters**

- [in] `vec_in`: pointer to input vector
- [in] `dim_vec`: input vector dimension
- [out] `p_out`: pointer to output vector

$$y_i = 2^{(x_i)} / \text{sum}(2^{x_j})$$

The relative output will be different here. But mathematically, the gradient will be the same with a  $\log(2)$  scaling factor.

void **riscv\_softmax\_q7** (**const** q7\_t \**vec\_in*, **const** uint16\_t *dim\_vec*, q7\_t \**p\_out*)  
Q7 softmax function.

Here, instead of typical natural logarithm  $e$  based softmax, we use 2-based softmax here, i.e.,:

#### Parameters

- [in] `vec_in`: pointer to input vector
- [in] `dim_vec`: input vector dimension
- [out] `p_out`: pointer to output vector

$$y_i = 2^{(x_i)} / \text{sum}(2^{x_j})$$

The relative output will be different here. But mathematically, the gradient will be the same with a  $\log(2)$  scaling factor.

void **riscv\_softmax\_s8** (**const** int8\_t \**input*, **const** int32\_t *num\_rows*, **const** int32\_t *row\_size*,  
**const** int32\_t *mult*, **const** int32\_t *shift*, **const** int32\_t *diff\_min*, int8\_t  
\**output*)  
S8 softmax function.

**Note** Supported framework: TensorFlow Lite micro (bit-accurate)

#### Parameters

- [in] `input`: Pointer to the input tensor
- [in] `num_rows`: Number of rows in the input tensor
- [in] `row_size`: Number of elements in each input row
- [in] `mult`: Input quantization multiplier
- [in] `shift`: Input quantization shift within the range [0, 31]
- [in] `diff_min`: Minimum difference with max in row. Used to check if the quantized exponential operation can be performed
- [out] `output`: Pointer to the output tensor

void **riscv\_softmax\_u8** (**const** uint8\_t \**input*, **const** int32\_t *num\_rows*, **const** int32\_t *row\_size*,  
**const** int32\_t *mult*, **const** int32\_t *shift*, **const** int32\_t *diff\_min*, uint8\_t  
\**output*)  
U8 softmax function.

**Note** Supported framework: TensorFlow Lite micro (bit-accurate)

#### Parameters

- [in] `input`: Pointer to the input tensor
- [in] `num_rows`: Number of rows in the input tensor
- [in] `row_size`: Number of elements in each input row



- [in] `mult`: Input quantization multiplier
- [in] `shift`: Input quantization shift within the range [0, 31]
- [in] `diff_min`: Minimum difference with max in row. Used to check if the quantized exponential operation can be performed
- [out] `output`: Pointer to the output tensor

void **riscv\_softmax\_with\_batch\_q7** (**const** q7\_t \**vec\_in*, **const** uint16\_t *nb\_batches*, **const** uint16\_t *dim\_vec*, q7\_t \**p\_out*)

Q7 softmax function with batch parameter.

Here, instead of typical natural logarithm  $e$  based softmax, we use 2-based softmax here, i.e.,:

#### Parameters

- [in] `vec_in`: pointer to input vector
- [in] `nb_batches`: number of batches
- [in] `dim_vec`: input vector dimension
- [out] `p_out`: pointer to output vector

$$y_i = 2^{x_i} / \sum(2^{x_j})$$

The relative output will be different here. But mathematically, the gradient will be the same with a  $\log(2)$  scaling factor.

## SVDF Layer Functions

```
riscv_status riscv_svdf_s8 (const nmsis_nn_context *input_ctx, const nmsis_nn_context *output_ctx, const nmsis_nn_svdf_params *svdf_params, const nmsis_nn_per_tensor_quant_params *input_quant_params, const nmsis_nn_per_tensor_quant_params *output_quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *state_dims, q15_t *state_data, const nmsis_nn_dims *weights_feature_dims, const q7_t *weights_feature_data, const nmsis_nn_dims *weights_time_dims, const q15_t *weights_time_data, const nmsis_nn_dims *bias_dims, const q31_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

group **SVDF**

#### Functions

```
riscv_status riscv_svdf_s8 (const nmsis_nn_context *input_ctx, const nmsis_nn_context *output_ctx, const nmsis_nn_svdf_params *svdf_params, const nmsis_nn_per_tensor_quant_params *input_quant_params, const nmsis_nn_per_tensor_quant_params *output_quant_params, const nmsis_nn_dims *input_dims, const q7_t *input_data, const nmsis_nn_dims *state_dims, q15_t *state_data, const nmsis_nn_dims *weights_feature_dims, const q7_t *weights_feature_data, const nmsis_nn_dims *weights_time_dims, const q15_t *weights_time_data, const nmsis_nn_dims *bias_dims, const q31_t *bias_data, const nmsis_nn_dims *output_dims, q7_t *output_data)
```

s8 SVDF function

1. Supported framework: TensorFlow Lite micro
2. q7 is used as data type eventhough it is s8 data. It is done so to be consistent with existing APIs.

**Return** The function returns RISC\_V\_MATH\_SUCCESS

**Parameters**

- [in] input\_ctx: Temporary scratch buffer
- [in] output\_ctx: Temporary output scratch buffer
- [in] svdf\_params: SVDF Parameters Range of svdf\_params->input\_offset : [-128, 127]  
Range of svdf\_params->output\_offset : [-128, 127]
- [in] input\_quant\_params: Input quantization parameters
- [in] output\_quant\_params: Output quantization parameters
- [in] input\_dims: Input tensor dimensions
- [in] input\_data: Pointer to input tensor
- [in] state\_dims: State tensor dimensions
- [in] state\_data: Pointer to state tensor
- [in] weights\_feature\_dims: Weights (feature) tensor dimensions
- [in] weights\_feature\_data: Pointer to the weights (feature) tensor
- [in] weights\_time\_dims: Weights (time) tensor dimensions
- [in] weights\_time\_data: Pointer to the weights (time) tensor
- [in] bias\_dims: Bias tensor dimensions
- [in] bias\_data: Pointer to bias tensor
- [in] output\_dims: Output tensor dimensions
- [out] output\_data: Pointer to the output tensor

*group* **groupNN**

A collection of functions to perform basic operations for neural network layers. Functions with a \_s8 suffix support TensorFlow Lite framework.

### 4.3.2 Neural Network Data Conversion Functions

void **riscv\_q7\_to\_q15\_no\_shift** (const q7\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_q7\_to\_q15\_reordered\_no\_shift** (const q7\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

void **riscv\_q7\_to\_q15\_reordered\_with\_offset** (const q7\_t \*src, q15\_t \*dst, uint32\_t  
block\_size, q15\_t offset)

void **riscv\_q7\_to\_q15\_with\_offset** (const q7\_t \*src, q15\_t \*dst, uint32\_t block\_size, q15\_t offset)

void **riscv\_q7\_to\_q7\_no\_shift** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

void **riscv\_q7\_to\_q7\_reordered\_no\_shift** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

*group* **ndata\_convert**

Perform data type conversion in-between neural network operations

## Functions

void **riscv\_q7\_to\_q15\_no\_shift** (const q7\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

Converts the elements of the Q7 vector to Q15 vector without left-shift.

Converts the elements of the q7 vector to q15 vector without left-shift.

The equation used for the conversion process is:

### Description:

#### Parameters

- [in] \*pSrc: points to the Q7 input vector
- [out] \*pDst: points to the Q15 output vector
- [in] blockSize: length of the input vector

void **riscv\_q7\_to\_q15\_reordered\_no\_shift** (const q7\_t \*pSrc, q15\_t \*pDst, uint32\_t blockSize)

Converts the elements of the Q7 vector to reordered Q15 vector without left-shift.

Converts the elements of the q7 vector to reordered q15 vector without left-shift.

This function does the q7 to q15 expansion with re-ordering

#### Parameters

- [in] \*pSrc: points to the Q7 input vector
- [out] \*pDst: points to the Q15 output vector
- [in] blockSize: length of the input vector

is converted into:

This looks strange but is natural considering how sign-extension is done at assembly level.

The expansion of other other operand will follow the same rule so that the end results are the same.

The tail (i.e., last (N % 4) elements) will still be in original order.

void **riscv\_q7\_to\_q15\_reordered\_with\_offset** (const q7\_t \*src, q15\_t \*dst, uint32\_t blockSize, q15\_t offset)

Converts the elements of the Q7 vector to a reordered Q15 vector with an added offset.

Converts the elements of the q7 vector to reordered q15 vector with an added offset.

**Note** Refer header file for details.

void **riscv\_q7\_to\_q15\_with\_offset** (const q7\_t \*src, q15\_t \*dst, uint32\_t blockSize, q15\_t offset)

Converts the elements from a q7 vector to a q15 vector with an added offset.

The equation used for the conversion process is:

### Description:

#### Parameters

- [in] src: pointer to the q7 input vector
- [out] dst: pointer to the q15 output vector
- [in] block\_size: length of the input vector
- [in] offset: q7 offset to be added to each input vector element.

void **riscv\_q7\_to\_q7\_no\_shift** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

Converts the elements of the Q7 vector to Q7 vector without left-shift.

The equation used for the conversion process is:

**Return** none.

**Description:**

**Parameters**

- [in] \*pSrc: points to the Q7 input vector
- [out] \*pDst: points to the Q7 output vector
- [in] blockSize: length of the input vector

void **riscv\_q7\_to\_q7\_reordered\_no\_shift** (const q7\_t \*pSrc, q7\_t \*pDst, uint32\_t blockSize)

Converts the elements of the Q7 vector to reordered Q7 vector without left-shift.

This function does the q7 to q7 expansion with re-ordering

**Return** none.

**Parameters**

- [in] \*pSrc: points to the Q7 input vector
- [out] \*pDst: points to the Q7 output vector
- [in] blockSize: length of the input vector

is converted into:

This looks strange but is natural considering how sign-extension is done at assembly level.

The expansion of other other oprand will follow the same rule so that the end results are the same.

The tail (i.e., last (N % 4) elements) will still be in original order.

### 4.3.3 Basic Math Functions for Neural Network Computation

void **riscv\_nn\_accumulate\_q7\_to\_q15** (q15\_t \*dst, const q7\_t \*src, uint32\_t block\_size)

void **riscv\_nn\_accumulate\_q7\_to\_q7** (q7\_t \*dst, const q7\_t \*src, uint32\_t block\_size)

void **riscv\_nn\_add\_q7** (const q7\_t \*input, q31\_t \*output, uint32\_t block\_size)

q7\_t \***riscv\_nn\_depthwise\_conv\_nt\_t\_padded\_s8** (const q7\_t \*lhs, const q7\_t \*rhs, const int32\_t lhs\_offset, const uint16\_t num\_ch, const int32\_t \*out\_shift, const int32\_t \*out\_mult, const int32\_t out\_offset, const int32\_t activation\_min, const int32\_t activation\_max, const uint16\_t row\_x\_col, const int32\_t \*const output\_bias, q7\_t \*out)

q7\_t \***riscv\_nn\_depthwise\_conv\_nt\_t\_s8** (const q7\_t \*lhs, const q7\_t \*rhs, const int32\_t lhs\_offset, const uint16\_t num\_ch, const int32\_t \*out\_shift, const int32\_t \*out\_mult, const int32\_t out\_offset, const int32\_t activation\_min, const int32\_t activation\_max, const uint16\_t row\_x\_col, const int32\_t \*const output\_bias, q7\_t \*out)

```

riscv_status riscv_nn_mat_mul_core_1x_s8 (int32_t row_elements, const int8_t *row_base, const
int8_t *col_base, int32_t *const sum_col, int32_t
*const output)

riscv_status riscv_nn_mat_mul_core_4x_s8 (const int32_t row_elements, const int32_t offset,
const int8_t *row_base, const int8_t *col_base,
int32_t *const sum_col, int32_t *const output)

riscv_status riscv_nn_mat_mult_nt_t_s8 (const q7_t *lhs, const q7_t *rhs, const q31_t
*bias, q7_t *dst, const int32_t *dst_multipliers, const
int32_t *dst_shifts, const int32_t lhs_rows, const
int32_t rhs_rows, const int32_t rhs_cols, const int32_t
lhs_offset, const int32_t dst_offset, const int32_t activa-
tion_min, const int32_t activation_max)

void riscv_nn_mult_q15 (q15_t *pSrcA, q15_t *pSrcB, q15_t *pDst, const uint16_t out_shift, uint32_t
blockSize)

void riscv_nn_mult_q7 (q7_t *pSrcA, q7_t *pSrcB, q7_t *pDst, const uint16_t out_shift, uint32_t block-
Size)

riscv_status riscv_nn_vec_mat_mult_t_s8 (const q7_t *lhs, const q7_t *rhs, const q31_t *bias,
q7_t *dst, const int32_t lhs_offset, const int32_t
rhs_offset, const int32_t dst_offset, const int32_t
dst_multiplier, const int32_t dst_shift, const int32_t
rhs_cols, const int32_t rhs_rows, const int32_t acti-
vation_min, const int32_t activation_max)

riscv_status riscv_nn_vec_mat_mult_t_svd_s8 (const q7_t *lhs, const q7_t *rhs, q15_t
*dst, const int32_t lhs_offset, const int32_t
rhs_offset, const int32_t scatter_offset, const
int32_t dst_multiplier, const int32_t dst_shift,
const int32_t rhs_cols, const int32_t rhs_rows,
const int32_t activation_min, const int32_t ac-
tivation_max)

```

#### group **NNBasicMath**

Basic Math Functions for Neural Network Computation

### Functions

void **riscv\_nn\_accumulate\_q7\_to\_q15** (q15\_t \*dst, const q7\_t \*src, uint32\_t block\_size)  
 Converts the elements from a q7 vector and accumulate to a q15 vector.

The equation used for the conversion process is:

#### Description:

#### Parameters

- [in] \*src: points to the q7 input vector
- [out] \*dst: points to the q15 output vector
- [in] block\_size: length of the input vector

void **riscv\_nn\_accumulate\_q7\_to\_q7** (q7\_t \*dst, const q7\_t \*src, uint32\_t block\_size)  
 Converts the elements from a q7 vector and accumulate to a q7 vector.

The equation used for the conversion process is:

#### Description:

**Parameters**

- [in] *\*src*: points to the q7 input vector
- [out] *\*dst*: points to the q7 output vector
- [in] *block\_size*: length of the input vector

void **riscv\_nn\_add\_q7**(**const** q7\_t *\*input*, q31\_t *\*output*, uint32\_t *block\_size*)

Non-saturating addition of elements of a q7 vector.

2<sup>24</sup> samples can be added without saturating the result.

**Description:****Parameters**

- [in] *\*input*: Pointer to the q7 input vector
- [out] *\*output*: Pointer to the q31 output variable.
- [in] *block\_size*: length of the input vector

The equation used for the conversion process is:

```
q7_t *riscv_nn_depthwise_conv_nt_t_padded_s8(const q7_t *lhs, const q7_t *rhs,  
                                              const int32_t lhs_offset, const  
                                              uint16_t num_ch, const int32_t  
                                              *out_shift, const int32_t *out_mult,  
                                              const int32_t out_offset, const  
                                              int32_t activation_min, const int32_t  
                                              activation_max, const uint16_t  
                                              row_x_col, const int32_t *const  
                                              output_bias, q7_t *out)
```

Depthwise convolution of transposed rhs matrix with 4 lhs matrices. To be used in padded cases where the padding is -lhs\_offset(Range: int8). Dimensions are the same for lhs and rhs.

**Return** The function returns one of the two

- Updated output pointer if an implementation is available
- NULL if no implementation is available.

**Note** If number of channels is not a multiple of 4, upto 3 elements outside the boundary will be read out for the following.

- Output shift
- Output multiplier
- Output bias
- rhs

**Parameters**

- [in] *lhs*: Input left-hand side matrix
- [in] *rhs*: Input right-hand side matrix (transposed)
- [in] *lhs\_offset*: LHS matrix offset(input offset). Range: -127 to 128
- [in] *num\_ch*: Number of channels in LHS/RHS
- [in] *out\_shift*: Per channel output shift. Length of vector is equal to number of channels

- [in] out\_mult: Per channel output multiplier. Length of vector is equal to number of channels
- [in] out\_offset: Offset to be added to the output values. Range: -127 to 128
- [in] activation\_min: Minimum value to clamp the output to. Range: int8
- [in] activation\_max: Maximum value to clamp the output to. Range: int8
- [in] row\_x\_col: (row\_dimension \* col\_dimension) of LHS/RHS matrix
- [in] output\_bias: Per channel output bias. Length of vector is equal to number of channels
- [in] out: Output pointer

```
q7_t *riscv_nn_depthwise_conv_nt_t_s8(const q7_t *lhs, const q7_t *rhs, const
                                     int32_t lhs_offset, const uint16_t num_ch,
                                     const int32_t *out_shift, const int32_t
                                     *out_mult, const int32_t out_offset, const
                                     int32_t activation_min, const int32_t activa-
                                     tion_max, const uint16_t row_x_col, const
                                     int32_t *const output_bias, q7_t *out)
```

Depthwise convolution of transposed rhs matrix with 4 lhs matrices. To be used in non-padded cases. Dimensions are the same for lhs and rhs.

**Return** The function returns one of the two

- Updated output pointer if an implementation is available
- NULL if no implementation is available.

**Note** If number of channels is not a multiple of 4, upto 3 elements outside the boundary will be read out for the following.

- Output shift
- Output multiplier
- Output bias
- rhs

#### Parameters

- [in] lhs: Input left-hand side matrix
- [in] rhs: Input right-hand side matrix (transposed)
- [in] lhs\_offset: LHS matrix offset(input offset). Range: -127 to 128
- [in] num\_ch: Number of channels in LHS/RHS
- [in] out\_shift: Per channel output shift. Length of vector is equal to number of channels.
- [in] out\_mult: Per channel output multiplier. Length of vector is equal to number of channels.
- [in] out\_offset: Offset to be added to the output values. Range: -127 to 128
- [in] activation\_min: Minimum value to clamp the output to. Range: int8
- [in] activation\_max: Maximum value to clamp the output to. Range: int8
- [in] row\_x\_col: (row\_dimension \* col\_dimension) of LHS/RHS matrix
- [in] output\_bias: Per channel output bias. Length of vector is equal to number of channels.

- [in] out: Output pointer

```
riscv_status riscv_nn_mat_mul_core_1x_s8 (int32_t row_elements, const int8_t *row_base,  
                                          const int8_t *col_base, int32_t *const  
                                          sum_col, int32_t *const output)
```

General Matrix-multiplication without requantization for one row & one column.

Pseudo-code  $*output = 0$   $sum\_col = 0$  for ( $i = 0$ ;  $i < row\_elements$ ;  $i++$ )  $*output += row\_base[i] * col\_base[i]$   $sum\_col += col\_base[i]$

**Return** The function returns the multiply-accumulated result of the row by column.

#### Parameters

- [in] row\_elements: number of row elements
- [in] row\_base: pointer to row operand
- [in] col\_base: pointer to col operand
- [out] sum\_col: pointer to store sum of column elements
- [out] output: pointer to store result of multiply-accumulate

```
riscv_status riscv_nn_mat_mul_core_4x_s8 (const int32_t row_elements, const int32_t off-  
                                          set, const int8_t *row_base, const int8_t  
                                          *col_base, int32_t *const sum_col, int32_t  
                                          *const output)
```

General Matrix-multiplication without requantization for four rows and one column.

Pseudo-code  $output[0] = 0 .. output[3] = 0$   $sum\_col = 0$  for ( $i = 0$ ;  $i < row\_elements$ ;  $i++$ )  $output[0] += row\_base[i] * col\_base[i] .. output[3] += row\_base[i + (row\_elements * 3)] * col\_base[i]$   $sum\_col += col\_base[i]$

**Return** The function returns the multiply-accumulated result of the row by column

#### Parameters

- [in] row\_elements: number of row elements
- [in] offset: offset between rows. Can be the same as row\_elements. For e.g, in a 1x1 conv scenario with stride as 1.
- [in] row\_base: pointer to row operand
- [in] col\_base: pointer to col operand
- [out] sum\_col: pointer to store sum of column elements
- [out] output: pointer to store result(4 int32's) of multiply-accumulate

```
riscv_status riscv_nn_mat_mult_nt_t_s8 (const q7_t *lhs, const q7_t *rhs, const q31_t  
                                          *bias, q7_t *dst, const int32_t *dst_multipliers,  
                                          const int32_t *dst_shifts, const int32_t lhs_rows,  
                                          const int32_t rhs_rows, const int32_t rhs_cols,  
                                          const int32_t lhs_offset, const int32_t dst_offset,  
                                          const int32_t activation_min, const int32_t acti-  
                                          vation_max)
```

General Matrix-multiplication function with per-channel requantization. This function assumes:

- LHS input matrix NOT transposed (nt)
- RHS input matrix transposed (t)



**Note** This operation also performs the broadcast bias addition before the requantization

**Return** The function returns `RISCV_MATH_SUCCESS`

**Parameters**

- [in] `lhs`: Pointer to the LHS input matrix
- [in] `rhs`: Pointer to the RHS input matrix
- [in] `bias`: Pointer to the bias vector. The length of this vector is equal to the number of output columns (or RHS input rows)
- [out] `dst`: Pointer to the output matrix with “m” rows and “n” columns
- [in] `dst_multipliers`: Pointer to the multipliers vector needed for the per-channel requantization. The length of this vector is equal to the number of output columns (or RHS input rows)
- [in] `dst_shifts`: Pointer to the shifts vector needed for the per-channel requantization. The length of this vector is equal to the number of output columns (or RHS input rows)
- [in] `lhs_rows`: Number of LHS input rows
- [in] `rhs_rows`: Number of RHS input rows
- [in] `rhs_cols`: Number of LHS/RHS input columns
- [in] `lhs_offset`: Offset to be applied to the LHS input value
- [in] `dst_offset`: Offset to be applied the output result
- [in] `activation_min`: Minimum value to clamp down the output. Range : int8
- [in] `activation_max`: Maximum value to clamp up the output. Range : int8

void **riscv\_nn\_mult\_q15** (q15\_t \*pSrcA, q15\_t \*pSrcB, q15\_t \*pDst, **const** uint16\_t out\_shift, uint32\_t blockSize)

Q7 vector multiplication with variable output shifts.

q7 vector multiplication with variable output shifts

**Scaling and Overflow Behavior:**

The function uses saturating arithmetic. Results outside of the allowable Q15 range [0x8000 0x7FFF] will be saturated.

**Parameters**

- [in] `*pSrcA`: pointer to the first input vector
- [in] `*pSrcB`: pointer to the second input vector
- [out] `*pDst`: pointer to the output vector
- [in] `out_shift`: amount of right-shift for output
- [in] `blockSize`: number of samples in each vector

void **riscv\_nn\_mult\_q7** (q7\_t \*pSrcA, q7\_t \*pSrcB, q7\_t \*pDst, **const** uint16\_t out\_shift, uint32\_t blockSize)

Q7 vector multiplication with variable output shifts.

q7 vector multiplication with variable output shifts

**Scaling and Overflow Behavior:**

The function uses saturating arithmetic. Results outside of the allowable Q7 range [0x80 0x7F] will be saturated.

#### Parameters

- [in] \*pSrcA: pointer to the first input vector
- [in] \*pSrcB: pointer to the second input vector
- [out] \*pDst: pointer to the output vector
- [in] out\_shift: amount of right-shift for output
- [in] blockSize: number of samples in each vector

```
riscv_status riscv_nn_vec_mat_mult_t_s8 (const q7_t *lhs, const q7_t *rhs, const q31_t  
                                         *bias, q7_t *dst, const int32_t lhs_offset, const  
                                         int32_t rhs_offset, const int32_t dst_offset, const  
                                         int32_t dst_multiplier, const int32_t dst_shift,  
                                         const int32_t rhs_cols, const int32_t rhs_rows,  
                                         const int32_t activation_min, const int32_t acti-  
                                         vation_max)
```

s8 Vector by Matrix (transposed) multiplication

**Return** The function returns RISC\_V\_MATH\_SUCCESS

#### Parameters

- [in] lhs: Input left-hand side vector
- [in] rhs: Input right-hand side matrix (transposed)
- [in] bias: Input bias
- [out] dst: Output vector
- [in] lhs\_offset: Offset to be added to the input values of the left-hand side vector. Range: -127 to 128
- [in] rhs\_offset: Not used
- [in] dst\_offset: Offset to be added to the output values. Range: -127 to 128
- [in] dst\_multiplier: Output multiplier
- [in] dst\_shift: Output shift
- [in] rhs\_cols: Number of columns in the right-hand side input matrix
- [in] rhs\_rows: Number of rows in the right-hand side input matrix
- [in] activation\_min: Minimum value to clamp the output to. Range: int8
- [in] activation\_max: Maximum value to clamp the output to. Range: int8

```
riscv_status riscv_nn_vec_mat_mult_t_svd_s8 (const q7_t *lhs, const q7_t *rhs, q15_t  
                                         *dst, const int32_t lhs_offset, const  
                                         int32_t rhs_offset, const int32_t scat-  
                                         ter_offset, const int32_t dst_multiplier,  
                                         const int32_t dst_shift, const int32_t  
                                         rhs_cols, const int32_t rhs_rows, const  
                                         int32_t activation_min, const int32_t acti-  
                                         vation_max)
```

s8 Vector by Matrix (transposed) multiplication with s16 output

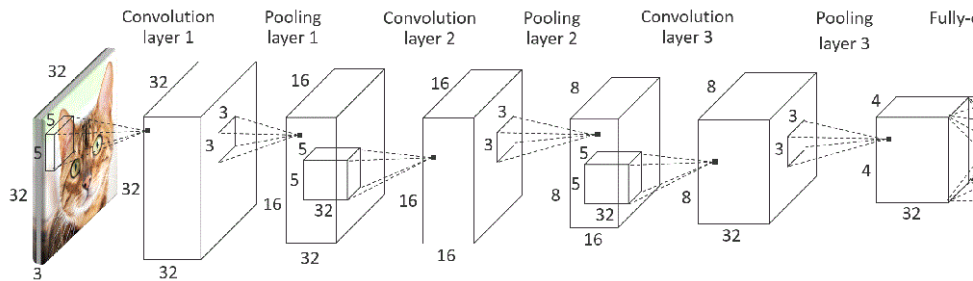
**Return** The function returns `RISCV_MATH_SUCCESS`

**Parameters**

- [in] `lhs`: Input left-hand side vector
- [in] `rhs`: Input right-hand side matrix (transposed)
- [out] `dst`: Output vector
- [in] `lhs_offset`: Offset to be added to the input values of the left-hand side vector. Range: -127 to 128
- [in] `rhs_offset`: Not used
- [in] `scatter_offset`: Address offset for `dst`. First output is stored at 'dst', the second at 'dst + scatter\_offset' and so on.
- [in] `dst_multiplier`: Output multiplier
- [in] `dst_shift`: Output shift
- [in] `rhs_cols`: Number of columns in the right-hand side input matrix
- [in] `rhs_rows`: Number of rows in the right-hand side input matrix
- [in] `activation_min`: Minimum value to clamp the output to. Range: int16
- [in] `activation_max`: Maximum value to clamp the output to. Range: int16

### 4.3.4 Convolutional Neural Network Example

group **CNNexample**



**Refer** `riscv_nnexamples_cifar10.cpp`

**Description:**

Demonstrates a convolutional neural network (CNN) example with the use of convolution, ReLU activation, pooling and fully-connected functions.

**Model definition:**

The CNN used in this example is based on CIFAR-10 example from Caffe [1]. The neural network consists of 3 convolution layers interspersed by ReLU activation and max pooling layers, followed by a fully-connected layer at the end. The input to the network is a 32x32 pixel color image, which will be classified into one of the 10 output classes. This example model implementation needs 32.3 KB to store weights, 40 KB for activations and 3.1 KB for storing the `im2col` data.

**Variables Description:**

- `conv1_wt`, `conv2_wt`, `conv3_wt` are convolution layer weight matrices
- `conv1_bias`, `conv2_bias`, `conv3_bias` are convolution layer bias arrays

- `ip1_wt`, `ip1_bias` point to fully-connected layer weights and biases
- `input_data` points to the input image data
- `output_data` points to the classification output
- `col_buffer` is a buffer to store the `im2col` output
- `scratch_buffer` is used to store the activation data (intermediate layer outputs)

#### NMSIS DSP Software Library Functions Used:

- `riscv_convolve_HWC_q7_RGB()`
- `riscv_convolve_HWC_q7_fast()`
- `riscv_relu_q7()`
- `riscv_maxpool_q7_HWC()`
- `riscv_avepool_q7_HWC()`
- `riscv_fully_connected_q7_opt()`
- `riscv_fully_connected_q7()`

[1] <https://github.com/BVLC/caffe>

### 4.3.5 Gated Recurrent Unit Example

#### group GRUExample

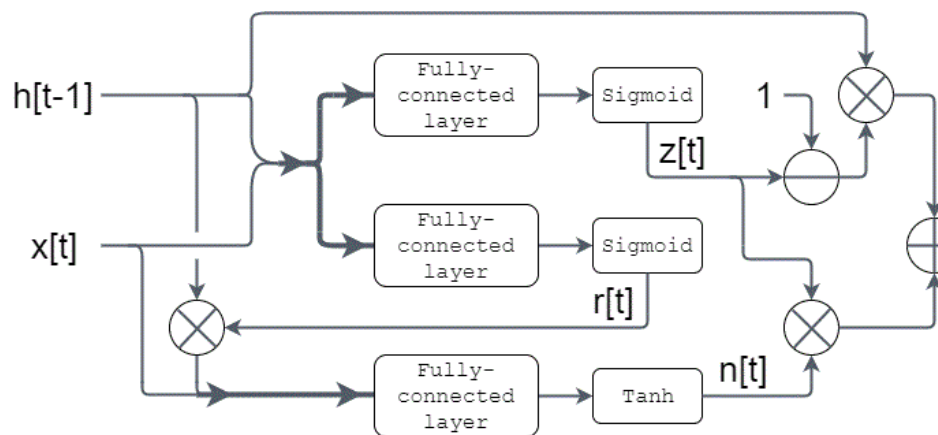
Refer `riscv_nnexamples_gru.cpp`

#### Description:

Demonstrates a gated recurrent unit (GRU) example with the use of fully-connected, Tanh/Sigmoid activation functions.

#### Model definition:

GRU is a type of recurrent neural network (RNN). It contains two sigmoid gates and one hidden state.



The computation can be summarized as:

#### Variables Description:

- `update_gate_weights`, `reset_gate_weights`, `hidden_state_weights` are weights corresponding to update gate ( $W_z$ ), reset gate ( $W_r$ ), and hidden state ( $W_n$ ).
- `update_gate_bias`, `reset_gate_bias`, `hidden_state_bias` are layer bias arrays
- `test_input1`, `test_input2`, `test_history` are the inputs and initial history

The buffer is allocated as:

| reset | input | history | update | hidden\_state |

In this way, the concatenation is automatically done since (reset, input) and (input, history) are physically concatenated in memory.

The ordering of the weight matrix should be adjusted accordingly.

#### NMSIS DSP Software Library Functions Used:

- `riscv_fully_connected_mat_q7_vec_q15_opt()`
- `riscv_nn_activations_direct_q15()`
- `riscv_mult_q15()`
- `riscv_offset_q15()`
- `riscv_sub_q15()`
- `riscv_copy_q15()`

## 4.4 Changelog

### 4.4.1 V1.0.2

This is release 1.0.2 version of NMSIS-NN library.

- Sync up to CMSIS NN library 3.0.0
- Initial support for RISC-V vector extension support

### 4.4.2 V1.0.1

This is release V1.0.1 version of NMSIS-DSP library.

- Both Nuclei RISC-V 32 and 64 bit cores are supported now.
- Libraries are optimized for RISC-V 32 and 64 bit DSP instructions.
- The DSP examples are now using Nuclei SDK as running environment.

### 4.4.3 V1.0.0

This is the first version of NMSIS-NN library.

We adapt the CMSIS-NN v1.0.0 library to use RISC-V DSP instructions, all the API names now are renamed from `arm_xxx` to `riscv_xxx`.



## CHANGELOG

### 5.1 V1.0.2

This is the official release version V1.0.2 release of Nuclei MCU Software Interface Standard(NMSIS).

The following changes has been made since V1.0.1.

- **Documentation**
  - Update NMSIS Core/DSP/NN related documentation
- **Device Templates**
  - DOWNLOAD\_MODE\_XXX macros are removed from riscv\_encoding.h, it is now defined as enum in <Device.h>, and can be customized by soc vendor.
  - startup code now don't rely on DOWNLOAD\_MODE macro, instead it now rely on a new macro called VECTOR\_TABLE\_REMAPPED, when VECTOR\_TABLE\_REMAPPED is defined, it means the vector table's lma != vma, such as vector table need to be copied from flash to ilm when boot up
  - Add more customized csr of Nuclei RISC-V Core
  - Add **BIT**, **BITS**, **REG**, **ADDR** related macros in <Device.h>
- **NMSIS-Core**
  - Nuclei Cache CCM operation APIs are now introduced in core\_feature\_cache.h
  - Update NMSIS-Core header files
- **NMSIS-DSP/NN**
  - Merged the official CMSIS 5.8.0 release, CMSIS-DSP 1.9.0, CMSIS-NN 3.0.0
  - RISC-V Vector extension and P-extension support for DSP/NN libraries are added

### 5.2 V1.0.2-RC2

This is the release candidate version V1.0.2-RC2 release of Nuclei MCU Software Interface Standard(NMSIS).

The following changes has been made since V1.0.2-RC1.

- **Documentation**
  - Update NMSIS Core/DSP/NN related documentation

## 5.3 V1.0.2-RC1

This is the release candidate version V1.0.2-RC1 release of Nuclei MCU Software Interface Standard(NMSIS).

The following changes has been made since V1.0.1.

- **Device Templates**
  - DOWNLOAD\_MODE\_XXX macros are removed from riscv\_encoding.h, it is now defined as enum in <Device.h>, and can be customized by soc vendor.
  - startup code now don't rely on DOWNLOAD\_MODE macro, instead it now rely on a new macro called VECTOR\_TABLE\_REMAPPED, when VECTOR\_TABLE\_REMAPPED is defined, it means the vector table's lma != vma, such as vector table need to be copied from flash to ilm when boot up
  - Add **BIT**, **BITS**, **REG**, **ADDR** related macros in <Device.h>
- **NMSIS-Core**
  - Nuclei Cache CCM operation APIs are now introduced in core\_feature\_cache.h
- **NMSIS-DSP/NN**
  - Merged the official CMSIS 5.8.0 release, CMSIS-DSP 1.9.0, CMSIS-NN 3.0.0
  - RISC-V Vector extension and P-extension support for DSP/NN libraries are added

## 5.4 V1.0.1

This is the official V1.0.1 release of Nuclei MCU Software Interface Standard(NMSIS).

The following changes has been maded since V1.0.1-RC1.

- **Device Templates**
  - I/D Cache enable assemble code in startup\_<Device>.S are removed now
  - Cache control updates in System\_<Device>.c
    - \* I-Cache will be enabled if \_\_ICACHE\_PRESENT = 1 defined in <Device.h>
    - \* D-Cache will be enabled if \_\_DCACHE\_PRESENT = 1 defined in <Device.h>

## 5.5 V1.0.1-RC1

This is release candidate version V1.0.1-RC1 of NMSIS.

- **NMSIS-Core**
  - Add RISC-V DSP 64bit intrinsic functions in core\_feature\_dsp.h
  - Add more CSR definitions in riscv\_encoding.h
  - Update arm compatible functions for RISC-V dsp instruction cases in core\_compatible.h
- **NMSIS-DSP**
  - Optimize RISC-V 32bit DSP library implementation
  - Add support for Nuclei RISC-V 64bit DSP SIMD instruction for DSP library
  - Add test cases used for DSP library testing, mainly for internal usage



- Change the examples and tests to use Nuclei SDK as running environment
- **NMSIS-NN**
  - Add support for Nuclei RISC-V 64bit DSP SIMD instruction for NN library
  - Change the examples and tests to use Nuclei SDK as running environment
- **Device Templates**
  - Add DDR\_DOWNLOAD\_MODE in device templates
  - Modifications to `startup_<Device>.S` files
    - \* `_premain_init` is added to replace `_init`
    - \* `_postmain_fini` is added to replace `_fini`
  - If you have implemented your init or de-init functions through `_init` or `_fini`, please use `_premain_init` and `_postmain_fini` functions defined `system_<Device>.c` now

## 5.6 V1.0.0-beta1

Main changes in release **V1.0.0-beta1**.

- **NMSIS-Core**
  - Fix `SysTick_Reload` implementation
  - Update `ECLIC_Register_IRQ` implementation to allow handler == NULL
  - Fix MTH offset from 0x8 to 0xB, this will affect function of `ECLIC_GetMth` and `ECLIC_SetMth`
  - Fix wrong macro check in cache function
  - Add missing `SOC_INT_MAX` enum definition in Device template
  - In `System_<Device>.c`, `ECLIC_NLBits` set to `__ECLIC_INTCTLBITS`, which means all the bits are for level, no bits for priority

## 5.7 V1.0.0-beta

Main changes in release **V1.0.0-beta**.

- **NMSIS-Core**
  - Fix error typedef of `CSR_MCAUSE_Type`
  - Change `CSR_MCACHE_CTL_DE` to future value 0x00010000
  - Fix names in CSR naming, `CSR_SCRATCHCSW` -> `CSR_MSCRATCHCSW`, and `CSR_SCRATCHCSWL` -> `CSR_MSCRATCHCSWL`
  - Add macros in `riscv_encoding.h`: `MSTATUS_FS_INITIAL`, `MSTATUS_FS_CLEAN`, `MSTATUS_FS_DIRTY`
- **Documentation**
  - Fix an typo in `core_template_intexc.rst`
  - Add cross references of Nuclei ISA Spec
  - Update appendix

- Refines tables and figures

## 5.8 V1.0.0-alpha.1

API changes has been made to system timer.

- Start from Nuclei N core version 1.4, MSTOP register is renamed to MTIMECTL to provide more features
- Changes made to NMSIS/Core/core\_feature\_timer.h
  - MSTOP register name changed to MTIMECTL due to core spec changes
  - SysTimer\_SetMstopValue renamed to SysTimer\_SetControlValue
  - SysTimer\_GetMstopValue renamed to SysTimer\_GetControlValue
  - Add SysTimer\_Start and SysTimer\_Stop to start or stop system timer counter
  - SysTick\_Reload function is introduced to reload system timer
  - Macro names started with SysTimer\_xxx are changed, please check in the code.
- Removed unused lines of code in DSP and NN library source code which has unused macros which will not work for RISC-V cores.
- Fix some documentation issues, mainly typos and invalid cross references.

## 5.9 V1.0.0-alpha

This is the V1.0.0-alpha release of Nuclei MCU Software Interface Standard(NMSIS).

In this release, we have release three main compoments:

- **NMSIS-Core:** Standardized API for the Nuclei processor core and peripherals.
- **NMSIS-DSP:** DSP library collection optimized for the Nuclei Processors which has RISC-V SIMD instruction set.
- **NMSIS-NN:** Efficient neural network library developed to maximize the performance and minimize the memory footprint Nuclei Processors which has RISC-V SIMD instruction set.

We also released totally new [Nuclei-SDK](https://github.com/Nuclei-Software/nuclei-sdk)<sup>24</sup> which is an SDK implementation based on the **NMSIS-Core** for Nuclei N/NX evaluation cores running on HummingBird Evaluation Kit.

---

<sup>24</sup> <https://github.com/Nuclei-Software/nuclei-sdk>

**GLOSSARY**

**API** (Application Program Interface) A defined set of routines and protocols for building application software.

**DSP** (Digital Signal Processing) is the use of digital processing, such as by computers or more specialized digital signal processors, to perform a wide variety of signal processing operations.

**ISR** (Interrupt Service Routine) Also known as an interrupt handler, an ISR is a callback function whose execution is triggered by a hardware interrupt (or software interrupt instructions) and is used to handle high-priority conditions that require interrupting the current code executing on the processor.

**NN** (Neural Network) is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes.

**XIP** (eXecute In Place) a method of executing programs directly from long term storage rather than copying it into RAM, saving writable memory for dynamic data and not the static program code.



## APPENDIX

- **Nuclei Tools and Documents:** <https://nucleisys.com/download.php>
- **Nuclei riscv-openocd Repo:** <https://github.com/riscv-mcu/riscv-openocd>
- **Nuclei riscv-binutils-gdb:** <https://github.com/riscv-mcu/riscv-binutils-gdb>
- **Nuclei riscv-gnu-toolchain:** <https://github.com/riscv-mcu/riscv-gnu-toolchain>
- **Nuclei riscv-newlib:** <https://github.com/riscv-mcu/riscv-newlib>
- **Nuclei riscv-gcc:** <https://github.com/riscv-mcu/riscv-gcc>
- **Nuclei SDK:** <https://github.com/Nuclei-Software/nuclei-sdk>
- **NMSIS:** <https://doc.nucleisys.com/nmsis/>
- **Nuclei Bumblebee Core Document:** [https://github.com/nucleisys/Bumblebee\\_Core\\_Doc](https://github.com/nucleisys/Bumblebee_Core_Doc)
- **Nuclei RISC-V IP Products:** <https://www.nucleisys.com/product.php>
- **RISC-V MCU Community Website:** <https://www.riscv-mcu.com/>
- **Nuclei Spec:** [https://doc.nucleisys.com/nuclei\\_spec](https://doc.nucleisys.com/nuclei_spec)
- **RISC-V Packed SIMD(P) Extension Spec:** <https://github.com/riscv/riscv-p-spec>
- **RISC-V Vector(V) Extension Spec:** <https://github.com/riscv/riscv-v-spec>



## INDICES AND TABLES

- `genindex`
- `search`





## Symbols

- `__FLD2VAL` (*C macro*), 356, 357
- `__VAL2FLD` (*C macro*), 356
- `__ALIGNED` (*C macro*), 60
- `__ASM` (*C macro*), 59
- `__CLZ` (*C macro*), 410
- `__COMPILER_BARRIER` (*C macro*), 60
- `__CPU_RELAX` (*C macro*), 94
- `__DMB` (*C macro*), 409
- `__DSB` (*C macro*), 409
- `__FENCE` (*C macro*), 94
- `__I` (*C macro*), 356
- `__IM` (*C macro*), 356
- `__INLINE` (*C macro*), 59
- `__INTERRUPT` (*C macro*), 60
- `__IO` (*C macro*), 356
- `__IOM` (*C macro*), 356
- `__ISB` (*C macro*), 409
- `__LDRBT` (*C macro*), 409
- `__LDRHT` (*C macro*), 409
- `__LDRT` (*C macro*), 409
- `__NMSIS_VERSION` (*C macro*), 58
- `__NMSIS_VERSION_MAJOR` (*C macro*), 58
- `__NMSIS_VERSION_MINOR` (*C macro*), 58
- `__NMSIS_VERSION_PATCH` (*C macro*), 58
- `__NO_RETURN` (*C macro*), 59
- `__NUCLEI_NX_REV` (*C macro*), 58
- `__NUCLEI_N_REV` (*C macro*), 58
- `__O` (*C macro*), 356
- `__OM` (*C macro*), 356
- `__PACKED` (*C macro*), 59
- `__PACKED_STRUCT` (*C macro*), 59
- `__PACKED_UNION` (*C macro*), 59
- `__RARELY` (*C macro*), 60
- `__RBIT` (*C macro*), 410
- `__RESTRICT` (*C macro*), 60
- `__RISCV_FLEN` (*C macro*), 382
- `__RISCV_XLEN` (*C macro*), 78
- `__RMB` (*C macro*), 94
- `__RV_BITREVI` (*C macro*), 205, 206
- `__RV_CSR_CLEAR` (*C macro*), 62
- `__RV_CSR_READ` (*C macro*), 61
- `__RV_CSR_READ_CLEAR` (*C macro*), 61
- `__RV_CSR_READ_SET` (*C macro*), 61
- `__RV_CSR_SET` (*C macro*), 61
- `__RV_CSR_SWAP` (*C macro*), 60
- `__RV_CSR_WRITE` (*C macro*), 61
- `__RV_FLD` (*C macro*), 383
- `__RV_FLOAD` (*C macro*), 384
- `__RV_FLW` (*C macro*), 383
- `__RV_FSD` (*C macro*), 384
- `__RV_FSTORE` (*C macro*), 384
- `__RV_FSW` (*C macro*), 383
- `__RV_INSB` (*C macro*), 205, 206
- `__RV_KSLLI16` (*C macro*), 125, 126
- `__RV_KSLLI8` (*C macro*), 136
- `__RV_KSLLIW` (*C macro*), 188, 189
- `__RV_SCLIP16` (*C macro*), 165, 166
- `__RV_SCLIP32` (*C macro*), 254
- `__RV_SCLIP8` (*C macro*), 171
- `__RV_SLLI16` (*C macro*), 125, 126
- `__RV_SLLI8` (*C macro*), 136, 137
- `__RV_SRAI16` (*C macro*), 125, 127
- `__RV_SRAI16_U` (*C macro*), 125, 127
- `__RV_SRAI8` (*C macro*), 136, 137
- `__RV_SRAI8_U` (*C macro*), 136, 138
- `__RV_SRAI_U` (*C macro*), 206, 207
- `__RV_SRLI16` (*C macro*), 125, 128
- `__RV_SRLI16_U` (*C macro*), 125, 129
- `__RV_SRLI8` (*C macro*), 136, 139
- `__RV_SRLI8_U` (*C macro*), 136, 139
- `__RV_UCLIP16` (*C macro*), 165, 166
- `__RV_UCLIP32` (*C macro*), 254, 255
- `__RV_UCLIP8` (*C macro*), 171, 172
- `__RV_WEXTI` (*C macro*), 206, 207
- `__RWMB` (*C macro*), 94
- `__SMP_RMB` (*C macro*), 94
- `__SMP_RWMB` (*C macro*), 94
- `__SMP_WMB` (*C macro*), 94
- `__SSAT` (*C macro*), 410
- `__STATIC_FORCEINLINE` (*C macro*), 59
- `__STATIC_INLINE` (*C macro*), 59
- `__STRBT` (*C macro*), 409
- `__STRHT` (*C macro*), 409

\_\_STRT (*C macro*), 410  
 \_\_UNALIGNED\_UINT16\_READ (*C macro*), 59  
 \_\_UNALIGNED\_UINT16\_WRITE (*C macro*), 59  
 \_\_UNALIGNED\_UINT32\_READ (*C macro*), 59  
 \_\_UNALIGNED\_UINT32\_WRITE (*C macro*), 59  
 \_\_USAT (*C macro*), 410  
 \_\_USED (*C macro*), 59  
 \_\_USUALLY (*C macro*), 60  
 \_\_VECTOR\_SIZE (*C macro*), 59  
 \_\_WEAK (*C macro*), 59  
 \_\_WMB (*C macro*), 94  
 \_\_disable\_FPU (*C macro*), 382  
 \_\_enable\_FPU (*C macro*), 382  
 \_\_get\_FCSR (*C macro*), 382  
 \_\_get\_FFLAGS (*C macro*), 382  
 \_\_get\_FRM (*C macro*), 382  
 \_\_has\_builtin (*C macro*), 59  
 \_\_set\_FCSR (*C macro*), 382  
 \_\_set\_FFLAGS (*C macro*), 382  
 \_\_set\_FRM (*C macro*), 382  
 \_fini (*C++ function*), 406  
 \_init (*C++ function*), 406  
 \_postmain\_fini (*C++ function*), 406  
 \_premain\_init (*C++ function*), 406

## A

API, 741

## C

CacheInfo\_Type (*C++ class*), 390  
 CAUSE\_BREAKPOINT (*C macro*), 78  
 CAUSE\_FAULT\_FETCH (*C macro*), 78  
 CAUSE\_FAULT\_LOAD (*C macro*), 78  
 CAUSE\_FAULT\_STORE (*C macro*), 78  
 CAUSE\_HYPERVISOR\_ECALL (*C macro*), 78  
 CAUSE\_ILLEGAL\_INSTRUCTION (*C macro*), 78  
 CAUSE\_MACHINE\_ECALL (*C macro*), 78  
 CAUSE\_MISALIGNED\_FETCH (*C macro*), 77  
 CAUSE\_MISALIGNED\_LOAD (*C macro*), 78  
 CAUSE\_MISALIGNED\_STORE (*C macro*), 78  
 CAUSE\_SUPERVISOR\_ECALL (*C macro*), 78  
 CAUSE\_USER\_ECALL (*C macro*), 78  
 CCM\_COMMAND\_COMMAND (*C macro*), 76  
 CCM\_DATA\_DATA (*C macro*), 76  
 CCM\_SUEN\_SUEN (*C macro*), 76  
 CCM\_SUEN\_SUEN\_Msk (*C macro*), 388  
 CCM\_SUEN\_SUEN\_Pos (*C macro*), 388  
 CLIC\_CLICCFG\_NLBIT\_Msk (*C macro*), 89  
 CLIC\_CLICCFG\_NLBIT\_Pos (*C macro*), 89  
 CLIC\_CLICINFO\_CTLBIT\_Msk (*C macro*), 89  
 CLIC\_CLICINFO\_CTLBIT\_Pos (*C macro*), 89  
 CLIC\_CLICINFO\_NUM\_Msk (*C macro*), 90  
 CLIC\_CLICINFO\_NUM\_Pos (*C macro*), 90  
 CLIC\_CLICINFO\_VER\_Msk (*C macro*), 90

CLIC\_CLICINFO\_VER\_Pos (*C macro*), 90  
 CLIC\_CTRL\_Type (*C++ class*), 91  
 CLIC\_INTATTR\_SHV\_Msk (*C macro*), 90  
 CLIC\_INTATTR\_SHV\_Pos (*C macro*), 90  
 CLIC\_INTATTR\_TRIG\_Msk (*C macro*), 90  
 CLIC\_INTATTR\_TRIG\_Pos (*C macro*), 90  
 CLIC\_INTIE\_IE\_Msk (*C macro*), 90  
 CLIC\_INTIE\_IE\_Pos (*C macro*), 90  
 CLIC\_INTIP\_IP\_Msk (*C macro*), 90  
 CLIC\_INTIP\_IP\_Pos (*C macro*), 90  
 CLICCFG\_Type (*C++ union*), 91  
 CLICCFG\_Type::\_reserved0 (*C++ member*), 91  
 CLICCFG\_Type::\_reserved1 (*C++ member*), 91  
 CLICCFG\_Type::\_reserved2 (*C++ member*), 91  
 CLICCFG\_Type::b (*C++ member*), 91  
 CLICCFG\_Type::nlbits (*C++ member*), 91  
 CLICCFG\_Type::w (*C++ member*), 91  
 CLICINFO\_Type (*C++ union*), 91  
 CLICINFO\_Type::\_reserved0 (*C++ member*), 91  
 CLICINFO\_Type::b (*C++ member*), 91  
 CLICINFO\_Type::intctlbits (*C++ member*), 91  
 CLICINFO\_Type::numint (*C++ member*), 91  
 CLICINFO\_Type::version (*C++ member*), 91  
 CLICINFO\_Type::w (*C++ member*), 91  
 core\_exception\_handler (*C++ function*), 408  
 cos\_factors\_128 (*C++ member*), 652, 653  
 cos\_factors\_2048 (*C++ member*), 652, 653  
 cos\_factors\_512 (*C++ member*), 652, 653  
 cos\_factors\_8192 (*C++ member*), 652, 653  
 cos\_factorsQ31\_128 (*C++ member*), 652, 653  
 cos\_factorsQ31\_2048 (*C++ member*), 652, 653  
 cos\_factorsQ31\_512 (*C++ member*), 652, 653  
 cos\_factorsQ31\_8192 (*C++ member*), 652, 653  
 CSR\_CCM\_FPIPE (*C macro*), 71  
 CSR\_CCM\_MBEGINADDR (*C macro*), 70  
 CSR\_CCM\_MCOMMAND (*C macro*), 70  
 CSR\_CCM\_MDATA (*C macro*), 70  
 CSR\_CCM\_SBEGINADDR (*C macro*), 71  
 CSR\_CCM\_SCOMMAND (*C macro*), 71  
 CSR\_CCM\_SDATA (*C macro*), 71  
 CSR\_CCM\_SUEN (*C macro*), 71  
 CSR\_CCM\_UBEGINADDR (*C macro*), 71  
 CSR\_CCM\_UCOMMAND (*C macro*), 71  
 CSR\_CCM\_UDATA (*C macro*), 71  
 CSR\_CYCLE (*C macro*), 63  
 CSR\_CYCLEH (*C macro*), 67  
 CSR\_DCSR (*C macro*), 65  
 CSR\_DPC (*C macro*), 65  
 CSR\_DSCRATCH (*C macro*), 65  
 CSR\_FCSR (*C macro*), 63  
 CSR\_FFLAGS (*C macro*), 63  
 CSR\_FRM (*C macro*), 63  
 CSR\_HPMCOUNTER10 (*C macro*), 63  
 CSR\_HPMCOUNTER10H (*C macro*), 67

CSR\_HPMCOUNTER11 (*C macro*), 63  
 CSR\_HPMCOUNTER11H (*C macro*), 67  
 CSR\_HPMCOUNTER12 (*C macro*), 63  
 CSR\_HPMCOUNTER12H (*C macro*), 67  
 CSR\_HPMCOUNTER13 (*C macro*), 63  
 CSR\_HPMCOUNTER13H (*C macro*), 67  
 CSR\_HPMCOUNTER14 (*C macro*), 63  
 CSR\_HPMCOUNTER14H (*C macro*), 67  
 CSR\_HPMCOUNTER15 (*C macro*), 63  
 CSR\_HPMCOUNTER15H (*C macro*), 67  
 CSR\_HPMCOUNTER16 (*C macro*), 63  
 CSR\_HPMCOUNTER16H (*C macro*), 67  
 CSR\_HPMCOUNTER17 (*C macro*), 63  
 CSR\_HPMCOUNTER17H (*C macro*), 67  
 CSR\_HPMCOUNTER18 (*C macro*), 63  
 CSR\_HPMCOUNTER18H (*C macro*), 67  
 CSR\_HPMCOUNTER19 (*C macro*), 63  
 CSR\_HPMCOUNTER19H (*C macro*), 67  
 CSR\_HPMCOUNTER20 (*C macro*), 63  
 CSR\_HPMCOUNTER20H (*C macro*), 67  
 CSR\_HPMCOUNTER21 (*C macro*), 63  
 CSR\_HPMCOUNTER21H (*C macro*), 67  
 CSR\_HPMCOUNTER22 (*C macro*), 63  
 CSR\_HPMCOUNTER22H (*C macro*), 67  
 CSR\_HPMCOUNTER23 (*C macro*), 63  
 CSR\_HPMCOUNTER23H (*C macro*), 67  
 CSR\_HPMCOUNTER24 (*C macro*), 63  
 CSR\_HPMCOUNTER24H (*C macro*), 68  
 CSR\_HPMCOUNTER25 (*C macro*), 63  
 CSR\_HPMCOUNTER25H (*C macro*), 68  
 CSR\_HPMCOUNTER26 (*C macro*), 63  
 CSR\_HPMCOUNTER26H (*C macro*), 68  
 CSR\_HPMCOUNTER27 (*C macro*), 64  
 CSR\_HPMCOUNTER27H (*C macro*), 68  
 CSR\_HPMCOUNTER28 (*C macro*), 64  
 CSR\_HPMCOUNTER28H (*C macro*), 68  
 CSR\_HPMCOUNTER29 (*C macro*), 64  
 CSR\_HPMCOUNTER29H (*C macro*), 68  
 CSR\_HPMCOUNTER3 (*C macro*), 63  
 CSR\_HPMCOUNTER30 (*C macro*), 64  
 CSR\_HPMCOUNTER30H (*C macro*), 68  
 CSR\_HPMCOUNTER31 (*C macro*), 64  
 CSR\_HPMCOUNTER31H (*C macro*), 68  
 CSR\_HPMCOUNTER3H (*C macro*), 67  
 CSR\_HPMCOUNTER4 (*C macro*), 63  
 CSR\_HPMCOUNTER4H (*C macro*), 67  
 CSR\_HPMCOUNTER5 (*C macro*), 63  
 CSR\_HPMCOUNTER5H (*C macro*), 67  
 CSR\_HPMCOUNTER6 (*C macro*), 63  
 CSR\_HPMCOUNTER6H (*C macro*), 67  
 CSR\_HPMCOUNTER7 (*C macro*), 63  
 CSR\_HPMCOUNTER7H (*C macro*), 67  
 CSR\_HPMCOUNTER8 (*C macro*), 63  
 CSR\_HPMCOUNTER8H (*C macro*), 67  
 CSR\_HPMCOUNTER9 (*C macro*), 63  
 CSR\_HPMCOUNTER9H (*C macro*), 67  
 CSR\_INSTRET (*C macro*), 63  
 CSR\_INSTRETH (*C macro*), 67  
 CSR\_JALMNXTI (*C macro*), 70  
 CSR\_JALSNXTI (*C macro*), 69  
 CSR\_LSTEPFORC (*C macro*), 70  
 CSR\_MARCHID (*C macro*), 67  
 CSR\_MBADADDR (*C macro*), 64  
 CSR\_MCACHE\_CTL (*C macro*), 70  
 CSR\_MCACHE\_CTL\_DE (*C macro*), 72  
 CSR\_MCACHE\_CTL\_IE (*C macro*), 72  
 CSR\_MCACHECTL\_Type (*C++ union*), 83  
 CSR\_MCACHECTL\_Type::\_reserved0 (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::\_reserved1 (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::b (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::d (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::dc\_ecc\_en (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::dc\_ecc\_excp\_en (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::dc\_en (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::dc\_rwdecc (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::dc\_rwtecc (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::ic\_ecc\_en (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::ic\_ecc\_excp\_en (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::ic\_en (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::ic\_rwdecc (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::ic\_rwtecc (*C++ member*), 84  
 CSR\_MCACHECTL\_Type::ic\_scpd\_mod (*C++ member*), 84  
 CSR\_MCAUSE (*C macro*), 64  
 CSR\_MCAUSE\_Type (*C++ union*), 81  
 CSR\_MCAUSE\_Type::\_reserved0 (*C++ member*), 81  
 CSR\_MCAUSE\_Type::\_reserved1 (*C++ member*), 81  
 CSR\_MCAUSE\_Type::b (*C++ member*), 82  
 CSR\_MCAUSE\_Type::d (*C++ member*), 82  
 CSR\_MCAUSE\_Type::exccode (*C++ member*), 81  
 CSR\_MCAUSE\_Type::interrupt (*C++ member*), 82  
 CSR\_MCAUSE\_Type::minhv (*C++ member*), 82  
 CSR\_MCAUSE\_Type::mpie (*C++ member*), 81  
 CSR\_MCAUSE\_Type::mpil (*C++ member*), 81  
 CSR\_MCAUSE\_Type::mpp (*C++ member*), 82

CSR\_MCFG\_INFO (*C macro*), 70  
 CSR\_MCFGINFO\_Type (*C++ union*), 86  
 CSR\_MCFGINFO\_Type::\_reserved0 (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::b (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::clic (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::d (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::dcache (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::dlm (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::ecc (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::fio (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::icache (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::ilm (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::nice (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::plic (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::ppi (*C++ member*), 86  
 CSR\_MCFGINFO\_Type::tee (*C++ member*), 86  
 CSR\_MCLICBASE (*C macro*), 69  
 CSR\_MCOUNTEREN (*C macro*), 64  
 CSR\_MCOUNTINHIBIT (*C macro*), 69  
 CSR\_MCOUNTINHIBIT\_Type (*C++ union*), 82  
 CSR\_MCOUNTINHIBIT\_Type::\_reserved0 (*C++ member*), 82  
 CSR\_MCOUNTINHIBIT\_Type::\_reserved1 (*C++ member*), 82  
 CSR\_MCOUNTINHIBIT\_Type::b (*C++ member*), 82  
 CSR\_MCOUNTINHIBIT\_Type::cy (*C++ member*), 82  
 CSR\_MCOUNTINHIBIT\_Type::d (*C++ member*), 82  
 CSR\_MCOUNTINHIBIT\_Type::ir (*C++ member*), 82  
 CSR\_MCYCLE (*C macro*), 65  
 CSR\_MCYCLEH (*C macro*), 68  
 CSR\_MDCAUSE (*C macro*), 70  
 CSR\_MDCAUSE\_Type (*C++ union*), 83  
 CSR\_MDCAUSE\_Type::\_reserved0 (*C++ member*), 83  
 CSR\_MDCAUSE\_Type::b (*C++ member*), 83  
 CSR\_MDCAUSE\_Type::d (*C++ member*), 83  
 CSR\_MDCAUSE\_Type::mdcause (*C++ member*), 83  
 CSR\_MDCFG\_INFO (*C macro*), 70  
 CSR\_MDCFGINFO\_Type (*C++ union*), 87  
 CSR\_MDCFGINFO\_Type::\_reserved0 (*C++ member*), 87  
 CSR\_MDCFGINFO\_Type::\_reserved1 (*C++ member*), 88  
 CSR\_MDCFGINFO\_Type::b (*C++ member*), 88  
 CSR\_MDCFGINFO\_Type::cache\_ecc (*C++ member*), 87  
 CSR\_MDCFGINFO\_Type::d (*C++ member*), 88  
 CSR\_MDCFGINFO\_Type::lm\_ecc (*C++ member*), 87  
 CSR\_MDCFGINFO\_Type::lm\_size (*C++ member*), 87  
 CSR\_MDCFGINFO\_Type::lm\_xonly (*C++ member*), 87  
 CSR\_MDCFGINFO\_Type::lsize (*C++ member*), 87  
 CSR\_MDCFGINFO\_Type::set (*C++ member*), 87  
 CSR\_MDCFGINFO\_Type::way (*C++ member*), 87  
 CSR\_MDLM\_CTL (*C macro*), 70  
 CSR\_MDLMCTL\_Type (*C++ union*), 85  
 CSR\_MDLMCTL\_Type::\_reserved0 (*C++ member*), 86  
 CSR\_MDLMCTL\_Type::b (*C++ member*), 86  
 CSR\_MDLMCTL\_Type::d (*C++ member*), 86  
 CSR\_MDLMCTL\_Type::dlm\_bpa (*C++ member*), 86  
 CSR\_MDLMCTL\_Type::dlm\_ecc\_en (*C++ member*), 85  
 CSR\_MDLMCTL\_Type::dlm\_ecc\_excp\_en (*C++ member*), 85  
 CSR\_MDLMCTL\_Type::dlm\_en (*C++ member*), 85  
 CSR\_MDLMCTL\_Type::dlm\_rwecc (*C++ member*), 86  
 CSR\_MECC\_CODE (*C macro*), 70  
 CSR\_MECC\_LOCK (*C macro*), 70  
 CSR\_MECCCODE\_Type (*C++ union*), 89  
 CSR\_MECCCODE\_Type::\_reserved0 (*C++ member*), 89  
 CSR\_MECCCODE\_Type::\_reserved1 (*C++ member*), 89  
 CSR\_MECCCODE\_Type::\_reserved2 (*C++ member*), 89  
 CSR\_MECCCODE\_Type::b (*C++ member*), 89  
 CSR\_MECCCODE\_Type::code (*C++ member*), 89  
 CSR\_MECCCODE\_Type::d (*C++ member*), 89  
 CSR\_MECCCODE\_Type::ramid (*C++ member*), 89  
 CSR\_MECCCODE\_Type::sramid (*C++ member*), 89  
 CSR\_MECCLOCK\_Type (*C++ union*), 88  
 CSR\_MECCLOCK\_Type::\_reserved0 (*C++ member*), 89  
 CSR\_MECCLOCK\_Type::b (*C++ member*), 89  
 CSR\_MECCLOCK\_Type::d (*C++ member*), 89  
 CSR\_MECCLOCK\_Type::ecc\_lock (*C++ member*), 89  
 CSR\_MEDELEG (*C macro*), 64  
 CSR\_MEPC (*C macro*), 64  
 CSR\_MFIOCFG\_INFO (*C macro*), 70  
 CSR\_MFIOCFGINFO\_Type (*C++ union*), 88  
 CSR\_MFIOCFGINFO\_Type::\_reserved0 (*C++ member*), 88  
 CSR\_MFIOCFGINFO\_Type::\_reserved1 (*C++ member*), 88  
 CSR\_MFIOCFGINFO\_Type::b (*C++ member*), 88  
 CSR\_MFIOCFGINFO\_Type::d (*C++ member*), 88  
 CSR\_MFIOCFGINFO\_Type::fio\_bpa (*C++ member*), 88  
 CSR\_MFIOCFGINFO\_Type::fio\_size (*C++ member*), 88

CSR\_MFP16MODE (*C macro*), 70  
 CSR\_MHARTID (*C macro*), 67  
 CSR\_MHPMCOUNTER10 (*C macro*), 65  
 CSR\_MHPMCOUNTER10H (*C macro*), 68  
 CSR\_MHPMCOUNTER11 (*C macro*), 65  
 CSR\_MHPMCOUNTER11H (*C macro*), 68  
 CSR\_MHPMCOUNTER12 (*C macro*), 65  
 CSR\_MHPMCOUNTER12H (*C macro*), 68  
 CSR\_MHPMCOUNTER13 (*C macro*), 65  
 CSR\_MHPMCOUNTER13H (*C macro*), 68  
 CSR\_MHPMCOUNTER14 (*C macro*), 65  
 CSR\_MHPMCOUNTER14H (*C macro*), 68  
 CSR\_MHPMCOUNTER15 (*C macro*), 65  
 CSR\_MHPMCOUNTER15H (*C macro*), 68  
 CSR\_MHPMCOUNTER16 (*C macro*), 65  
 CSR\_MHPMCOUNTER16H (*C macro*), 68  
 CSR\_MHPMCOUNTER17 (*C macro*), 65  
 CSR\_MHPMCOUNTER17H (*C macro*), 68  
 CSR\_MHPMCOUNTER18 (*C macro*), 65  
 CSR\_MHPMCOUNTER18H (*C macro*), 68  
 CSR\_MHPMCOUNTER19 (*C macro*), 66  
 CSR\_MHPMCOUNTER19H (*C macro*), 68  
 CSR\_MHPMCOUNTER20 (*C macro*), 66  
 CSR\_MHPMCOUNTER20H (*C macro*), 68  
 CSR\_MHPMCOUNTER21 (*C macro*), 66  
 CSR\_MHPMCOUNTER21H (*C macro*), 68  
 CSR\_MHPMCOUNTER22 (*C macro*), 66  
 CSR\_MHPMCOUNTER22H (*C macro*), 68  
 CSR\_MHPMCOUNTER23 (*C macro*), 66  
 CSR\_MHPMCOUNTER23H (*C macro*), 68  
 CSR\_MHPMCOUNTER24 (*C macro*), 66  
 CSR\_MHPMCOUNTER24H (*C macro*), 68  
 CSR\_MHPMCOUNTER25 (*C macro*), 66  
 CSR\_MHPMCOUNTER25H (*C macro*), 68  
 CSR\_MHPMCOUNTER26 (*C macro*), 66  
 CSR\_MHPMCOUNTER26H (*C macro*), 68  
 CSR\_MHPMCOUNTER27 (*C macro*), 66  
 CSR\_MHPMCOUNTER27H (*C macro*), 68  
 CSR\_MHPMCOUNTER28 (*C macro*), 66  
 CSR\_MHPMCOUNTER28H (*C macro*), 68  
 CSR\_MHPMCOUNTER29 (*C macro*), 66  
 CSR\_MHPMCOUNTER29H (*C macro*), 69  
 CSR\_MHPMCOUNTER3 (*C macro*), 65  
 CSR\_MHPMCOUNTER30 (*C macro*), 66  
 CSR\_MHPMCOUNTER30H (*C macro*), 69  
 CSR\_MHPMCOUNTER31 (*C macro*), 66  
 CSR\_MHPMCOUNTER31H (*C macro*), 69  
 CSR\_MHPMCOUNTER3H (*C macro*), 68  
 CSR\_MHPMCOUNTER4 (*C macro*), 65  
 CSR\_MHPMCOUNTER4H (*C macro*), 68  
 CSR\_MHPMCOUNTER5 (*C macro*), 65  
 CSR\_MHPMCOUNTER5H (*C macro*), 68  
 CSR\_MHPMCOUNTER6 (*C macro*), 65  
 CSR\_MHPMCOUNTER6H (*C macro*), 68  
 CSR\_MHPMCOUNTER7 (*C macro*), 65  
 CSR\_MHPMCOUNTER7H (*C macro*), 68  
 CSR\_MHPMCOUNTER8 (*C macro*), 65  
 CSR\_MHPMCOUNTER8H (*C macro*), 68  
 CSR\_MHPMCOUNTER9 (*C macro*), 65  
 CSR\_MHPMCOUNTER9H (*C macro*), 68  
 CSR\_MHPMEVENT10 (*C macro*), 66  
 CSR\_MHPMEVENT11 (*C macro*), 66  
 CSR\_MHPMEVENT12 (*C macro*), 66  
 CSR\_MHPMEVENT13 (*C macro*), 66  
 CSR\_MHPMEVENT14 (*C macro*), 66  
 CSR\_MHPMEVENT15 (*C macro*), 66  
 CSR\_MHPMEVENT16 (*C macro*), 66  
 CSR\_MHPMEVENT17 (*C macro*), 66  
 CSR\_MHPMEVENT18 (*C macro*), 66  
 CSR\_MHPMEVENT19 (*C macro*), 66  
 CSR\_MHPMEVENT20 (*C macro*), 66  
 CSR\_MHPMEVENT21 (*C macro*), 66  
 CSR\_MHPMEVENT22 (*C macro*), 66  
 CSR\_MHPMEVENT23 (*C macro*), 66  
 CSR\_MHPMEVENT24 (*C macro*), 67  
 CSR\_MHPMEVENT25 (*C macro*), 67  
 CSR\_MHPMEVENT26 (*C macro*), 67  
 CSR\_MHPMEVENT27 (*C macro*), 67  
 CSR\_MHPMEVENT28 (*C macro*), 67  
 CSR\_MHPMEVENT29 (*C macro*), 67  
 CSR\_MHPMEVENT3 (*C macro*), 66  
 CSR\_MHPMEVENT30 (*C macro*), 67  
 CSR\_MHPMEVENT31 (*C macro*), 67  
 CSR\_MHPMEVENT4 (*C macro*), 66  
 CSR\_MHPMEVENT5 (*C macro*), 66  
 CSR\_MHPMEVENT6 (*C macro*), 66  
 CSR\_MHPMEVENT7 (*C macro*), 66  
 CSR\_MHPMEVENT8 (*C macro*), 66  
 CSR\_MHPMEVENT9 (*C macro*), 66  
 CSR\_MICFG\_INFO (*C macro*), 70  
 CSR\_MICFGINFO\_Type (*C++ union*), 87  
 CSR\_MICFGINFO\_Type::\_reserved0 (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::\_reserved1 (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::b (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::cache\_ecc (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::d (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::lm\_ecc (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::lm\_size (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::lm\_xonly (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::lsize (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::set (*C++ member*), 87  
 CSR\_MICFGINFO\_Type::way (*C++ member*), 87



CSR\_MIDELEG (*C macro*), 64  
 CSR\_MIE (*C macro*), 64  
 CSR\_MILM\_CTL (*C macro*), 69  
 CSR\_MILMCTL\_Type (*C++ union*), 85  
 CSR\_MILMCTL\_Type::\_reserved0 (*C++ member*), 85  
 CSR\_MILMCTL\_Type::b (*C++ member*), 85  
 CSR\_MILMCTL\_Type::d (*C++ member*), 85  
 CSR\_MILMCTL\_Type::ilm\_bpa (*C++ member*), 85  
 CSR\_MILMCTL\_Type::ilm\_ecc\_en (*C++ member*), 85  
 CSR\_MILMCTL\_Type::ilm\_ecc\_excp\_en (*C++ member*), 85  
 CSR\_MILMCTL\_Type::ilm\_en (*C++ member*), 85  
 CSR\_MILMCTL\_Type::ilm\_rwecc (*C++ member*), 85  
 CSR\_MIMPID (*C macro*), 67  
 CSR\_MINSTRET (*C macro*), 65  
 CSR\_MINSTRETH (*C macro*), 68  
 CSR\_MINTSTATUS (*C macro*), 69  
 CSR\_MIP (*C macro*), 64  
 CSR\_MISA (*C macro*), 64  
 CSR\_MISA\_Type (*C++ union*), 79  
 CSR\_MISA\_Type::\_reserved1 (*C++ member*), 79  
 CSR\_MISA\_Type::\_reserved2 (*C++ member*), 79  
 CSR\_MISA\_Type::\_reserved4 (*C++ member*), 80  
 CSR\_MISA\_Type::\_reserved5 (*C++ member*), 80  
 CSR\_MISA\_Type::\_resreved3 (*C++ member*), 79  
 CSR\_MISA\_Type::a (*C++ member*), 79  
 CSR\_MISA\_Type::b (*C++ member*), 79, 80  
 CSR\_MISA\_Type::c (*C++ member*), 79  
 CSR\_MISA\_Type::d (*C++ member*), 79  
 CSR\_MISA\_Type::e (*C++ member*), 79  
 CSR\_MISA\_Type::f (*C++ member*), 79  
 CSR\_MISA\_Type::g (*C++ member*), 79  
 CSR\_MISA\_Type::h (*C++ member*), 79  
 CSR\_MISA\_Type::i (*C++ member*), 79  
 CSR\_MISA\_Type::j (*C++ member*), 79  
 CSR\_MISA\_Type::l (*C++ member*), 79  
 CSR\_MISA\_Type::m (*C++ member*), 79  
 CSR\_MISA\_Type::mxl (*C++ member*), 80  
 CSR\_MISA\_Type::n (*C++ member*), 79  
 CSR\_MISA\_Type::p (*C++ member*), 79  
 CSR\_MISA\_Type::q (*C++ member*), 79  
 CSR\_MISA\_Type::s (*C++ member*), 80  
 CSR\_MISA\_Type::t (*C++ member*), 80  
 CSR\_MISA\_Type::u (*C++ member*), 80  
 CSR\_MISA\_Type::v (*C++ member*), 80  
 CSR\_MISA\_Type::x (*C++ member*), 80  
 CSR\_MMISC\_CTL (*C macro*), 70  
 CSR\_MMISCCTL\_Type (*C++ type*), 79  
 CSR\_MMISCCTRL\_Type (*C++ union*), 83  
 CSR\_MMISCCTRL\_Type::\_reserved0 (*C++ member*), 83  
 CSR\_MMISCCTRL\_Type::\_reserved1 (*C++ member*), 83  
 CSR\_MMISCCTRL\_Type::\_reserved2 (*C++ member*), 83  
 CSR\_MMISCCTRL\_Type::\_reserved3 (*C++ member*), 83  
 CSR\_MMISCCTRL\_Type::b (*C++ member*), 83  
 CSR\_MMISCCTRL\_Type::bpu (*C++ member*), 83  
 CSR\_MMISCCTRL\_Type::d (*C++ member*), 83  
 CSR\_MMISCCTRL\_Type::misalign (*C++ member*), 83  
 CSR\_MMISCCTRL\_Type::nmi\_cause (*C++ member*), 83  
 CSR\_MNVEC (*C macro*), 70  
 CSR\_MNXTI (*C macro*), 69  
 CSR\_MPPICFG\_INFO (*C macro*), 70  
 CSR\_MPPICFGINFO\_Type (*C++ union*), 88  
 CSR\_MPPICFGINFO\_Type::\_reserved0 (*C++ member*), 88  
 CSR\_MPPICFGINFO\_Type::\_reserved1 (*C++ member*), 88  
 CSR\_MPPICFGINFO\_Type::b (*C++ member*), 88  
 CSR\_MPPICFGINFO\_Type::d (*C++ member*), 88  
 CSR\_MPPICFGINFO\_Type::ppi\_bpa (*C++ member*), 88  
 CSR\_MPPICFGINFO\_Type::ppi\_size (*C++ member*), 88  
 CSR\_MSAVECAUSE1 (*C macro*), 70  
 CSR\_MSAVECAUSE2 (*C macro*), 70  
 CSR\_MSAVEDCAUSE1 (*C macro*), 70  
 CSR\_MSAVEDCAUSE2 (*C macro*), 70  
 CSR\_MSAVEEPC1 (*C macro*), 70  
 CSR\_MSAVEEPC2 (*C macro*), 70  
 CSR\_MSAVESTATUS (*C macro*), 70  
 CSR\_MSAVESTATUS\_Type (*C++ union*), 84  
 CSR\_MSAVESTATUS\_Type::\_reserved0 (*C++ member*), 84  
 CSR\_MSAVESTATUS\_Type::\_reserved1 (*C++ member*), 85  
 CSR\_MSAVESTATUS\_Type::\_reserved2 (*C++ member*), 85  
 CSR\_MSAVESTATUS\_Type::b (*C++ member*), 85  
 CSR\_MSAVESTATUS\_Type::mpiel (*C++ member*), 84  
 CSR\_MSAVESTATUS\_Type::mpie2 (*C++ member*), 84  
 CSR\_MSAVESTATUS\_Type::mpp1 (*C++ member*), 84  
 CSR\_MSAVESTATUS\_Type::mpp2 (*C++ member*), 85  
 CSR\_MSAVESTATUS\_Type::ptyp1 (*C++ member*), 84  
 CSR\_MSAVESTATUS\_Type::ptyp2 (*C++ member*), 85

CSR\_MSAVESTATUS\_Type::w (C++ member), 85  
 CSR\_MSCOUNTEREN (C macro), 66  
 CSR\_MSCRATCH (C macro), 64  
 CSR\_MSCRATCHCSW (C macro), 69  
 CSR\_MSCRATCHCSWL (C macro), 69  
 CSR\_MSMPCFG\_INFO (C macro), 70  
 CSR\_MSTATUS (C macro), 64  
 CSR\_MSTATUS\_Type (C++ union), 80  
 CSR\_MSTATUS\_Type::\_reserved0 (C++ member), 80  
 CSR\_MSTATUS\_Type::\_reserved1 (C++ member), 80  
 CSR\_MSTATUS\_Type::\_reserved2 (C++ member), 80  
 CSR\_MSTATUS\_Type::\_reserved3 (C++ member), 80  
 CSR\_MSTATUS\_Type::\_reserved4 (C++ member), 80  
 CSR\_MSTATUS\_Type::\_reserved6 (C++ member), 81  
 CSR\_MSTATUS\_Type::b (C++ member), 81  
 CSR\_MSTATUS\_Type::d (C++ member), 81  
 CSR\_MSTATUS\_Type::fs (C++ member), 81  
 CSR\_MSTATUS\_Type::mie (C++ member), 80  
 CSR\_MSTATUS\_Type::mpie (C++ member), 80  
 CSR\_MSTATUS\_Type::mpp (C++ member), 80  
 CSR\_MSTATUS\_Type::mprv (C++ member), 81  
 CSR\_MSTATUS\_Type::sd (C++ member), 81  
 CSR\_MSTATUS\_Type::sie (C++ member), 80  
 CSR\_MSTATUS\_Type::spie (C++ member), 80  
 CSR\_MSTATUS\_Type::sum (C++ member), 81  
 CSR\_MSTATUS\_Type::xs (C++ member), 81  
 CSR\_MSUBM (C macro), 70  
 CSR\_MSUBM\_Type (C++ union), 82  
 CSR\_MSUBM\_Type::\_reserved0 (C++ member), 82  
 CSR\_MSUBM\_Type::\_reserved1 (C++ member), 82  
 CSR\_MSUBM\_Type::b (C++ member), 82  
 CSR\_MSUBM\_Type::d (C++ member), 83  
 CSR\_MSUBM\_Type::ptyp (C++ member), 82  
 CSR\_MSUBM\_Type::typ (C++ member), 82  
 CSR\_MTLB\_CTL (C macro), 70  
 CSR\_MTLBCFG\_INFO (C macro), 70  
 CSR\_MTVAL (C macro), 64  
 CSR\_MTVEC (C macro), 64  
 CSR\_MTVEC\_Type (C++ union), 81  
 CSR\_MTVEC\_Type::addr (C++ member), 81  
 CSR\_MTVEC\_Type::b (C++ member), 81  
 CSR\_MTVEC\_Type::d (C++ member), 81  
 CSR\_MTVEC\_Type::mode (C++ member), 81  
 CSR\_MTVT (C macro), 69  
 CSR\_MTVT2 (C macro), 70  
 CSR\_MUCOUNTEREN (C macro), 66  
 CSR\_MVENDORID (C macro), 67  
 CSR\_PMPADDR0 (C macro), 64  
 CSR\_PMPADDR1 (C macro), 64  
 CSR\_PMPADDR10 (C macro), 65  
 CSR\_PMPADDR11 (C macro), 65  
 CSR\_PMPADDR12 (C macro), 65  
 CSR\_PMPADDR13 (C macro), 65  
 CSR\_PMPADDR14 (C macro), 65  
 CSR\_PMPADDR15 (C macro), 65  
 CSR\_PMPADDR2 (C macro), 64  
 CSR\_PMPADDR3 (C macro), 64  
 CSR\_PMPADDR4 (C macro), 64  
 CSR\_PMPADDR5 (C macro), 65  
 CSR\_PMPADDR6 (C macro), 65  
 CSR\_PMPADDR7 (C macro), 65  
 CSR\_PMPADDR8 (C macro), 65  
 CSR\_PMPADDR9 (C macro), 65  
 CSR\_PMPCFG0 (C macro), 64  
 CSR\_PMPCFG1 (C macro), 64  
 CSR\_PMPCFG2 (C macro), 64  
 CSR\_PMPCFG3 (C macro), 64  
 CSR\_PUSHMCAUSE (C macro), 70  
 CSR\_PUSHMEPC (C macro), 70  
 CSR\_PUSHMSUBM (C macro), 70  
 CSR\_PUSHSCAUSE (C macro), 69  
 CSR\_PUSHSEPC (C macro), 69  
 CSR\_SBADADDR (C macro), 64  
 CSR\_SCAUSE (C macro), 64  
 CSR\_SEPC (C macro), 64  
 CSR\_SIE (C macro), 64  
 CSR\_SIP (C macro), 64  
 CSR\_SLEEPVALUE (C macro), 70  
 CSR\_SPMPADDR0 (C macro), 69  
 CSR\_SPMPADDR1 (C macro), 69  
 CSR\_SPMPADDR10 (C macro), 69  
 CSR\_SPMPADDR11 (C macro), 69  
 CSR\_SPMPADDR12 (C macro), 69  
 CSR\_SPMPADDR13 (C macro), 69  
 CSR\_SPMPADDR14 (C macro), 69  
 CSR\_SPMPADDR15 (C macro), 69  
 CSR\_SPMPADDR2 (C macro), 69  
 CSR\_SPMPADDR3 (C macro), 69  
 CSR\_SPMPADDR4 (C macro), 69  
 CSR\_SPMPADDR5 (C macro), 69  
 CSR\_SPMPADDR6 (C macro), 69  
 CSR\_SPMPADDR7 (C macro), 69  
 CSR\_SPMPADDR8 (C macro), 69  
 CSR\_SPMPADDR9 (C macro), 69  
 CSR\_SPMPCFG0 (C macro), 69  
 CSR\_SPMPCFG1 (C macro), 69  
 CSR\_SPMPCFG2 (C macro), 69  
 CSR\_SPMPCFG3 (C macro), 69  
 CSR\_SPTBR (C macro), 64  
 CSR\_SSCRATCH (C macro), 64

CSR\_SSTATUS (*C macro*), 64  
CSR\_STVEC (*C macro*), 64  
CSR\_STVT2 (*C macro*), 69  
CSR\_TDATA1 (*C macro*), 65  
CSR\_TDATA2 (*C macro*), 65  
CSR\_TDATA3 (*C macro*), 65  
CSR\_TIME (*C macro*), 63  
CSR\_TIMEH (*C macro*), 67  
CSR\_TSELECT (*C macro*), 65  
CSR\_TXEVT (*C macro*), 70  
CSR\_UCODE (*C macro*), 69  
CSR\_USTATUS (*C macro*), 63  
CSR\_WFE (*C macro*), 70

## D

DCAUSE\_FAULT\_FETCH\_INST (*C macro*), 78  
DCAUSE\_FAULT\_FETCH\_PMP (*C macro*), 78  
DCAUSE\_FAULT\_LOAD\_INST (*C macro*), 78  
DCAUSE\_FAULT\_LOAD\_NICE (*C macro*), 78  
DCAUSE\_FAULT\_LOAD\_PMP (*C macro*), 78  
DCAUSE\_FAULT\_STORE\_INST (*C macro*), 78  
DCAUSE\_FAULT\_STORE\_PMP (*C macro*), 78  
DCSR\_CAUSE (*C macro*), 72  
DCSR\_CAUSE\_DEBUGINT (*C macro*), 72  
DCSR\_CAUSE\_HALT (*C macro*), 72  
DCSR\_CAUSE\_HWBP (*C macro*), 72  
DCSR\_CAUSE\_NONE (*C macro*), 72  
DCSR\_CAUSE\_STEP (*C macro*), 72  
DCSR\_CAUSE\_SWBP (*C macro*), 72  
DCSR\_DEBUGINT (*C macro*), 72  
DCSR\_EBREAKH (*C macro*), 72  
DCSR\_EBREAKM (*C macro*), 72  
DCSR\_EBREAKS (*C macro*), 72  
DCSR\_EBREAKU (*C macro*), 72  
DCSR\_FULLRESET (*C macro*), 72  
DCSR\_HALT (*C macro*), 72  
DCSR\_NDRESET (*C macro*), 72  
DCSR\_PRV (*C macro*), 72  
DCSR\_STEP (*C macro*), 72  
DCSR\_STOPCYCLE (*C macro*), 72  
DCSR\_STOPTIME (*C macro*), 72  
DCSR\_XDEBUGVER (*C macro*), 72  
depthwise\_conv\_s8\_generic (*C++ function*), 687, 703  
depthwise\_conv\_s8\_mult\_4 (*C++ function*), 687, 703  
depthwise\_conv\_u8\_generic (*C++ function*), 688, 705  
depthwise\_conv\_u8\_mult\_4 (*C++ function*), 687, 705  
DSP, 741

## E

ECLIC (*C macro*), 90

ECLIC\_BASE (*C macro*), 90  
ECLIC\_ClearPendingIRQ (*C macro*), 370, 371  
ECLIC\_DisableIRQ (*C macro*), 370, 371  
ECLIC\_EnableIRQ (*C macro*), 370, 371  
ECLIC\_GetCfgNlbits (*C macro*), 370, 371  
ECLIC\_GetCtrlIRQ (*C macro*), 370, 371  
ECLIC\_GetEnableIRQ (*C macro*), 370, 371  
ECLIC\_GetInfoCtlbits (*C macro*), 370, 371  
ECLIC\_GetInfoNum (*C macro*), 370, 371  
ECLIC\_GetInfoVer (*C macro*), 370, 371  
ECLIC\_GetLevelIRQ (*C macro*), 370, 371  
ECLIC\_GetMth (*C macro*), 370, 371  
ECLIC\_GetPendingIRQ (*C macro*), 370, 371  
ECLIC\_GetPriorityIRQ (*C macro*), 370, 371  
ECLIC\_GetShvIRQ (*C macro*), 370, 371  
ECLIC\_GetTrigIRQ (*C macro*), 370, 371  
ECLIC\_GetVector (*C macro*), 370, 371  
ECLIC\_Init (*C++ function*), 408  
ECLIC\_MAX\_NLBITS (*C macro*), 90  
ECLIC\_MODE\_MTVEC\_Msk (*C macro*), 90  
ECLIC\_NON\_VECTOR\_INTERRUPT (*C macro*), 90  
ECLIC\_Register\_IRQ (*C++ function*), 408  
ECLIC\_SetCfgNlbits (*C macro*), 370, 371  
ECLIC\_SetCtrlIRQ (*C macro*), 370, 371  
ECLIC\_SetLevelIRQ (*C macro*), 370, 371  
ECLIC\_SetMth (*C macro*), 370, 371  
ECLIC\_SetPendingIRQ (*C macro*), 370, 371  
ECLIC\_SetPriorityIRQ (*C macro*), 370, 371  
ECLIC\_SetShvIRQ (*C macro*), 370, 371  
ECLIC\_SetTrigIRQ (*C macro*), 370, 371  
ECLIC\_SetVector (*C macro*), 370, 371  
ECLIC\_VECTOR\_INTERRUPT (*C macro*), 90  
EXC\_HANDLER (*C++ type*), 407  
Exception\_DumpFrame (*C++ function*), 407  
Exception\_Get\_EXC (*C++ function*), 408  
Exception\_Init (*C++ function*), 407  
Exception\_Register\_EXC (*C++ function*), 407

## F

FFLAGS\_AE\_DZ (*C macro*), 77  
FFLAGS\_AE\_NV (*C macro*), 77  
FFLAGS\_AE\_NX (*C macro*), 77  
FFLAGS\_AE\_OF (*C macro*), 77  
FFLAGS\_AE\_UF (*C macro*), 77  
FREG (*C macro*), 77  
FRM\_RNDMODE\_DYN (*C macro*), 76  
FRM\_RNDMODE\_RDN (*C macro*), 76  
FRM\_RNDMODE\_RMM (*C macro*), 76  
FRM\_RNDMODE\_RNE (*C macro*), 76  
FRM\_RNDMODE\_RTZ (*C macro*), 76  
FRM\_RNDMODE\_RUP (*C macro*), 76

## I

IRQ\_COP (*C macro*), 76



IRQ\_H\_EXT (*C macro*), 76  
 IRQ\_H\_SOFT (*C macro*), 76  
 IRQ\_H\_TIMER (*C macro*), 76  
 IRQ\_HOST (*C macro*), 76  
 IRQ\_M\_EXT (*C macro*), 76  
 IRQ\_M\_SOFT (*C macro*), 76  
 IRQ\_M\_TIMER (*C macro*), 76  
 IRQ\_S\_EXT (*C macro*), 76  
 IRQ\_S\_SOFT (*C macro*), 76  
 IRQ\_S\_TIMER (*C macro*), 76  
 ISR, 741

## M

MAX\_SYSTEM\_EXCEPTION\_NUM (*C macro*), 407  
 MCACHE\_CTL\_DC\_ECC\_EN (*C macro*), 74  
 MCACHE\_CTL\_DC\_ECC\_EXCP\_EN (*C macro*), 75  
 MCACHE\_CTL\_DC\_EN (*C macro*), 74  
 MCACHE\_CTL\_DC\_RWDECC (*C macro*), 75  
 MCACHE\_CTL\_DC\_RWTECC (*C macro*), 75  
 MCACHE\_CTL\_IC\_ECC\_EN (*C macro*), 74  
 MCACHE\_CTL\_IC\_ECC\_EXCP\_EN (*C macro*), 74  
 MCACHE\_CTL\_IC\_EN (*C macro*), 74  
 MCACHE\_CTL\_IC\_RWDECC (*C macro*), 74  
 MCACHE\_CTL\_IC\_RWTECC (*C macro*), 74  
 MCACHE\_CTL\_IC\_SCPD\_MOD (*C macro*), 74  
 MCFG\_INFO\_CLIC (*C macro*), 75  
 MCFG\_INFO\_DCACHE (*C macro*), 75  
 MCFG\_INFO\_DLM (*C macro*), 75  
 MCFG\_INFO\_ECC (*C macro*), 75  
 MCFG\_INFO\_FIO (*C macro*), 75  
 MCFG\_INFO\_ICACHE (*C macro*), 75  
 MCFG\_INFO\_ILM (*C macro*), 75  
 MCFG\_INFO\_NICE (*C macro*), 75  
 MCFG\_INFO\_PLIC (*C macro*), 75  
 MCFG\_INFO\_PPI (*C macro*), 75  
 MCFG\_INFO\_TEE (*C macro*), 75  
 MCONTROL\_ACTION (*C macro*), 73  
 MCONTROL\_ACTION\_DEBUG\_EXCEPTION (*C macro*), 73  
 MCONTROL\_ACTION\_DEBUG\_MODE (*C macro*), 73  
 MCONTROL\_ACTION\_TRACE\_EMIT (*C macro*), 73  
 MCONTROL\_ACTION\_TRACE\_START (*C macro*), 73  
 MCONTROL\_ACTION\_TRACE\_STOP (*C macro*), 73  
 MCONTROL\_CHAIN (*C macro*), 73  
 MCONTROL\_DMODE (*C macro*), 72  
 MCONTROL\_EXECUTE (*C macro*), 73  
 MCONTROL\_H (*C macro*), 73  
 MCONTROL\_LOAD (*C macro*), 73  
 MCONTROL\_M (*C macro*), 73  
 MCONTROL\_MASKMAX (*C macro*), 72  
 MCONTROL\_MATCH (*C macro*), 73  
 MCONTROL\_MATCH\_EQUAL (*C macro*), 73  
 MCONTROL\_MATCH\_GE (*C macro*), 73  
 MCONTROL\_MATCH\_LT (*C macro*), 73  
 MCONTROL\_MATCH\_MASK\_HIGH (*C macro*), 73  
 MCONTROL\_MATCH\_MASK\_LOW (*C macro*), 73  
 MCONTROL\_MATCH\_NAPOT (*C macro*), 73  
 MCONTROL\_S (*C macro*), 73  
 MCONTROL\_SELECT (*C macro*), 72  
 MCONTROL\_STORE (*C macro*), 73  
 MCONTROL\_TIMING (*C macro*), 73  
 MCONTROL\_TYPE (*C macro*), 72  
 MCONTROL\_TYPE\_MATCH (*C macro*), 73  
 MCONTROL\_TYPE\_NONE (*C macro*), 73  
 MCONTROL\_U (*C macro*), 73  
 MCOUNTINHIBIT\_CY (*C macro*), 74  
 MCOUNTINHIBIT\_IR (*C macro*), 74  
 MDCAUSE\_MDCAUSE (*C macro*), 74  
 MDCFG\_DC\_ECC (*C macro*), 75  
 MDCFG\_DC\_LSIZE (*C macro*), 75  
 MDCFG\_DC\_SET (*C macro*), 75  
 MDCFG\_DC\_WAY (*C macro*), 75  
 MDCFG\_DLM\_ECC (*C macro*), 75  
 MDCFG\_DLM\_SIZE (*C macro*), 75  
 MDLM\_CTL\_DLM\_BPA (*C macro*), 74  
 MDLM\_CTL\_DLM\_ECC\_EN (*C macro*), 74  
 MDLM\_CTL\_DLM\_ECC\_EXCP\_EN (*C macro*), 74  
 MDLM\_CTL\_DLM\_EN (*C macro*), 74  
 MDLM\_CTL\_DLM\_RWECC (*C macro*), 74  
 MECC\_CODE\_CODE (*C macro*), 75  
 MECC\_CODE\_RAMID (*C macro*), 75  
 MECC\_CODE\_SRAMID (*C macro*), 76  
 MECC\_LOCK\_ECC\_LOCK (*C macro*), 75  
 MFIOCFG\_INFO\_FIO\_BPA (*C macro*), 75  
 MFIOCFG\_INFO\_FIO\_SIZE (*C macro*), 75  
 MICFG\_IC\_ECC (*C macro*), 75  
 MICFG\_IC\_LSIZE (*C macro*), 75  
 MICFG\_IC\_SET (*C macro*), 75  
 MICFG\_IC\_WAY (*C macro*), 75  
 MICFG\_ILM\_ECC (*C macro*), 75  
 MICFG\_ILM\_SIZE (*C macro*), 75  
 MICFG\_ILM\_XONLY (*C macro*), 75  
 MIE\_HEIE (*C macro*), 74  
 MIE\_HSIE (*C macro*), 73  
 MIE\_HTIE (*C macro*), 74  
 MIE\_MEIE (*C macro*), 74  
 MIE\_MSIE (*C macro*), 73  
 MIE\_MTIE (*C macro*), 74  
 MIE\_SEIE (*C macro*), 74  
 MIE\_SSIE (*C macro*), 73  
 MIE\_STIE (*C macro*), 74  
 MILM\_CTL\_ILM\_BPA (*C macro*), 74  
 MILM\_CTL\_ILM\_ECC\_EN (*C macro*), 74  
 MILM\_CTL\_ILM\_ECC\_EXCP\_EN (*C macro*), 74  
 MILM\_CTL\_ILM\_EN (*C macro*), 74  
 MILM\_CTL\_ILM\_RWECC (*C macro*), 74  
 MIP\_HEIP (*C macro*), 73  
 MIP\_HSIP (*C macro*), 73

MIP\_HTIP (*C macro*), 73  
 MIP\_MEIP (*C macro*), 73  
 MIP\_MSIP (*C macro*), 73  
 MIP\_MTIP (*C macro*), 73  
 MIP\_SEIP (*C macro*), 73  
 MIP\_SSID (*C macro*), 73  
 MIP\_STIP (*C macro*), 73  
 MMISC\_CTL\_BPU (*C macro*), 74  
 MMISC\_CTL\_MISALIGN (*C macro*), 74  
 MMISC\_CTL\_NMI\_CAUSE\_FFF (*C macro*), 74  
 MPPICFG\_INFO\_PPI\_BPA (*C macro*), 75  
 MPPICFG\_INFO\_PPI\_SIZE (*C macro*), 75  
 MSTATUS32\_SD (*C macro*), 71  
 MSTATUS64\_SD (*C macro*), 71  
 MSTATUS\_FS (*C macro*), 71  
 MSTATUS\_FS\_CLEAN (*C macro*), 71  
 MSTATUS\_FS\_DIRTY (*C macro*), 71  
 MSTATUS\_FS\_INITIAL (*C macro*), 71  
 MSTATUS\_HIE (*C macro*), 71  
 MSTATUS\_HPIE (*C macro*), 71  
 MSTATUS\_MIE (*C macro*), 71  
 MSTATUS\_MPIE (*C macro*), 71  
 MSTATUS\_MPP (*C macro*), 71  
 MSTATUS\_MPRV (*C macro*), 71  
 MSTATUS\_MXR (*C macro*), 71  
 MSTATUS\_PUM (*C macro*), 71  
 MSTATUS\_SIE (*C macro*), 71  
 MSTATUS\_SPIE (*C macro*), 71  
 MSTATUS\_SPP (*C macro*), 71  
 MSTATUS\_UIE (*C macro*), 71  
 MSTATUS\_UPIE (*C macro*), 71  
 MSTATUS\_VM (*C macro*), 71  
 MSTATUS\_XS (*C macro*), 71  
 MSUBM\_PTyp (*C macro*), 74  
 MSUBM\_Typ (*C macro*), 74  
 MTVT2\_COMMON\_CODE\_ENTRY (*C macro*), 75  
 MTVT2\_MTVT2EN (*C macro*), 75

## N

NMSIS\_Core\_Cache::CCM\_CMD\_Type (*C++ enum*), 387, 388  
 NMSIS\_Core\_Cache::CCM\_DC\_INVAL (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_DC\_INVAL\_ALL (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_DC\_LOCK (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_DC\_UNLOCK (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_DC\_WB (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_DC\_WB\_ALL (*C++ enumerator*), 387, 389

NMSIS\_Core\_Cache::CCM\_DC\_WBINVAL (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_DC\_WBINVAL\_ALL (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_IC\_INVAL (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_IC\_INVAL\_ALL (*C++ enumerator*), 388, 389  
 NMSIS\_Core\_Cache::CCM\_IC\_LOCK (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_IC\_UNLOCK (*C++ enumerator*), 387, 389  
 NMSIS\_Core\_Cache::CCM\_OP\_ECC\_ERR (*C++ enumerator*), 387, 388  
 NMSIS\_Core\_Cache::CCM\_OP\_EXCEED\_ERR (*C++ enumerator*), 387, 388  
 NMSIS\_Core\_Cache::CCM\_OP\_FINFO\_Type (*C++ enum*), 387, 388  
 NMSIS\_Core\_Cache::CCM\_OP\_PERM\_CHECK\_ERR (*C++ enumerator*), 387, 388  
 NMSIS\_Core\_Cache::CCM\_OP\_REFILL\_BUS\_ERR (*C++ enumerator*), 387, 388  
 NMSIS\_Core\_Cache::CCM\_OP\_SUCCESS (*C++ enumerator*), 387, 388  
 NMSIS\_Core\_CPU\_Intrinsic::WFI\_DEEP\_SLEEP (*C++ enumerator*), 93, 95  
 NMSIS\_Core\_CPU\_Intrinsic::WFI\_SHALLOW\_SLEEP (*C++ enumerator*), 93, 95  
 NMSIS\_Core\_CPU\_Intrinsic::WFI\_SleepMode\_Type (*C++ enum*), 93, 95  
 NMSIS\_Core\_ECLIC\_Registers::ECLIC\_LEVEL\_TRIGGER (*C++ enumerator*), 90  
 NMSIS\_Core\_ECLIC\_Registers::ECLIC\_MAX\_TRIGGER (*C++ enumerator*), 91  
 NMSIS\_Core\_ECLIC\_Registers::ECLIC\_NEGTIVE\_EDGE\_TRIGGER (*C++ enumerator*), 91  
 NMSIS\_Core\_ECLIC\_Registers::ECLIC\_POSTIVE\_EDGE\_TRIGGER (*C++ enumerator*), 91  
 NMSIS\_Core\_ECLIC\_Registers::ECLIC\_TRIGGER\_Type (*C++ enum*), 90  
 NMSIS\_Core\_IntExc::FirstDeviceSpecificInterrupt\_IRQn (*C++ enumerator*), 369, 373  
 NMSIS\_Core\_IntExc::IRQn\_Type (*C++ enum*), 368, 372  
 NMSIS\_Core\_IntExc::Reserved0\_IRQn (*C++ enumerator*), 368, 372  
 NMSIS\_Core\_IntExc::Reserved10\_IRQn (*C++ enumerator*), 369, 373  
 NMSIS\_Core\_IntExc::Reserved11\_IRQn (*C++ enumerator*), 369, 373  
 NMSIS\_Core\_IntExc::Reserved12\_IRQn (*C++ enumerator*), 369, 373  
 NMSIS\_Core\_IntExc::Reserved13\_IRQn (*C++ enumerator*), 369, 373

NMSIS\_Core\_IntExc::Reserved14\_IRQn (C++  
enumerator), 369, 373

NMSIS\_Core\_IntExc::Reserved15\_IRQn (C++  
enumerator), 369, 373

NMSIS\_Core\_IntExc::Reserved16\_IRQn (C++  
enumerator), 369, 373

NMSIS\_Core\_IntExc::Reserved1\_IRQn (C++  
enumerator), 368, 372

NMSIS\_Core\_IntExc::Reserved2\_IRQn (C++  
enumerator), 368, 372

NMSIS\_Core\_IntExc::Reserved3\_IRQn (C++  
enumerator), 368, 372

NMSIS\_Core\_IntExc::Reserved4\_IRQn (C++  
enumerator), 368, 372

NMSIS\_Core\_IntExc::Reserved5\_IRQn (C++  
enumerator), 368, 372

NMSIS\_Core\_IntExc::Reserved6\_IRQn (C++  
enumerator), 368, 373

NMSIS\_Core\_IntExc::Reserved7\_IRQn (C++  
enumerator), 368, 373

NMSIS\_Core\_IntExc::Reserved8\_IRQn (C++  
enumerator), 368, 373

NMSIS\_Core\_IntExc::Reserved9\_IRQn (C++  
enumerator), 368, 373

NMSIS\_Core\_IntExc::SOC\_INT\_MAX (C++ enu-  
merator), 369, 373

NMSIS\_Core\_IntExc::SysTimer\_IRQn (C++  
enumerator), 368, 373

NMSIS\_Core\_IntExc::SysTimerSW\_IRQn (C++  
enumerator), 368, 372

NN, 741

## P

PMP\_A (C macro), 77

PMP\_A\_NA4 (C macro), 77

PMP\_A\_NAPOT (C macro), 77

PMP\_A\_TOR (C macro), 77

PMP\_COUNT (C macro), 77

PMP\_L (C macro), 77

PMP\_R (C macro), 77

PMP\_SHIFT (C macro), 77

PMP\_W (C macro), 77

PMP\_X (C macro), 77

PRV\_H (C macro), 76

PRV\_M (C macro), 76

PRV\_S (C macro), 76

PRV\_U (C macro), 76

PTE\_A (C macro), 77

PTE\_D (C macro), 77

PTE\_G (C macro), 77

PTE\_PPN\_SHIFT (C macro), 77

PTE\_R (C macro), 77

PTE\_SOFT (C macro), 77

PTE\_TABLE (C macro), 77

PTE\_U (C macro), 77

PTE\_V (C macro), 77

PTE\_W (C macro), 77

PTE\_X (C macro), 77

## R

realCoefA (C++ member), 658, 659

realCoefAQ31 (C++ member), 658, 659

realCoefB (C++ member), 658, 659

realCoefBQ31 (C++ member), 658, 659

RESTORE\_FPU\_CONTEXT (C macro), 385

RESTORE\_IRQ\_CSR\_CONTEXT (C macro), 370, 372

riscv\_abs\_f16 (C++ function), 430, 431

riscv\_abs\_f32 (C++ function), 430, 431

riscv\_abs\_q15 (C++ function), 430, 431

riscv\_abs\_q31 (C++ function), 430, 431

riscv\_abs\_q7 (C++ function), 430, 432

riscv\_absmax\_f16 (C++ function), 592

riscv\_absmax\_f32 (C++ function), 592, 593

riscv\_absmax\_q15 (C++ function), 592, 593

riscv\_absmax\_q31 (C++ function), 592, 593

riscv\_absmax\_q7 (C++ function), 592, 593

riscv\_absmin\_f16 (C++ function), 594

riscv\_absmin\_f32 (C++ function), 594

riscv\_absmin\_q15 (C++ function), 594

riscv\_absmin\_q31 (C++ function), 594, 595

riscv\_absmin\_q7 (C++ function), 594, 595

riscv\_add\_f16 (C++ function), 432

riscv\_add\_f32 (C++ function), 432

riscv\_add\_q15 (C++ function), 432, 433

riscv\_add\_q31 (C++ function), 432, 433

riscv\_add\_q7 (C++ function), 432, 433

riscv\_and\_u16 (C++ function), 434

riscv\_and\_u32 (C++ function), 434

riscv\_and\_u8 (C++ function), 434

riscv\_avepool\_q7\_HWC (C++ function), 718, 720

riscv\_avgpool\_s8 (C++ function), 718, 719

riscv\_avgpool\_s8\_get\_buffer\_size (C++  
function), 718, 719

riscv\_barycenter\_f16 (C++ function), 610

riscv\_barycenter\_f32 (C++ function), 610

riscv\_bilinear\_interp\_f16 (C++ function),  
557, 558

riscv\_bilinear\_interp\_f32 (C++ function),  
557, 558

riscv\_bilinear\_interp\_q15 (C++ function),  
557, 558

riscv\_bilinear\_interp\_q31 (C++ function),  
557, 558

riscv\_bilinear\_interp\_q7 (C++ function), 557,  
558

riscv\_biquad\_cas\_df1\_32x64\_init\_q31  
(C++ function), 485, 487

riscv\_biquad\_cas\_df1\_32x64\_q31 (C++ function), 485, 488  
 riscv\_biquad\_cascade\_df1\_f16 (C++ function), 488, 491  
 riscv\_biquad\_cascade\_df1\_f32 (C++ function), 488, 492  
 riscv\_biquad\_cascade\_df1\_fast\_q15 (C++ function), 489, 492  
 riscv\_biquad\_cascade\_df1\_fast\_q31 (C++ function), 489, 492  
 riscv\_biquad\_cascade\_df1\_init\_f16 (C++ function), 489, 493  
 riscv\_biquad\_cascade\_df1\_init\_f32 (C++ function), 489, 493  
 riscv\_biquad\_cascade\_df1\_init\_q15 (C++ function), 489, 494  
 riscv\_biquad\_cascade\_df1\_init\_q31 (C++ function), 489, 495  
 riscv\_biquad\_cascade\_df1\_q15 (C++ function), 489, 495  
 riscv\_biquad\_cascade\_df1\_q31 (C++ function), 489, 495  
 riscv\_biquad\_cascade\_df2T\_init\_f16 (C++ function), 496, 499  
 riscv\_biquad\_cascade\_df2T\_init\_f32 (C++ function), 496, 499  
 riscv\_biquad\_cascade\_df2T\_init\_f64 (C++ function), 496, 500  
 riscv\_biquad\_cascade\_stereo\_df2T\_init\_f16 (C++ function), 496, 501  
 riscv\_biquad\_cascade\_stereo\_df2T\_init\_f32 (C++ function), 496, 501  
 riscv\_bitonic\_sort\_f32 (C++ function), 611  
 riscv\_braycurtis\_distance\_f16 (C++ function), 471  
 riscv\_braycurtis\_distance\_f32 (C++ function), 471, 472  
 riscv\_bubble\_sort\_f32 (C++ function), 611  
 riscv\_canberra\_distance\_f16 (C++ function), 472  
 riscv\_canberra\_distance\_f32 (C++ function), 472  
 riscv\_cfft\_f16 (C++ function), 639, 642  
 riscv\_cfft\_f32 (C++ function), 639, 642  
 riscv\_cfft\_f64 (C++ function), 639, 643  
 riscv\_cfft\_init\_f16 (C++ function), 639, 643  
 riscv\_cfft\_init\_f32 (C++ function), 639, 643  
 riscv\_cfft\_init\_f64 (C++ function), 639, 643  
 riscv\_cfft\_init\_q15 (C++ function), 639, 644  
 riscv\_cfft\_init\_q31 (C++ function), 639, 644  
 riscv\_cfft\_q15 (C++ function), 639, 644  
 riscv\_cfft\_q31 (C++ function), 639, 645  
 riscv\_cfft\_radix2\_f16 (C++ function), 639, 645  
 riscv\_cfft\_radix2\_f32 (C++ function), 639, 645  
 riscv\_cfft\_radix2\_init\_f16 (C++ function), 639, 645  
 riscv\_cfft\_radix2\_init\_f32 (C++ function), 639, 646  
 riscv\_cfft\_radix2\_init\_q15 (C++ function), 639, 647  
 riscv\_cfft\_radix2\_init\_q31 (C++ function), 639, 647  
 riscv\_cfft\_radix2\_q15 (C++ function), 639, 648  
 riscv\_cfft\_radix2\_q31 (C++ function), 639, 648  
 riscv\_cfft\_radix4\_f16 (C++ function), 639, 648  
 riscv\_cfft\_radix4\_f32 (C++ function), 639, 648  
 riscv\_cfft\_radix4\_init\_f16 (C++ function), 639, 648  
 riscv\_cfft\_radix4\_init\_f32 (C++ function), 639, 649  
 riscv\_cfft\_radix4\_init\_q15 (C++ function), 639, 650  
 riscv\_cfft\_radix4\_init\_q31 (C++ function), 639, 650  
 riscv\_cfft\_radix4\_q15 (C++ function), 639, 651  
 riscv\_cfft\_radix4\_q31 (C++ function), 639, 651  
 riscv\_cfft\_radix4by2\_f16 (C++ function), 639, 648  
 riscv\_chebyshev\_distance\_f16 (C++ function), 473  
 riscv\_chebyshev\_distance\_f32 (C++ function), 473  
 riscv\_cityblock\_distance\_f16 (C++ function), 473, 474  
 riscv\_cityblock\_distance\_f32 (C++ function), 473, 474  
 riscv\_clip\_f16 (C++ function), 435  
 riscv\_clip\_f32 (C++ function), 435  
 riscv\_clip\_q15 (C++ function), 435  
 riscv\_clip\_q31 (C++ function), 435, 436  
 riscv\_clip\_q7 (C++ function), 435, 436  
 riscv\_cmplx\_conj\_f16 (C++ function), 452  
 riscv\_cmplx\_conj\_f32 (C++ function), 452, 453  
 riscv\_cmplx\_conj\_q15 (C++ function), 452, 453  
 riscv\_cmplx\_conj\_q31 (C++ function), 452, 453  
 riscv\_cmplx\_dot\_prod\_f16 (C++ function), 453, 454  
 riscv\_cmplx\_dot\_prod\_f32 (C++ function), 453, 454  
 riscv\_cmplx\_dot\_prod\_q15 (C++ function), 453, 454  
 riscv\_cmplx\_dot\_prod\_q31 (C++ function), 453, 455  
 riscv\_cmplx\_mag\_f16 (C++ function), 455  
 riscv\_cmplx\_mag\_f32 (C++ function), 455, 456  
 riscv\_cmplx\_mag\_q15 (C++ function), 455, 456  
 riscv\_cmplx\_mag\_q31 (C++ function), 455, 456  
 riscv\_cmplx\_mag\_squared\_f16 (C++ function),

456, 457  
 riscv\_cmplx\_mag\_squared\_f32 (C++ *function*),  
 456, 457  
 riscv\_cmplx\_mag\_squared\_q15 (C++ *function*),  
 456, 457  
 riscv\_cmplx\_mag\_squared\_q31 (C++ *function*),  
 456, 457  
 riscv\_cmplx\_mult\_cmplx\_f16 (C++ *function*),  
 458  
 riscv\_cmplx\_mult\_cmplx\_f32 (C++ *function*),  
 458  
 riscv\_cmplx\_mult\_cmplx\_q15 (C++ *function*),  
 458, 459  
 riscv\_cmplx\_mult\_cmplx\_q31 (C++ *function*),  
 458, 459  
 riscv\_cmplx\_mult\_real\_f16 (C++ *function*),  
 459, 460  
 riscv\_cmplx\_mult\_real\_f32 (C++ *function*),  
 459, 460  
 riscv\_cmplx\_mult\_real\_q15 (C++ *function*),  
 459, 460  
 riscv\_cmplx\_mult\_real\_q31 (C++ *function*),  
 459, 460  
 riscv\_concatenation\_s8\_w (C++ *function*), 681  
 riscv\_concatenation\_s8\_x (C++ *function*), 681,  
 682  
 riscv\_concatenation\_s8\_y (C++ *function*), 681,  
 682  
 riscv\_concatenation\_s8\_z (C++ *function*), 681,  
 683  
 riscv\_conv\_f32 (C++ *function*), 502, 503  
 riscv\_conv\_fast\_opt\_q15 (C++ *function*), 502,  
 503  
 riscv\_conv\_fast\_q15 (C++ *function*), 502, 504  
 riscv\_conv\_fast\_q31 (C++ *function*), 502, 504  
 riscv\_conv\_opt\_q15 (C++ *function*), 502, 505  
 riscv\_conv\_opt\_q7 (C++ *function*), 502, 505  
 riscv\_conv\_partial\_f32 (C++ *function*), 507,  
 508  
 riscv\_conv\_partial\_fast\_opt\_q15 (C++  
*function*), 507, 509  
 riscv\_conv\_partial\_fast\_q15 (C++ *function*),  
 507, 509  
 riscv\_conv\_partial\_fast\_q31 (C++ *function*),  
 507, 510  
 riscv\_conv\_partial\_opt\_q15 (C++ *function*),  
 507, 510  
 riscv\_conv\_partial\_opt\_q7 (C++ *function*),  
 508, 511  
 riscv\_conv\_partial\_q15 (C++ *function*), 508,  
 511  
 riscv\_conv\_partial\_q31 (C++ *function*), 508,  
 512  
 riscv\_conv\_partial\_q7 (C++ *function*), 508, 512  
 riscv\_conv\_q15 (C++ *function*), 502, 506  
 riscv\_conv\_q31 (C++ *function*), 502, 506  
 riscv\_conv\_q7 (C++ *function*), 502, 507  
 riscv\_convolve\_1\_x\_n\_s8 (C++ *function*), 684,  
 690  
 riscv\_convolve\_1\_x\_n\_s8\_get\_buffer\_size  
 (C++ *function*), 684, 690  
 riscv\_convolve\_1x1\_HWC\_q7\_fast\_nonsquare  
 (C++ *function*), 684, 691  
 riscv\_convolve\_1x1\_s8\_fast (C++ *function*),  
 684, 692  
 riscv\_convolve\_1x1\_s8\_fast\_get\_buffer\_size  
 (C++ *function*), 684, 692  
 riscv\_convolve\_HWC\_q15\_basic (C++ *func-*  
*tion*), 684, 693  
 riscv\_convolve\_HWC\_q15\_fast (C++ *function*),  
 685, 693  
 riscv\_convolve\_HWC\_q15\_fast\_nonsquare  
 (C++ *function*), 685, 694  
 riscv\_convolve\_HWC\_q7\_basic (C++ *function*),  
 685, 696  
 riscv\_convolve\_HWC\_q7\_basic\_nonsquare  
 (C++ *function*), 685, 696  
 riscv\_convolve\_HWC\_q7\_fast (C++ *function*),  
 685, 697  
 riscv\_convolve\_HWC\_q7\_fast\_nonsquare  
 (C++ *function*), 686, 698  
 riscv\_convolve\_HWC\_q7\_RGB (C++ *function*),  
 686, 699  
 riscv\_convolve\_s8 (C++ *function*), 686, 700  
 riscv\_convolve\_s8\_get\_buffer\_size (C++  
*function*), 686, 701  
 riscv\_convolve\_wrapper\_s8 (C++ *function*),  
 686, 701  
 riscv\_convolve\_wrapper\_s8\_get\_buffer\_size  
 (C++ *function*), 686, 702  
 riscv\_copy\_f16 (C++ *function*), 614  
 riscv\_copy\_f32 (C++ *function*), 614  
 riscv\_copy\_q15 (C++ *function*), 614  
 riscv\_copy\_q31 (C++ *function*), 614  
 riscv\_copy\_q7 (C++ *function*), 614, 615  
 riscv\_correlate\_f16 (C++ *function*), 513, 514  
 riscv\_correlate\_f32 (C++ *function*), 513, 514  
 riscv\_correlate\_fast\_opt\_q15 (C++ *func-*  
*tion*), 513, 515  
 riscv\_correlate\_fast\_q15 (C++ *function*), 513,  
 515  
 riscv\_correlate\_fast\_q31 (C++ *function*), 513,  
 516  
 riscv\_correlate\_opt\_q15 (C++ *function*), 513,  
 516  
 riscv\_correlate\_opt\_q7 (C++ *function*), 513,  
 517  
 riscv\_correlate\_q15 (C++ *function*), 513, 517



riscv\_correlate\_q31 (C++ function), 513, 518  
 riscv\_correlate\_q7 (C++ function), 513, 518  
 riscv\_correlation\_distance\_f16 (C++ function), 474  
 riscv\_correlation\_distance\_f32 (C++ function), 474  
 riscv\_cos\_f32 (C++ function), 480, 481  
 riscv\_cos\_q15 (C++ function), 481  
 riscv\_cos\_q31 (C++ function), 480, 481  
 riscv\_cosine\_distance\_f16 (C++ function), 475  
 riscv\_cosine\_distance\_f32 (C++ function), 475  
 riscv\_dct4\_f32 (C++ function), 653, 655  
 riscv\_dct4\_init\_f32 (C++ function), 653, 655  
 riscv\_dct4\_init\_q15 (C++ function), 653, 656  
 riscv\_dct4\_init\_q31 (C++ function), 653, 656  
 riscv\_dct4\_q15 (C++ function), 654, 657  
 riscv\_dct4\_q31 (C++ function), 654, 658  
 riscv\_depthwise\_conv\_3x3\_s8 (C++ function), 686, 702  
 riscv\_depthwise\_conv\_s8 (C++ function), 687, 703  
 riscv\_depthwise\_conv\_s8\_opt (C++ function), 687, 704  
 riscv\_depthwise\_conv\_s8\_opt\_get\_buffer\_size (C++ function), 687, 705  
 riscv\_depthwise\_conv\_u8\_basic\_ver1 (C++ function), 688, 705  
 riscv\_depthwise\_conv\_wrapper\_s8 (C++ function), 688, 707  
 riscv\_depthwise\_conv\_wrapper\_s8\_get\_buffer\_size (C++ function), 688, 708  
 riscv\_depthwise\_separable\_conv\_HWC\_q7 (C++ function), 689, 708  
 riscv\_depthwise\_separable\_conv\_HWC\_q7\_non\_square (C++ function), 689, 709  
 riscv\_dice\_distance (C++ function), 478  
 riscv\_divide\_q15 (C++ function), 482  
 riscv\_dot\_prod\_f16 (C++ function), 436, 437  
 riscv\_dot\_prod\_f32 (C++ function), 436, 437  
 riscv\_dot\_prod\_q15 (C++ function), 436, 437  
 riscv\_dot\_prod\_q31 (C++ function), 436, 437  
 riscv\_dot\_prod\_q7 (C++ function), 436, 438  
 riscv\_elementwise\_add\_s8 (C++ function), 679, 680  
 riscv\_elementwise\_mul\_s8 (C++ function), 679, 680  
 riscv\_entropy\_f16 (C++ function), 595  
 riscv\_entropy\_f32 (C++ function), 595, 596  
 riscv\_entropy\_f64 (C++ function), 595, 596  
 riscv\_euclidean\_distance\_f16 (C++ function), 475, 476  
 riscv\_euclidean\_distance\_f32 (C++ function), 475, 476  
 riscv\_f16\_to\_float (C++ function), 615  
 riscv\_f16\_to\_q15 (C++ function), 615  
 riscv\_fill\_f16 (C++ function), 616  
 riscv\_fill\_f32 (C++ function), 616  
 riscv\_fill\_q15 (C++ function), 616  
 riscv\_fill\_q31 (C++ function), 616, 617  
 riscv\_fill\_q7 (C++ function), 616, 617  
 riscv\_fir\_decimate\_f32 (C++ function), 519, 520  
 riscv\_fir\_decimate\_fast\_q15 (C++ function), 519, 520  
 riscv\_fir\_decimate\_fast\_q31 (C++ function), 519, 521  
 riscv\_fir\_decimate\_init\_f32 (C++ function), 519, 521  
 riscv\_fir\_decimate\_init\_q15 (C++ function), 519, 522  
 riscv\_fir\_decimate\_init\_q31 (C++ function), 519, 522  
 riscv\_fir\_decimate\_q15 (C++ function), 519, 523  
 riscv\_fir\_decimate\_q31 (C++ function), 519, 523  
 riscv\_fir\_f16 (C++ function), 524, 526  
 riscv\_fir\_f32 (C++ function), 524, 526  
 riscv\_fir\_fast\_q15 (C++ function), 524, 526  
 riscv\_fir\_init\_f16 (C++ function), 524, 527  
 riscv\_fir\_init\_f32 (C++ function), 524, 528  
 riscv\_fir\_init\_q15 (C++ function), 524, 528  
 riscv\_fir\_init\_q31 (C++ function), 524, 529  
 riscv\_fir\_init\_q7 (C++ function), 524, 529  
 riscv\_fir\_interpolate\_f32 (C++ function), 552, 554  
 riscv\_fir\_interpolate\_init\_f32 (C++ function), 552, 554  
 riscv\_fir\_interpolate\_init\_q15 (C++ function), 553, 555  
 riscv\_fir\_interpolate\_init\_q31 (C++ function), 553, 555  
 riscv\_fir\_interpolate\_q15 (C++ function), 553, 556  
 riscv\_fir\_interpolate\_q31 (C++ function), 553, 556  
 riscv\_fir\_lattice\_f32 (C++ function), 531, 533  
 riscv\_fir\_lattice\_init\_f32 (C++ function), 531, 533  
 riscv\_fir\_lattice\_init\_q15 (C++ function), 531, 533  
 riscv\_fir\_lattice\_init\_q31 (C++ function), 531, 533  
 riscv\_fir\_lattice\_q15 (C++ function), 531, 534  
 riscv\_fir\_lattice\_q31 (C++ function), 531, 534

riscv\_fir\_q15 (C++ *function*), 524, 530  
 riscv\_fir\_q31 (C++ *function*), 524, 530  
 riscv\_fir\_q7 (C++ *function*), 524, 531  
 riscv\_fir\_sparse\_f32 (C++ *function*), 534, 536  
 riscv\_fir\_sparse\_init\_f32 (C++ *function*), 534, 536  
 riscv\_fir\_sparse\_init\_q15 (C++ *function*), 534, 536  
 riscv\_fir\_sparse\_init\_q31 (C++ *function*), 534, 537  
 riscv\_fir\_sparse\_init\_q7 (C++ *function*), 534, 537  
 riscv\_fir\_sparse\_q15 (C++ *function*), 534, 538  
 riscv\_fir\_sparse\_q31 (C++ *function*), 534, 538  
 riscv\_fir\_sparse\_q7 (C++ *function*), 534, 538  
 riscv\_float\_to\_f16 (C++ *function*), 617  
 riscv\_float\_to\_q15 (C++ *function*), 617  
 riscv\_float\_to\_q31 (C++ *function*), 617, 618  
 riscv\_float\_to\_q7 (C++ *function*), 617, 618  
 riscv\_fully\_connected\_mat\_q7\_vec\_q15 (C++ *function*), 711, 713  
 riscv\_fully\_connected\_mat\_q7\_vec\_q15\_opt (C++ *function*), 711, 713  
 riscv\_fully\_connected\_q15 (C++ *function*), 711, 714  
 riscv\_fully\_connected\_q15\_opt (C++ *function*), 711, 714  
 riscv\_fully\_connected\_q7 (C++ *function*), 711, 715  
 riscv\_fully\_connected\_q7\_opt (C++ *function*), 711, 716  
 riscv\_fully\_connected\_s8 (C++ *function*), 711, 717  
 riscv\_fully\_connected\_s8\_get\_buffer\_size (C++ *function*), 712, 718  
 riscv\_gaussian\_naive\_bayes\_predict\_f16 (C++ *function*), 451  
 riscv\_gaussian\_naive\_bayes\_predict\_f32 (C++ *function*), 451, 452  
 riscv\_hamming\_distance (C++ *function*), 478, 479  
 riscv\_heap\_sort\_f32 (C++ *function*), 611  
 riscv\_iir\_lattice\_f32 (C++ *function*), 539, 540  
 riscv\_iir\_lattice\_init\_f32 (C++ *function*), 539, 541  
 riscv\_iir\_lattice\_init\_q15 (C++ *function*), 539, 541  
 riscv\_iir\_lattice\_init\_q31 (C++ *function*), 539, 541  
 riscv\_iir\_lattice\_q15 (C++ *function*), 539, 542  
 riscv\_iir\_lattice\_q31 (C++ *function*), 539, 542  
 riscv\_insertion\_sort\_f32 (C++ *function*), 612  
 riscv\_jaccard\_distance (C++ *function*), 478, 479  
 riscv\_jensenshannon\_distance\_f16 (C++ *function*), 476  
 riscv\_jensenshannon\_distance\_f32 (C++ *function*), 476, 477  
 riscv\_kullback\_leibler\_f16 (C++ *function*), 596  
 riscv\_kullback\_leibler\_f32 (C++ *function*), 596  
 riscv\_kullback\_leibler\_f64 (C++ *function*), 596, 597  
 riscv\_kulsinski\_distance (C++ *function*), 478, 479  
 riscv\_levinson\_durbin\_f16 (C++ *function*), 542, 543  
 riscv\_levinson\_durbin\_f32 (C++ *function*), 542, 543  
 riscv\_levinson\_durbin\_q31 (C++ *function*), 542, 543  
 riscv\_linear\_interp\_f16 (C++ *function*), 559, 561  
 riscv\_linear\_interp\_f32 (C++ *function*), 559, 560  
 riscv\_linear\_interp\_q15 (C++ *function*), 559, 561  
 riscv\_linear\_interp\_q31 (C++ *function*), 559, 561  
 riscv\_linear\_interp\_q7 (C++ *function*), 559, 561  
 riscv\_lms\_f32 (C++ *function*), 543, 545  
 riscv\_lms\_init\_f32 (C++ *function*), 543, 546  
 riscv\_lms\_init\_q15 (C++ *function*), 543, 546  
 riscv\_lms\_init\_q31 (C++ *function*), 543, 546  
 riscv\_lms\_norm\_f32 (C++ *function*), 548, 550  
 riscv\_lms\_norm\_init\_f32 (C++ *function*), 548, 550  
 riscv\_lms\_norm\_init\_q15 (C++ *function*), 548, 551  
 riscv\_lms\_norm\_init\_q31 (C++ *function*), 548, 551  
 riscv\_lms\_norm\_q15 (C++ *function*), 548, 551  
 riscv\_lms\_norm\_q31 (C++ *function*), 548, 552  
 riscv\_lms\_q15 (C++ *function*), 543, 547  
 riscv\_lms\_q31 (C++ *function*), 544, 547  
 riscv\_logsumexp\_dot\_prod\_f16 (C++ *function*), 597  
 riscv\_logsumexp\_dot\_prod\_f32 (C++ *function*), 597, 598  
 riscv\_logsumexp\_f16 (C++ *function*), 597, 598  
 riscv\_logsumexp\_f32 (C++ *function*), 597, 598  
 riscv\_mat\_add\_f16 (C++ *function*), 563, 564  
 riscv\_mat\_add\_f32 (C++ *function*), 563, 564  
 riscv\_mat\_add\_q15 (C++ *function*), 563, 564  
 riscv\_mat\_add\_q31 (C++ *function*), 563, 565  
 riscv\_mat\_cholesky\_f16 (C++ *function*), 565

riscv\_mat\_cholesky\_f32 (C++ function), 565, 566  
 riscv\_mat\_cholesky\_f64 (C++ function), 565, 566  
 riscv\_mat\_cmplx\_mult\_f16 (C++ function), 568  
 riscv\_mat\_cmplx\_mult\_f32 (C++ function), 568  
 riscv\_mat\_cmplx\_mult\_q15 (C++ function), 568, 569  
 riscv\_mat\_cmplx\_mult\_q31 (C++ function), 568, 569  
 riscv\_mat\_cmplx\_trans\_f16 (C++ function), 570  
 riscv\_mat\_cmplx\_trans\_f32 (C++ function), 570  
 riscv\_mat\_cmplx\_trans\_q15 (C++ function), 570, 571  
 riscv\_mat\_cmplx\_trans\_q31 (C++ function), 570, 571  
 riscv\_mat\_init\_f16 (C++ function), 571, 572  
 riscv\_mat\_init\_f32 (C++ function), 571, 572  
 riscv\_mat\_init\_f64 (C++ function), 571, 572  
 riscv\_mat\_init\_q15 (C++ function), 571, 572  
 riscv\_mat\_init\_q31 (C++ function), 571, 573  
 riscv\_mat\_inverse\_f16 (C++ function), 573, 574  
 riscv\_mat\_inverse\_f32 (C++ function), 573, 574  
 riscv\_mat\_inverse\_f64 (C++ function), 573, 574  
 riscv\_mat\_ldlt\_f32 (C++ function), 565, 567  
 riscv\_mat\_ldlt\_f64 (C++ function), 565, 567  
 riscv\_mat\_mult\_f16 (C++ function), 576, 577  
 riscv\_mat\_mult\_f32 (C++ function), 576, 577  
 riscv\_mat\_mult\_f64 (C++ function), 576, 578  
 riscv\_mat\_mult\_fast\_q15 (C++ function), 576, 578  
 riscv\_mat\_mult\_fast\_q31 (C++ function), 577, 578  
 riscv\_mat\_mult\_q15 (C++ function), 577, 579  
 riscv\_mat\_mult\_q31 (C++ function), 577, 579  
 riscv\_mat\_mult\_q7 (C++ function), 577, 580  
 riscv\_mat\_scale\_f16 (C++ function), 580, 581  
 riscv\_mat\_scale\_f32 (C++ function), 580, 581  
 riscv\_mat\_scale\_q15 (C++ function), 580, 581  
 riscv\_mat\_scale\_q31 (C++ function), 580, 582  
 riscv\_mat\_solve\_lower\_triangular\_f16 (C++ function), 573, 575  
 riscv\_mat\_solve\_lower\_triangular\_f32 (C++ function), 573, 575  
 riscv\_mat\_solve\_lower\_triangular\_f64 (C++ function), 573, 575  
 riscv\_mat\_solve\_upper\_triangular\_f16 (C++ function), 573, 575  
 riscv\_mat\_solve\_upper\_triangular\_f32 (C++ function), 573, 576  
 riscv\_mat\_solve\_upper\_triangular\_f64 (C++ function), 573, 576  
 riscv\_mat\_sub\_f16 (C++ function), 582, 583  
 riscv\_mat\_sub\_f32 (C++ function), 582, 583  
 riscv\_mat\_sub\_f64 (C++ function), 582, 583  
 riscv\_mat\_sub\_q15 (C++ function), 582, 584  
 riscv\_mat\_sub\_q31 (C++ function), 582, 584  
 riscv\_mat\_trans\_f16 (C++ function), 584, 585  
 riscv\_mat\_trans\_f32 (C++ function), 584, 585  
 riscv\_mat\_trans\_f64 (C++ function), 584, 585  
 riscv\_mat\_trans\_q15 (C++ function), 584, 586  
 riscv\_mat\_trans\_q31 (C++ function), 584, 586  
 riscv\_mat\_trans\_q7 (C++ function), 585, 586  
 riscv\_mat\_vec\_mult\_f16 (C++ function), 587  
 riscv\_mat\_vec\_mult\_f32 (C++ function), 587  
 riscv\_mat\_vec\_mult\_q15 (C++ function), 587  
 riscv\_mat\_vec\_mult\_q31 (C++ function), 587  
 riscv\_mat\_vec\_mult\_q7 (C++ function), 587, 588  
 riscv\_max\_f16 (C++ function), 599  
 riscv\_max\_f32 (C++ function), 599  
 riscv\_max\_no\_idx\_f16 (C++ function), 599  
 riscv\_max\_no\_idx\_f32 (C++ function), 599, 600  
 riscv\_max\_pool\_s8 (C++ function), 718, 719  
 riscv\_max\_q15 (C++ function), 599, 600  
 riscv\_max\_q31 (C++ function), 599, 600  
 riscv\_max\_q7 (C++ function), 599, 600  
 riscv\_maxpool\_q7\_HWC (C++ function), 718, 720  
 riscv\_mean\_f16 (C++ function), 601  
 riscv\_mean\_f32 (C++ function), 601  
 riscv\_mean\_q15 (C++ function), 601  
 riscv\_mean\_q31 (C++ function), 601, 602  
 riscv\_mean\_q7 (C++ function), 601, 602  
 riscv\_merge\_sort\_f32 (C++ function), 611, 612  
 riscv\_merge\_sort\_init\_f32 (C++ function), 611, 612  
 riscv\_min\_f16 (C++ function), 602  
 riscv\_min\_f32 (C++ function), 602, 603  
 riscv\_min\_q15 (C++ function), 602, 603  
 riscv\_min\_q31 (C++ function), 602, 603  
 riscv\_min\_q7 (C++ function), 602, 603  
 riscv\_minkowski\_distance\_f16 (C++ function), 477  
 riscv\_minkowski\_distance\_f32 (C++ function), 477  
 riscv\_mult\_f16 (C++ function), 438  
 riscv\_mult\_f32 (C++ function), 438, 439  
 riscv\_mult\_q15 (C++ function), 438, 439  
 riscv\_mult\_q31 (C++ function), 438, 439  
 riscv\_mult\_q7 (C++ function), 438, 439  
 riscv\_negate\_f16 (C++ function), 440  
 riscv\_negate\_f32 (C++ function), 440  
 riscv\_negate\_q15 (C++ function), 440, 441  
 riscv\_negate\_q31 (C++ function), 440, 441  
 riscv\_negate\_q7 (C++ function), 440, 441  
 riscv\_nn\_accumulate\_q7\_to\_q15 (C++ function), 726, 727



[riscv\\_nn\\_accumulate\\_q7\\_to\\_q7 \(C++ function\), 726, 727](#)  
[riscv\\_nn\\_activations\\_direct\\_q15 \(C++ function\), 678](#)  
[riscv\\_nn\\_activations\\_direct\\_q7 \(C++ function\), 678](#)  
[riscv\\_nn\\_add\\_q7 \(C++ function\), 726, 728](#)  
[riscv\\_nn\\_depthwise\\_conv\\_nt\\_t\\_padded\\_s8 \(C++ function\), 726, 728](#)  
[riscv\\_nn\\_depthwise\\_conv\\_nt\\_t\\_s8 \(C++ function\), 726, 729](#)  
[riscv\\_nn\\_mat\\_mul\\_core\\_1x\\_s8 \(C++ function\), 726, 730](#)  
[riscv\\_nn\\_mat\\_mul\\_core\\_4x\\_s8 \(C++ function\), 727, 730](#)  
[riscv\\_nn\\_mat\\_mult\\_nt\\_t\\_s8 \(C++ function\), 727, 730](#)  
[riscv\\_nn\\_mult\\_q15 \(C++ function\), 727, 731](#)  
[riscv\\_nn\\_mult\\_q7 \(C++ function\), 727, 731](#)  
[riscv\\_nn\\_vec\\_mat\\_mult\\_t\\_s8 \(C++ function\), 727, 732](#)  
[riscv\\_nn\\_vec\\_mat\\_mult\\_t\\_svd\\_s8 \(C++ function\), 727, 732](#)  
[riscv\\_not\\_u16 \(C++ function\), 441, 442](#)  
[riscv\\_not\\_u32 \(C++ function\), 441, 442](#)  
[riscv\\_not\\_u8 \(C++ function\), 441, 442](#)  
[riscv\\_offset\\_f16 \(C++ function\), 442, 443](#)  
[riscv\\_offset\\_f32 \(C++ function\), 442, 443](#)  
[riscv\\_offset\\_q15 \(C++ function\), 442, 443](#)  
[riscv\\_offset\\_q31 \(C++ function\), 442, 443](#)  
[riscv\\_offset\\_q7 \(C++ function\), 442, 444](#)  
[riscv\\_or\\_u16 \(C++ function\), 444](#)  
[riscv\\_or\\_u32 \(C++ function\), 444](#)  
[riscv\\_or\\_u8 \(C++ function\), 444, 445](#)  
[riscv\\_pid\\_init\\_f32 \(C++ function\), 461, 463](#)  
[riscv\\_pid\\_init\\_q15 \(C++ function\), 461, 463](#)  
[riscv\\_pid\\_init\\_q31 \(C++ function\), 461, 464](#)  
[riscv\\_pid\\_reset\\_f32 \(C++ function\), 461, 464](#)  
[riscv\\_pid\\_reset\\_q15 \(C++ function\), 461, 464](#)  
[riscv\\_pid\\_reset\\_q31 \(C++ function\), 461, 464](#)  
[riscv\\_power\\_f16 \(C++ function\), 604](#)  
[riscv\\_power\\_f32 \(C++ function\), 604](#)  
[riscv\\_power\\_q15 \(C++ function\), 604](#)  
[riscv\\_power\\_q31 \(C++ function\), 604, 605](#)  
[riscv\\_power\\_q7 \(C++ function\), 604, 605](#)  
[riscv\\_q15\\_to\\_f16 \(C++ function\), 619](#)  
[riscv\\_q15\\_to\\_float \(C++ function\), 619](#)  
[riscv\\_q15\\_to\\_q31 \(C++ function\), 619](#)  
[riscv\\_q15\\_to\\_q7 \(C++ function\), 619](#)  
[riscv\\_q31\\_to\\_float \(C++ function\), 620](#)  
[riscv\\_q31\\_to\\_q15 \(C++ function\), 620](#)  
[riscv\\_q31\\_to\\_q7 \(C++ function\), 620](#)  
[riscv\\_q7\\_to\\_float \(C++ function\), 621](#)  
[riscv\\_q7\\_to\\_q15 \(C++ function\), 621](#)  
[riscv\\_q7\\_to\\_q15\\_no\\_shift \(C++ function\), 724, 725](#)  
[riscv\\_q7\\_to\\_q15\\_reordered\\_no\\_shift \(C++ function\), 724, 725](#)  
[riscv\\_q7\\_to\\_q15\\_reordered\\_with\\_offset \(C++ function\), 724, 725](#)  
[riscv\\_q7\\_to\\_q15\\_with\\_offset \(C++ function\), 724, 725](#)  
[riscv\\_q7\\_to\\_q31 \(C++ function\), 621](#)  
[riscv\\_q7\\_to\\_q7\\_no\\_shift \(C++ function\), 724, 725](#)  
[riscv\\_q7\\_to\\_q7\\_reordered\\_no\\_shift \(C++ function\), 724, 726](#)  
[riscv\\_quaternion2rotation\\_f32 \(C++ function\), 588, 589](#)  
[riscv\\_quaternion\\_conjugate\\_f32 \(C++ function\), 589, 590](#)  
[riscv\\_quaternion\\_inverse\\_f32 \(C++ function\), 590](#)  
[riscv\\_quaternion\\_norm\\_f32 \(C++ function\), 590](#)  
[riscv\\_quaternion\\_normalize\\_f32 \(C++ function\), 591](#)  
[riscv\\_quaternion\\_product\\_f32 \(C++ function\), 591](#)  
[riscv\\_quaternion\\_product\\_single\\_f32 \(C++ function\), 592](#)  
[riscv\\_quick\\_sort\\_f32 \(C++ function\), 612](#)  
[riscv\\_relu6\\_s8 \(C++ function\), 678, 679](#)  
[riscv\\_relu\\_q15 \(C++ function\), 678, 679](#)  
[riscv\\_relu\\_q7 \(C++ function\), 678, 679](#)  
[riscv\\_reshape\\_s8 \(C++ function\), 721](#)  
[riscv\\_rfft\\_1024\\_fast\\_init\\_f16 \(C++ function\), 665](#)  
[riscv\\_rfft\\_1024\\_fast\\_init\\_f32 \(C++ function\), 667](#)  
[riscv\\_rfft\\_1024\\_fast\\_init\\_f64 \(C++ function\), 660, 669](#)  
[riscv\\_rfft\\_128\\_fast\\_init\\_f16 \(C++ function\), 664](#)  
[riscv\\_rfft\\_128\\_fast\\_init\\_f32 \(C++ function\), 666](#)  
[riscv\\_rfft\\_128\\_fast\\_init\\_f64 \(C++ function\), 660, 668](#)  
[riscv\\_rfft\\_2048\\_fast\\_init\\_f16 \(C++ function\), 665](#)  
[riscv\\_rfft\\_2048\\_fast\\_init\\_f32 \(C++ function\), 667](#)  
[riscv\\_rfft\\_2048\\_fast\\_init\\_f64 \(C++ function\), 660, 669](#)  
[riscv\\_rfft\\_256\\_fast\\_init\\_f16 \(C++ function\), 664](#)  
[riscv\\_rfft\\_256\\_fast\\_init\\_f32 \(C++ function\), 666](#)

[riscv\\_rfft\\_256\\_fast\\_init\\_f64 \(C++ function\), 660, 668](#)  
[riscv\\_rfft\\_32\\_fast\\_init\\_f16 \(C++ function\), 664](#)  
[riscv\\_rfft\\_32\\_fast\\_init\\_f32 \(C++ function\), 666](#)  
[riscv\\_rfft\\_32\\_fast\\_init\\_f64 \(C++ function\), 659, 668](#)  
[riscv\\_rfft\\_4096\\_fast\\_init\\_f16 \(C++ function\), 665](#)  
[riscv\\_rfft\\_4096\\_fast\\_init\\_f32 \(C++ function\), 667](#)  
[riscv\\_rfft\\_4096\\_fast\\_init\\_f64 \(C++ function\), 660, 669](#)  
[riscv\\_rfft\\_512\\_fast\\_init\\_f16 \(C++ function\), 665](#)  
[riscv\\_rfft\\_512\\_fast\\_init\\_f32 \(C++ function\), 667](#)  
[riscv\\_rfft\\_512\\_fast\\_init\\_f64 \(C++ function\), 660, 668](#)  
[riscv\\_rfft\\_64\\_fast\\_init\\_f16 \(C++ function\), 664](#)  
[riscv\\_rfft\\_64\\_fast\\_init\\_f32 \(C++ function\), 666](#)  
[riscv\\_rfft\\_64\\_fast\\_init\\_f64 \(C++ function\), 659, 668](#)  
[riscv\\_rfft\\_f32 \(C++ function\), 659, 663](#)  
[riscv\\_rfft\\_fast\\_f16 \(C++ function\), 659, 663](#)  
[riscv\\_rfft\\_fast\\_f32 \(C++ function\), 659, 663](#)  
[riscv\\_rfft\\_fast\\_f64 \(C++ function\), 659, 663](#)  
[riscv\\_rfft\\_fast\\_init\\_f16 \(C++ function\), 659, 665](#)  
[riscv\\_rfft\\_fast\\_init\\_f32 \(C++ function\), 659, 667](#)  
[riscv\\_rfft\\_fast\\_init\\_f64 \(C++ function\), 660, 669](#)  
[riscv\\_rfft\\_init\\_f32 \(C++ function\), 660, 670](#)  
[riscv\\_rfft\\_init\\_q15 \(C++ function\), 660, 670](#)  
[riscv\\_rfft\\_init\\_q31 \(C++ function\), 660, 671](#)  
[riscv\\_rfft\\_q15 \(C++ function\), 660, 671](#)  
[riscv\\_rfft\\_q31 \(C++ function\), 660, 672](#)  
[riscv\\_rms\\_f16 \(C++ function\), 605, 606](#)  
[riscv\\_rms\\_f32 \(C++ function\), 605, 606](#)  
[riscv\\_rms\\_q15 \(C++ function\), 605, 606](#)  
[riscv\\_rms\\_q31 \(C++ function\), 605, 606](#)  
[riscv\\_rogerstanimoto\\_distance \(C++ function\), 478, 479](#)  
[riscv\\_rotation2quaternion\\_f32 \(C++ function\), 589](#)  
[riscv\\_russellrao\\_distance \(C++ function\), 478, 479](#)  
[riscv\\_scale\\_f16 \(C++ function\), 445](#)  
[riscv\\_scale\\_f32 \(C++ function\), 445, 446](#)  
[riscv\\_scale\\_q15 \(C++ function\), 445, 446](#)  
[riscv\\_scale\\_q31 \(C++ function\), 445, 446](#)  
[riscv\\_scale\\_q7 \(C++ function\), 445, 447](#)  
[riscv\\_selection\\_sort\\_f32 \(C++ function\), 613](#)  
[riscv\\_shift\\_q15 \(C++ function\), 447](#)  
[riscv\\_shift\\_q31 \(C++ function\), 447](#)  
[riscv\\_shift\\_q7 \(C++ function\), 447, 448](#)  
[riscv\\_sin\\_cos\\_f32 \(C++ function\), 470](#)  
[riscv\\_sin\\_cos\\_q31 \(C++ function\), 470, 471](#)  
[riscv\\_sin\\_f32 \(C++ function\), 482, 483](#)  
[riscv\\_sin\\_q15 \(C++ function\), 482, 483](#)  
[riscv\\_sin\\_q31 \(C++ function\), 482, 483](#)  
[riscv\\_softmax\\_q15 \(C++ function\), 721](#)  
[riscv\\_softmax\\_q7 \(C++ function\), 721, 722](#)  
[riscv\\_softmax\\_s8 \(C++ function\), 721, 722](#)  
[riscv\\_softmax\\_u8 \(C++ function\), 721, 722](#)  
[riscv\\_softmax\\_with\\_batch\\_q7 \(C++ function\), 721, 723](#)  
[riscv\\_sokalmichener\\_distance \(C++ function\), 478, 480](#)  
[riscv\\_sokalsneath\\_distance \(C++ function\), 478, 480](#)  
[riscv\\_sort\\_f32 \(C++ function\), 611, 613](#)  
[riscv\\_sort\\_init\\_f32 \(C++ function\), 611, 613](#)  
[riscv\\_spline\\_f32 \(C++ function\), 562, 563](#)  
[riscv\\_spline\\_init\\_f32 \(C++ function\), 562, 563](#)  
[riscv\\_sqrt\\_q15 \(C++ function\), 483, 484](#)  
[riscv\\_sqrt\\_q31 \(C++ function\), 483, 484](#)  
[riscv\\_std\\_f16 \(C++ function\), 607](#)  
[riscv\\_std\\_f32 \(C++ function\), 607](#)  
[riscv\\_std\\_q15 \(C++ function\), 607, 608](#)  
[riscv\\_std\\_q31 \(C++ function\), 607, 608](#)  
[riscv\\_sub\\_f16 \(C++ function\), 448, 449](#)  
[riscv\\_sub\\_f32 \(C++ function\), 448, 449](#)  
[riscv\\_sub\\_q15 \(C++ function\), 448, 449](#)  
[riscv\\_sub\\_q31 \(C++ function\), 448, 449](#)  
[riscv\\_sub\\_q7 \(C++ function\), 448, 450](#)  
[riscv\\_svd\\_f\\_s8 \(C++ function\), 723](#)  
[riscv\\_svm\\_linear\\_init\\_f16 \(C++ function\), 622, 623](#)  
[riscv\\_svm\\_linear\\_init\\_f32 \(C++ function\), 622, 623](#)  
[riscv\\_svm\\_linear\\_predict\\_f16 \(C++ function\), 623](#)  
[riscv\\_svm\\_linear\\_predict\\_f32 \(C++ function\), 623, 624](#)  
[riscv\\_svm\\_polynomial\\_init\\_f16 \(C++ function\), 624](#)  
[riscv\\_svm\\_polynomial\\_init\\_f32 \(C++ function\), 624, 625](#)  
[riscv\\_svm\\_polynomial\\_predict\\_f16 \(C++ function\), 624, 625](#)  
[riscv\\_svm\\_polynomial\\_predict\\_f32 \(C++ function\), 624, 626](#)  
[riscv\\_svm\\_rbf\\_init\\_f16 \(C++ function\), 626](#)

riscv\_svm\_rbf\_init\_f32 (C++ *function*), 626, 627  
 riscv\_svm\_rbf\_predict\_f16 (C++ *function*), 626, 627  
 riscv\_svm\_rbf\_predict\_f32 (C++ *function*), 626, 627  
 riscv\_svm\_sigmoid\_init\_f16 (C++ *function*), 628  
 riscv\_svm\_sigmoid\_init\_f32 (C++ *function*), 628  
 riscv\_svm\_sigmoid\_predict\_f16 (C++ *function*), 628, 629  
 riscv\_svm\_sigmoid\_predict\_f32 (C++ *function*), 628, 629  
 riscv\_var\_f16 (C++ *function*), 608, 609  
 riscv\_var\_f32 (C++ *function*), 608, 609  
 riscv\_var\_q15 (C++ *function*), 608, 609  
 riscv\_var\_q31 (C++ *function*), 608, 609  
 riscv\_vsqr\_t\_f32 (C++ *function*), 483, 484  
 riscv\_vsqr\_t\_q15 (C++ *function*), 483, 485  
 riscv\_vsqr\_t\_q31 (C++ *function*), 483, 485  
 riscv\_weighted\_sum\_f16 (C++ *function*), 622  
 riscv\_weighted\_sum\_f32 (C++ *function*), 622  
 riscv\_xor\_u16 (C++ *function*), 450  
 riscv\_xor\_u32 (C++ *function*), 450  
 riscv\_xor\_u8 (C++ *function*), 450, 451  
 riscv\_yule\_distance (C++ *function*), 478, 480  
 riscvBitRevTable (C++ *member*), 630, 631  
 rv\_csr\_t (C++ *type*), 78  
 rv\_fpu\_t (C++ *type*), 385

## S

SAVE\_FPU\_CONTEXT (C *macro*), 384  
 SAVE\_IRQ\_CSR\_CONTEXT (C *macro*), 370, 371  
 SIP\_SSIIP (C *macro*), 76  
 SIP\_STIP (C *macro*), 76  
 SLEEPVALUE\_SLEEPVALUE (C *macro*), 74  
 SSTATUS32\_SD (C *macro*), 72  
 SSTATUS64\_SD (C *macro*), 72  
 SSTATUS\_FS (C *macro*), 72  
 SSTATUS\_PUM (C *macro*), 72  
 SSTATUS\_SIE (C *macro*), 72  
 SSTATUS\_SPIE (C *macro*), 72  
 SSTATUS\_SPP (C *macro*), 72  
 SSTATUS\_UIE (C *macro*), 72  
 SSTATUS\_UPIE (C *macro*), 72  
 SSTATUS\_XS (C *macro*), 72  
 system\_default\_exception\_handler (C++ *function*), 407  
 SystemCoreClock (C++ *member*), 406  
 SystemCoreClockUpdate (C++ *function*), 406  
 SystemExceptionHandlers (C++ *member*), 409  
 SystemInit (C++ *function*), 406  
 SysTimer (C *macro*), 92

SysTimer\_BASE (C *macro*), 92  
 SysTimer\_MSFRST\_KEY (C *macro*), 92  
 SysTimer\_MSFRST\_Msk (C *macro*), 92  
 SysTimer\_MSIP\_MSIP\_Msk (C *macro*), 92  
 SysTimer\_MSIP\_MSIP\_Pos (C *macro*), 92  
 SysTimer\_MSIP\_Msk (C *macro*), 92  
 SysTimer\_MTIMECTL\_CLKSRC\_Msk (C *macro*), 92  
 SysTimer\_MTIMECTL\_CLKSRC\_Pos (C *macro*), 92  
 SysTimer\_MTIMECTL\_CMPCLREN\_Msk (C *macro*), 92  
 SysTimer\_MTIMECTL\_CMPCLREN\_Pos (C *macro*), 92  
 SysTimer\_MTIMECTL\_Msk (C *macro*), 92  
 SysTimer\_MTIMECTL\_TIMESTOP\_Msk (C *macro*), 92  
 SysTimer\_MTIMECTL\_TIMESTOP\_Pos (C *macro*), 92  
 SysTimer\_MTIMER\_Msk (C *macro*), 92  
 SysTimer\_MTIMERCMP\_Msk (C *macro*), 92  
 SysTimer\_Type (C++ *class*), 92

## T

T\_UINT16\_READ (C++ *member*), 60  
 T\_UINT16\_WRITE (C++ *member*), 60  
 T\_UINT32\_READ (C++ *member*), 60  
 T\_UINT32\_WRITE (C++ *member*), 60  
 twiddleCoef\_1024 (C++ *member*), 630, 633  
 twiddleCoef\_1024\_q15 (C++ *member*), 631, 636  
 twiddleCoef\_1024\_q31 (C++ *member*), 630, 635  
 twiddleCoef\_128 (C++ *member*), 630, 633  
 twiddleCoef\_128\_q15 (C++ *member*), 630, 636  
 twiddleCoef\_128\_q31 (C++ *member*), 630, 634  
 twiddleCoef\_16 (C++ *member*), 630, 633  
 twiddleCoef\_16\_q15 (C++ *member*), 630, 635  
 twiddleCoef\_16\_q31 (C++ *member*), 630, 634  
 twiddleCoef\_2048 (C++ *member*), 630, 634  
 twiddleCoef\_2048\_q15 (C++ *member*), 631, 637  
 twiddleCoef\_2048\_q31 (C++ *member*), 630, 635  
 twiddleCoef\_256 (C++ *member*), 630, 633  
 twiddleCoef\_256\_q15 (C++ *member*), 630, 636  
 twiddleCoef\_256\_q31 (C++ *member*), 630, 635  
 twiddleCoef\_32 (C++ *member*), 630, 633  
 twiddleCoef\_32\_q15 (C++ *member*), 630, 636  
 twiddleCoef\_32\_q31 (C++ *member*), 630, 634  
 twiddleCoef\_4096 (C++ *member*), 630, 634  
 twiddleCoef\_4096\_q15 (C++ *member*), 631, 637  
 twiddleCoef\_4096\_q31 (C++ *member*), 630, 635  
 twiddleCoef\_512 (C++ *member*), 630, 633  
 twiddleCoef\_512\_q15 (C++ *member*), 631, 636  
 twiddleCoef\_512\_q31 (C++ *member*), 630, 635  
 twiddleCoef\_64 (C++ *member*), 630, 633  
 twiddleCoef\_64\_q15 (C++ *member*), 630, 636  
 twiddleCoef\_64\_q31 (C++ *member*), 630, 634  
 twiddleCoefF16\_1024 (C++ *member*), 631, 638

twiddleCoeffF16\_128 (*C++ member*), 631, 637  
twiddleCoeffF16\_16 (*C++ member*), 631, 637  
twiddleCoeffF16\_2048 (*C++ member*), 631, 638  
twiddleCoeffF16\_256 (*C++ member*), 631, 638  
twiddleCoeffF16\_32 (*C++ member*), 631, 637  
twiddleCoeffF16\_4096 (*C++ member*), 631, 638  
twiddleCoeffF16\_512 (*C++ member*), 631, 638  
twiddleCoeffF16\_64 (*C++ member*), 631, 637  
twiddleCoeffF16\_rfft\_1024 (*C++ member*), 631, 638  
twiddleCoeffF16\_rfft\_128 (*C++ member*), 631, 638  
twiddleCoeffF16\_rfft\_2048 (*C++ member*), 631, 638  
twiddleCoeffF16\_rfft\_256 (*C++ member*), 631, 638  
twiddleCoeffF16\_rfft\_32 (*C++ member*), 631, 638  
twiddleCoeffF16\_rfft\_4096 (*C++ member*), 631, 638  
twiddleCoeffF16\_rfft\_512 (*C++ member*), 631, 638  
twiddleCoeffF16\_rfft\_64 (*C++ member*), 631, 638  
twiddleCoeffF64\_1024 (*C++ member*), 630, 632  
twiddleCoeffF64\_128 (*C++ member*), 630, 632  
twiddleCoeffF64\_16 (*C++ member*), 630, 631  
twiddleCoeffF64\_2048 (*C++ member*), 630, 632  
twiddleCoeffF64\_256 (*C++ member*), 630, 632  
twiddleCoeffF64\_32 (*C++ member*), 630, 632  
twiddleCoeffF64\_4096 (*C++ member*), 630, 632  
twiddleCoeffF64\_512 (*C++ member*), 630, 632  
twiddleCoeffF64\_64 (*C++ member*), 630, 632  
TXEVT\_TXEVT (*C macro*), 74

## U

UCODE\_OV (*C macro*), 74  
USE\_INTRINSIC (*C macro*), 712

## V

VM\_MBARE (*C macro*), 76  
VM\_MBB (*C macro*), 76  
VM\_MBBID (*C macro*), 76  
VM\_SV32 (*C macro*), 76  
VM\_SV39 (*C macro*), 76  
VM\_SV48 (*C macro*), 76

## W

Weights\_128 (*C++ member*), 652  
Weights\_2048 (*C++ member*), 652, 653  
Weights\_512 (*C++ member*), 652, 653  
Weights\_8192 (*C++ member*), 652, 653  
WeightsQ31\_128 (*C++ member*), 652, 653  
WeightsQ31\_2048 (*C++ member*), 652, 653

WeightsQ31\_512 (*C++ member*), 652, 653  
WeightsQ31\_8192 (*C++ member*), 652, 653  
WFE\_WFE (*C macro*), 74

## X

XIP, 741