Smart Computing Card SC5

User Guide

File Version V1.05

Release Date 2024.01.10

Preface

Introduction

This file introduces in detail the appearance characteristics, application scenarios, equipment parameters, electrical features, supporting software, use environment, etc. of the SC5 (X) series Smart Computing Card (including demo board). Users and developers can have a more comprehensive and in-depth understanding of the SC5 (X) series Smart Computing Card (including demo board) after reading the file. Equipment users and developers can carry out a series of work such as installation, debugging, deployment and maintenance of the equipment according to this file.

Targeted Readers

This file is mainly targeted at the following personnel:

- SOPHGO FAE and pre-sales engineer
- Developers of eco partners
- R&D engineer and pre-sales engineer of user enterprises

Agreed Symbols, Signs and Terms

The following symbols and signs may appear in this file, and their meanings are as follows:

⚠ Danger	Highly dangerous situation. If not avoided, it may cause casualties or serious injury.
⚠ Precaution	Moderate or low potential danger, which, if not avoided, may cause minor or moderate injury to personnel.
⚠ Attention	Potential risk. The ignorance of this part may cause equipment damage, data loss, performance degradation or unpredictable results.
⚠ Antistatic	Anti-static symbol, indicating electrostatic sensitive equipment or operation.
⚡ Electric Shock Warning	Electric shock warning, which indicates the danger of high voltage. Personnel shall be well protected.
⚷ Key	Keys that can help you solve a problem or save your time

Description	Additional information. Emphasis and supplement to the text.

Modification Record

File Version	Release Date	Description	Hardware	Software
V1.0	2020-03-06	First official release	SC5: V1.2 SC5H: V1.0 SC5-IO: V1.0 SC5+: V1.1	2.0.3
V1.01	2020-05-11	Parameters and errors correction		2.0.3
V1.02	2020-05-25	Add Docker and K8S usage		2.0.3
V1.03	2020-08-11	Add precautions		2.0.3 / 2.1.0
V1.04	2021-07-31	Amendments and		2.4.0
V1.05	2024-01-10	Add server adaptation and modify some descriptions.		2.4.0

Declaration

Contents

1 Safety

1.1 Common Safety Precautions

➔ Wear protective gloves to prevent your hands from being cut by sharp corners when handling the equipment;

➔ When carrying heavy objects, be prepared to bear the load. Keep your back straight and move smoothly, and avoid being crushed or sprained by heavy objects;

➔ The transported equipment shall be placed in a stable place to avoid hurting personnel;

➔ When the temperature difference between outdoor and indoor exceeds 15 ℃, let the equipment stand for 8 hours indoor before installation;

➔ If the temperature difference cannot be determined, it is recommended to place the equipment in the machine room for one night before installation, so as to ensure that there is no condensation or abnormal power on of the equipment;

➔ Avoid screws falling into the rack or box to prevent short circuit;

➔ If the product is an active cooling card, pay attention to the physical hazard protection of moving components (cooling fan);

➔ Please use tools to clean the dust of the active card fan to avoid finger scratching. If you need to clean the dust inside the board, please contact our FAE and other professionals for operation.

1.2 Electrical Safety Precautions

➔ Please wear anti-static gloves, wristbands and clothes to carry out installation;

➔ When the equipment is powered on, do not wear anti-static wristband;

➔ When handling and transporting components, special anti-static packaging must be used;

Power Safety:

➔ This product series does not support hot plug operation;

➔ Confirm whether the input voltage is within the rated voltage range of the equipment;

➔ Please use the supporting power cord. Do not use the cord for other equipment;

→ If the DC power supply system is used for power supply, please ensure that there is reinforced insulation or double insulation between DC power supply and AC power supply.

1.3 Product Packaging and Labeling

1.3.1 Naming Rule

S：SOPHON

C：Card

5：Based on the third generation intelligent vision deep learning processor BM1684

SOPHON is a sub-brand and product series name launched by SOPHON Technology for the field of deep learning.

SC5	It refers to the full-height and full-length Smart Computing Card equipped with single processor BM1684, demo board and active fan cooling.
SC5-IO	It refers to an I / O expansion card dock used with SC5 demo board, which can convert PCIe working mode to SoC working mode and expand the corresponding I/O peripheral interface (Part No. = SC5A111B0-S).
SC5H	It refers to the half height and half-length Smart Computing Card equipped with single processor BM1684 and active fan cooling.
SC5+	It refers to the half height and half-length Smart Computing Card equipped with three BM1684 and passive heat dissipation.

1.3.2 Product Label

As the example shown in the left, SC5 (X) board products are attached with product label identification, which is posted in the label groove of the back to indicate the PN code, SN code and relevant information. Please pay attention to protect the label from damage.

1.3.3 Product Packaging and Accessories

Accessory	SC5	SC5H	SC5-IO	SC5+
User Guide	✓	✓	✓	✓
Warranty Card	✓	✓	✓	✓

Certificate	✓	✓	✓	✓
Half Height Handle	——	✓	——	✓
Full Height Handle	✓	✓	——	✓
Screwdriver	——	✓	——	✓
Packing Box	✓	✓	✓	✓
4G Antenna	——	——	✓	——

Table 1-1 SC5 (X) Product Packaging and Accessories

2 Environmental Protection

2.1 RoHS Certificate

This product meets the requirements of RoHS 2.0 Directive (EU) 2015/863 and has obtained the certificate.

2.2 Toxic and Harmful Substances

This product meets the provisions of Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) No. 1907/2006, and has obtained the certificate according to the limit requirements of SVHC list (205 SVHC) issued by ECHA on January 16, 2020.

3 Product Introduction

3.1 Summary

⠀⠀⠀⠀⠀Sophon SC5 (X) series products are the latest generation of Smart Computing Card series launched by SOPHGO Technologies. The product models included are as follows: **In order to express the difference, "SC5 (X)" in this manual refers to SC5/SC5H/ SC5-IO/SC5 + and other series of accelerator cards and expansion products developed based on BM1684 Deep learning processor. The "SC5" or "SC5 accelerator card (demo board)" separately pointed out in this manual refers to the full height full-length demo board products with a single processor.**

Name	Part No.	Description	Targeted Personnel
Smart Computing Card SC5 (demo board)	SC5A111B0	Single width and full height active heat dissipation, EVB demo board 1*BM1684 processor	Developers, algorithm migration, function evaluation, scientific research institutes, etc.
SC5-IO I/O Expansion Card dock	SC5A111B0-S	Desktop expansion module, used with SC5 demo board, converts PCIe mode to SOC working mode	Developers, algorithm migration, function evaluation, scientific research institutes, etc.
Smart Computing Card SC5+	SC5P321B0	Single width, half height, half-length, passive heat dissipation. 3*BM1684 processor	Data center, Deep learning server
Smart Computing Card SC5H	SC5A121B0	Single width, half height, half-length, active heat dissipation. 1*BM1684 processor	Edge computing, industrial computer, PC

Table 3-1 Model List of SC5 (X) Series Products

3.1.1 Smart Computing Card SC5 (Demo Board)

⠀⠀⠀⠀⠀The overall product design of SC5 is in the form of demo board, equipped with a BM1684 high-performance computing processor, standard PCIE 3.0 X16 interface and active cooling fan design. It can adapt to the standard PC development and test environment and

various high-performance servers, and is compatible with all kinds of X86, ARM, Phytium, Shenwei PROCESSORPROCESSOR and other mainboard architectures.

The SC5 Smart Computing Card (demo board) is equipped with a BM1684 processor, which can support 32-channel HD video hardware decoding and over 16-channel video analysis. At the same time, it is equipped with RESET button and DEBUG interface. Its device driver and development test environment are consistent with SC5 + Smart Computing Card and support synchronous upgrading. Developers can easily verify, test, transplant and develop various index parameters of BM1684 processor at the lowest cost.

With the corresponding I/O expansion card dock, SC5 Smart Computing Card (demo board) can work in SoC mode and expand a variety of peripheral test interfaces, including USB, TCP/IP, SATA, 4G, etc.

The smart card SC5 supports the whole process of 16 channels of 1080P HD video processing and FP32/INT8 mixed precision reasoning.

BM1684 is an application specific integrated circuit (ASIC) acceleration solution for deep learning, mainly for the acceleration of network reasoning. On the acceleration platform based on BM1684, the performance of all mainstream networks, such as CNN/RNN/DNN, can be greatly improved.

BM1684 single processor has 64 neuron processor units (NPU), and each NPU contains 16 execution units (EU). BM1684 can support mixed reasoning with various accuracies such as FP32/INT8. When Winograd is enabled, the computing power of BM1684 further rises to 35.2TOPs. Deeply optimized NPU is a powerful scheduling engine, which can provide very high bandwidth data supply for neuron computing core. 32MB on- processor storage provides excellent programming flexibility for performance optimization and data reuse.

3.1.2 Smart Computing Card SC5H

The SC5H Smart Computing Card (half height and half-length standard card) is equipped with a BM1684 processor. Its computing performance is completely consistent with that of SC5. It can support 32-channel HD video hardware decoding and more than 16-channel video analysis. Its device driver and development test environment are consistent with deep leaning computing accelerators such as SC5 and SC5 +, and support synchronous upgrading.

The shell size of SC5H is smaller, which can be adapted to 2U height standard server or industrial computer equipment, and the application scenario is more extensive. The side suction fan design and ingenious air duct design can meet the needs of intensive deployment.

3.1.3 Smart Computing Card SC5+

SOPHON SC5+ is the latest generation of Smart Computing Card launched by SOPHGO Technologies. It is equipped with three BM1684 high-performance computing processor, standard PCIe 3.0 interface and fanless design to better adapt to various high-performance servers and x86 mainboard architecture.

SC5+ can provide up to 105.6T INT8 computing power (Winograd Enable) and 6.6 t FP32 computing power, supporting high-precision calculation.

SC5+ can be used in various scenarios including face recognition, video structuring, security monitoring, deepl learning and machine vision. Users can speed up the calculation of a variety of CNN/RNN/DNN and other neural network models through the tool chain software provided by computer.

3.2 Product Appearance

The appearance of each model is as follows:

Smart Computing Card SC5 (demo board)	SC5-IO Expansion Card Dock	Smart Computing Card SC5H (half height and half-length standard)	Smart Computing Card SC5 + (Half height, half length and single width)

<p align="center">Figure 3-1 SC5（X）Series Product Appearance</p>

3.3 System Framework

The system framework diagram of Smart Computing Card (demo board) SC5/SC5H is shown in Figure 3-2.
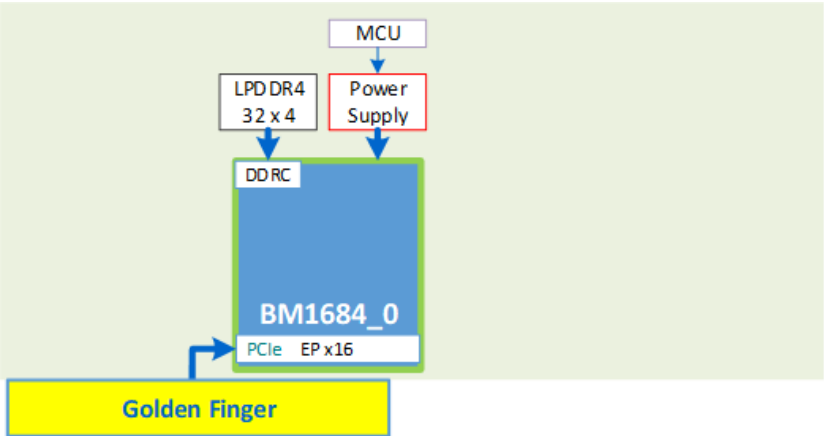
Figure 3-2 SC5/SC5H System Framework*

* As a demo board product, SC5 supports PCIe EP + RC mode. As a PCIe slave device, the data channel is actually PCIe X8.

The system framework diagram of Smart Computing Card SC5+ is shown in Figure 3-3.
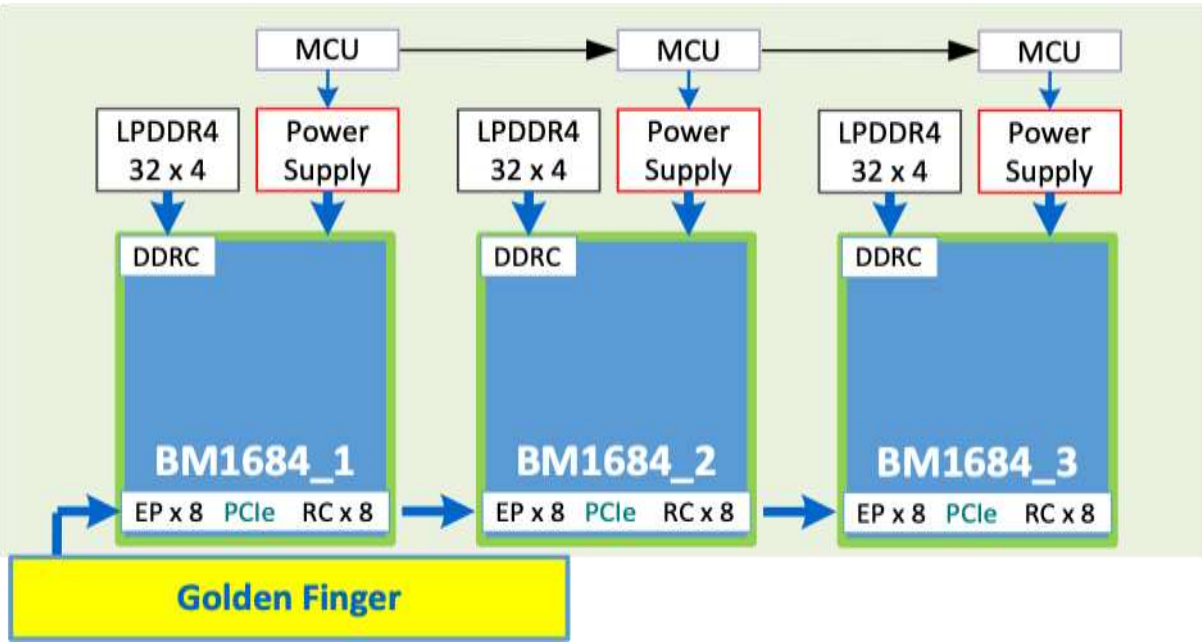
Figure 3-3 SC5+System Framework Diagram

4 Product Feature

4.1 Performance Feature

SC5 (X) series Smart Computing Card (or demo board) supports FP32, INT8 and other precision computing at the same time. The product series has a variety of forms of computing board products, which adopt one or more high-performance BM1684 Deeplearing processors respectively. As the third generation of mass-produced Deeplearing card, it ranks forefront in the world in terms of its reasonable, stable and reliable design.

The processor utilization of single BM1684 Deeplearing processor significantly exceeds other Deeplearing processor of the same level. It supports FP32 and INT8 precision calculations or mixed precision calculations, 32-channel hardware HD video decoding and H.264 and H.265 video formats. It is suitable for the data input of video stream and image stream of all kinds of ultra-HD network cameras with a wide resolution range.

The supporting BMNNSDK2 development kit od SC5 (X) series products is easy to use. It has a unified interface and supports a variety of Python/C++ development interfaces. The NNTC tool chain supports a variety of algorithm frameworks, such as Caffe/Tensorflow/Pytoch/Mxnet/Darknet.

4.2 Maintainability and Management Features

- Supports upgrade management. MCU Firmware of BM1684 processor can be upgraded online, which is convenient for daily maintenance;
- The SN of the board is directly stored in MCU, which can be easily read locally and remotely through SDK, and can also be queried and obtained on BM-SMI management tool;
- SN, SDK and driver version, temperature, power consumption, TPU occupation, memory occupation and other information of the board can be obtained through BM-SMI management tool;

4.3 Typical Application Scenario

- ❖ SC5 single- processor demo board accelerator card:

SC5 is suitable for developers, algorithm application partners and integrators to

understand the application and demo of BM1684 processor, and carry out various

function

development, testing, verification and migration of typical algorithm models based on the demo board.

The SC5 demo board, combined with the corresponding SC5-IO I/O dock, can be easily applied in the computer classroom of scientific research institutions to construct a fast and effective deep learning scientific experimental environment via full desktop deployment. The dual- mode split design of PCIe and SOC enables developers to deploy demo environment at the optimal hardware cost under the ordinary PC configuration environment.

❖ SC5H single- processor standard accelerator card:

The SC5H is equipped with a BM1684 processor and adopts the standard PCIe card size of half height and half length. At the same time, it is equipped with a side suction fan, which can well adapt to complex working conditions, such as embedded PC, Mini PC, etc.

SC5H is mainly used for edge deep learning computing and analysis, such as traffic, urban management, smart community, industrial detection and other scenarios requiring pre-deep leanring computing power. It can also be mixed with other forms of boards such as SC5+ for the same computing platform.

❖ SC5+ acceleration card III:

SC5+ accelerator card has strong performance and can be used with various servers and industrial computer platforms. It can form an intelligent analysis cluster through the superposition of multiple boards and cards, which is used for large-scale multi-channel deep learning computing and analysis. It can be applied in various scenarios such as face recognition, video structuring, industrial vision detection and analysis.

5 Product Specification

5.1 Basic Specification

Features		SC5	SC5H	SC5+	SC5-I/O
TPU Core Architecture		SOPHON			N/A
NPU Core Number		64 Cores		3x 64 Cores	N/A
Data processing		Supports FP32/INT8			
Processor		ARM 8 Core A53 @2.3GHz		3x ARM 8 Core A53 @2.3GHz	N/A
System/Power Interface	Data Channel	EP PCIE X8 RC PCIE X8	PCIE X16	PCIE X8	PCIE X2
	Working Mode	EP+RC	EP	EP	SOC Scalable
	Physical/Power Interface	PCI Express 3.0 X16 (Physical Interface)			12VDC-Jack
Memory	Standard Configuration	12GB		3x 12GB	N/A
	Scalable	16GB		3x 16GB	
	Memory	LPDDR4x, 4000MHz			
Power Consumption		30W MAX		75W MAX	No Load: 6W Loaded: 30W
Heat Dissipation		Active		Passive	N/A
Video Decoding	Capacity	H.264 & H.265: 1080P @960fps		H.264 & H.265: 1080P @2880fps	N/A

	Resolution	CIF/D1/720P/1080P/4K (3840×2160) /8K (8192×4096)/8192x8192[1]		
Video Encoding	Capacity	H.264 & H.265: 1080P @50fps	H.264 & H.265: 1080P@150fps	N/A
	Resolution	CIF/D1/720P/1080P/4K (3840×2160)[2]		N/A
JPEG Transcoding Capacity （1080P to CIF）		Max.18 channel	Max. 54 channel	N/A
JPEG Decoding and Encoding	JPEG Codec	480 pages/s @1080P	1440 pages/s @1080P	N/A
	Max. Resolution	32768 x 32768		

Features		SC5	SC5H	SC5+	SC5-I/O
SMBUS		Support SMBUS docking with X86 server[3]			N/A
Container Support		Support mainstream Docker Technology			N/A
Virtualization Support		Support KVM, Xen			N/A
Working Status Display		N/A	LED x1	LED x1	LED x3 (Power/ Hard Disk/Status)
I/O Scalable	SD-Card	1[4]	N/A	N/A	N/A
	RESET Button	1	N/A	N/A	N/A
	RJ45	2x 1000BASE-T[5]	N/A	N/A	N/A
	USB	N/A	N/A	N/A	4
	SATA	N/A	N/A	N/A	1
	4G/LTE	N/A	N/A	N/A	1[6]

[1] Video decoding of 8K and above resolution only supports non-real-time decoding;

[2] 4K and above resolution video coding only supports non-real-time coding;

[3] All functions can be realized only when SMBUS docking development is carried out on the host side

[4] It needs to be used with SC5-IO expansion card dock

[5] Same as 4

[6] Built in module, SIM card needs to be equipped separately

	Micro USB	1[7]	N/A	N/A	N/A
Deep Learning Framework		Caffe/TensorFlow/Pytorch/Mxnet/Darknet			N/A
Operation System Support		Ubuntu/CentOS/Debian			N/A
Compatibility		Compatible with mainstream x86 and ARM servers			N/A
Length x Height x Width（with bracket）	200 x 111.2 x 19.8mm	169.1 x 68.9 x 19mm	169.1x 68.9 x 19.5mm		206 x 28.5 x 59.5mm

Table 5-1 Basic Specifications and Parameters of SC5 (X) Series

5.2 Hardware Environment

SC5 (X) series boards adopt standard PCIe X16 Gen3.0 interface. The host side needs to provide a standard X16 slot or at least X8 in X16 PCIE slot (the tail is required to be open). the data transmission bandwidth between the board device and the host side will decrease if this condition is not met (such as PCIE X4 channel), which will affect the overall board performance.

Environment Indicator	Specification
Working Temperature	0°C~55°C (32°F ~ +131°F)
Storage Temperature	-40°C~+75°C (-40°F ~ +167°F)
Working Humidity	5%RH~90%RH (non-condensing)
Storage Humidity	5%RH~90%RH (non-condensing)
Altitude	Less than 3000m. When it is used above 900m, the maximum temperature decreases by 1°C for every 300m increase in altitude.

Table 5-2 SC5（X）Series Hardware Environment Condition

---

[7] For the bottom debugging port of reserve, please consult FAE for details

5.3 System Clock Requirements

The SC5 (X) series accelerator card complies with the standard PCIe standard card protocol (PCI Express® Card Electromechanical Specification Revision 3.0). The whole card only needs to provide standard PCIe 3.0 (downward compatible with 2.0 and 1.0) differential clock, and the signal quality meets the PCIe specification.

5.4 Hot Plug

SC5 (X) series accelerator cards do not support hot plug.

5.5 Power Management

Power Domain	Description	Note
PCIE12V	12V power is provided by the mainboard through PCIe slot. The standard specification is Max. 5.5A load capacity.	Needed by SC5/SC5H/SC5+
PCIE3.3V	Provided by the mainboard and PCIe slot. The standard specification is Max. 3A load capacity.	Only needed by SC5+

Table 5-3 SC5 (X) Series Power Management

5.6 Heat Dissipation Specification

5.6.1 Heat dissipation requirements (air volume)

When SC5 (X) series accelerator card is applied to server, PC, industrial computer, embedded PC and other hardware equipment, the hardware equipment used for carrying shall meet the following design and heat dissipation requirements.

Both SC5 and SC5H acceleration cards adopt active heat dissipation design. The use of the equipment has no special requirements on the air volume at the host side, but it needs to comply with the requirements of the working temperature range.

The SC5 + accelerator card adopts passive heat dissipation design, and the host side shall meet the design heat dissipation requirements described in Table 5-1 below. In practical application, it is recommended to install it into the standard server and adjust the speed of the card to ensure heat dissipation. It is not recommended to install it in ordinary

PC/industrial control computer. If it needs to be used, please add air guide structure and install high air volume and air pressure fan at the card inlet.

Average temperature of air inlet/ °C	Minimum wind speed required by card air inlet /CFM	Pressure drop/inch H2O
55	11	0.95
50	8.5	0.70
45	6.7	0.51
40	5.3	0.35
35	4.3	0.25
30	3.5	0.18

Table 5-4 Air Volume and Heat Dissipation Requirements of SC5 + Accelerator Card on the Host Side

5.6.2 Overtemperature protection mechanism

When the junction temperature of BM1684 is higher than 85 °C and the hysteresisis 5 °C, BM1684 processor will automatically reduce the frequency to 80% of the max frequency.

When the junction temperature of BM1684 is higher than 90 °C and the hysteresis is 5 °C, BM1684 processor will automatically reduce the frequency to the min frequency.

When the junction temperature of BM1684 is higher than 95 °C, the system will automatic power off.

Hysteresis of 5 °C means that when the junction temperature drops from above 90 °C to 85 °C, the working frequency of BM1684 will return to 80% of the maximum; When the junction temperature drops from above 85 °C to 80 °C, the working frequency of BM1684 will return to the maximum.

6 I/O Definition

6.1 Definition of Signal Pin of PCIe x16

The SC5 (X) series accelerator card complies with the standard PCIe standard card protocol (PCI Express® Card Electromechanical Specification Revision 3.0), which provides PCIe X16 golden finger physical interface to the outside. The actual data transmission is X8 or X16. Taking SC5 + as an example, its main data signals are: a group of PCIe 3.0 x8, with a maximum rate of 8Gbps/lane, which is used for interactive transmission of business data; a group of SMBUS (which can be connected with the host side through SMBUS protocol. For details, please refer to the definition of the host side), with a maximum speed of 100Kbps, which is used for SC5+ out-of-band management unit monitoring. The detailed signal distribution of SC5 (X) series accelerator card is shown in Table 6-1 ~ Table 6-5.

Table 6-1 SC5 (X) Gold Finger Pin Definition (Mechanical Key)

No.		Pin Name	Description	Process Method
Side B	1	+12V	12V Power	12V Power
	2	+12V	12V Power	
	3	+12V	12V Power	
	4	GND	Ground	Grounding
	5	SMCLK	SMBUS Clock	Input
	6	SMDAT	SMBUS Data	Input
	7	GND	Ground	Grounding
	8	+3.3V	3.3V Power	3.3V Power
	9	JTAG1	JTAG Interface TRST Signal	Suspend in Midair
	10	3.3Vaux	3.3V Auxiliary Power	
	11	WAKE#	Link Reactivated Signal	
No.		Pin Name	Description	Process Method
Side A	1	PRSNT1#	Hot Plug Presence Detection 1#	Connected to B48, in place detection

	2	+12V	12V Power	12V Power
	3	+12V	12V Power	
	4	GND	Ground	Grounding
	5	JTAG2	JTAG Interface TCK Signal	Suspend in Midair
	6	JTAG3	JTAG Interface TDI Signal	
	7	JTAG4	JTAG Interface TDO Signal	
	8	JTAG5	JTAG Interface TMS Signal	
	9	+3.3V	3.3V Power	3.3V Power
	10	+3.3V	3.3V Power	
	11	PERST#	Basic RESET	Whole board reset

Table 6-2 SC5(X) Gold Finger Pin Definition (End of the x1 connector)

No.		Pin Name	Description	Process Method
Side B	12	RSVD	Reserved	Suspend in Midair
	13	GND	Ground	Grounding
	14	PETp0	Send Differential Pair PCIe_ TX_ zero	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	15	PETn0		
	16	GND	Ground	Grounding
	17	PRSNT2#	Hot Plug Presence Detection 2#	Suspend in Midair
	18	GND	Ground	Grounding
Side A	12	GND	Ground	Grounding
	13	REFCLK+	Differential Clock	The baseboard provides 100M PCIe

	14	REFCLK-		differential clock, supports 3.0/2.0/1.0 and SSC
	15	GND	Ground	Grounding
	16	PERp0	Receive Differential Pair PCIe_RX_0	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	17	PERn0		
	18	GND	Ground	Grounding

Table 6-3 SC5(X) Gold Finger Pin Definition (End of the x4 connector)

No.		Pin Name	Description	Process Method
Side B	19	PETp1	Send Differential Pair PCIe_TX_1	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	20	PETn1		
	21	GND	Ground	Grounding
	22	GND	Ground	Grounding
	23	PETp2	Send Differential Pair PCIe_TX_2	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	24	PETn2		
	25	GND	Ground	Grounding
	26	GND	Ground	Grounding
	27	PETp3	Send Differential Pair PCIe_TX_3	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	28	PETn3		
	29	GND	Ground	Grounding
	30	RSVD	Reserved	Suspend in Midair
	31	PRSNT2#	Hot Plug Presence Detection 2#	Suspend in Midair
	32	GND	Ground	Grounding
Side A	19	RSVD	Suspend in Midair	
	20	GND	Ground	Grounding
	21	PERp1	Receive Differential Pair PCIe_RX_1	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	22	PERn1		

No.		Pin Name	Description	Process Method
	23	GND	Ground	Grounding
	24	GND	Ground	Grounding
	25	PERp2	Receive Differential Pair PCIe_RX_2	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	26	PERn2		
	27	GND	Ground	Grounding
	28	GND	Ground	Grounding
	29	PERp3	Receive Differential Pair PCIe_RX_3	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	30	PERn3		
	31	GND	Ground	Grounding
	32	RSVD	Reserved	Suspend in Midair

Table 6-4 SC5(X) Gold Finger Pin Definition (End of the x8 connector)

No.		Pin Name	Description	Process Method
Side B	33	PETp4	Send Differential Pair PCIe_TX_4	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	34	PETn4		
	35	GND	Ground	Grounding
	36	GND	Ground	Grounding
	37	PETp5	Send Differential Pair PCIe_TX_5	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	38	PETn5		
	39	GND	Ground	Grounding

	40	GND	Ground	Grounding
	41	PETp6	Send Differential Pair PCIe_TX_6	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	42	PETn6		
	43	GND	Ground	Grounding
	44	GND	Ground	Grounding
	45	PETp7	Send Differential Pair PCIe_TX_7	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	46	PETn7		
	47	GND	Ground	Grounding
	48	PRSNT2#	Hot Plug Presence Detection 2#	SC5, SC5H suspend in midair SC5+ connect with A1
	49	GND	Ground	Grounding
Side A	33	RSVD	Reserved	Suspend in Midair
	34	GND	Ground	Grounding
	35	PERp4	Receive Differential Pair PCIe_RX_4	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	36	PERn4		
	37	GND	Ground	Grounding
	38	GND	Ground	Grounding
	39	PERp5	Receive Differential Pair PCIe_RX_5	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	40	PERn5		
	41	GND	Ground	Grounding

	42	GND	Ground	Grounding
	43	PERp6	Receive Differential Pair PCIe_RX_6	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	44	PERn6		
	45	GND	Ground	Grounding
	46	GND	Ground	Grounding
	47	PERp7	Receive Differential Pair PCIe_RX_7	SC5, SC5H connect with BM1684 SC5+ connect with the first BM1684
	48	PERn7		
	49	GND	Ground	Grounding

Table 6-5 SC5(X) Gold Finger Pin Definition (End of the x16 connector)

No.		Pin Name	Description	Process Method
Side B	50	PETp8	Send Differential Pair PCIe_TX_8	SC5, SC5H connect with BM1684 SC5+ suspend in midair
	51	PETn8		
	52	GND	Ground	Grounding
	53	GND	Ground	Grounding
	54	PETp9	Send Differential Pair PCIe_TX_9	SC5, SC5H connect with BM1684 SC5+ suspend in midair
	55	PETn9		
	56	GND	Ground	Grounding
	57	GND	Ground	Grounding
	58	PETp10	Send Differential Pair	SC5, SC5H connect with BM1684

59	PETn10	PCIe_TX_10	SC5+ suspend in midair
60	GND	Ground	Grounding
61	GND	Ground	Grounding
62	PETp11	Send Differential Pair PCIe_TX_11	SC5, SC5H connect with BM1684 SC5+ suspend in midair
63	PETn11		
64	GND	Ground	Grounding
65	GND	Ground	Grounding
66	PETp12	Send Differential Pair PCIe_TX_12	SC5, SC5H connect with BM1684 SC5+ suspend in midair
67	PETn12		
68	GND	Ground	Grounding
69	GND	Ground	Grounding
70	PETp13	Send Differential Pair PCIe_TX_13	SC5, SC5H connect with BM1684 SC5+ suspend in midair
71	PETn13		
72	GND	Ground	Grounding
73	GND	Ground	Grounding
74	PETp14	Send Differential Pair PCIe_TX_14	SC5, SC5H connect with BM1684 SC5+ suspend in midair
75	PETn14		
76	GND	Ground	Grounding
77	GND	Ground	Grounding

	78	PETp15	Send Differential Pair PCIe_TX_15	SC5, SC5H connect with BM1684 SC5+ suspend in midair
	79	PETn15		
	80	GND	Ground	Grounding
	81	PRSNT2#	Hot Plug Presence Detection 2#	SC5, SC5H suspend in midair SC5+ connect with A1
	82	RSVD	Reserved	Suspend in Midair

No.		Pin Name	Description	Process Method
Side A	50	RSVD	Reserved	Suspend in Midair
	51	GND	Ground	Grounding
	52	PERp8	Receive Differential Pair PCIe_RX_8	SC5, SC5H connect with BM1684 SC5+ suspend in midair
	53	PERn8		
	54	GND	Ground	Grounding
	55	GND	Ground	Grounding
	56	PERp9	Receive Differential Pair PCIe_RX_9	SC5, SC5H connect with BM1684 SC5+ suspend in midair
	57	PERn9		
	58	GND	Ground	Grounding
	59	GND	Ground	Grounding
	60	PERp10	Receive Differential Pair PCIe_RX_10	SC5, SC5H connect with BM1684 SC5+ suspend in midair
	61	PERn10		
	62	GND	Ground	Grounding
	63	GND	Ground	Grounding
	64	PERp11	Receive Differential Pair PCIe_RX_11	SC5, SC5H connect with BM1684 SC5+ suspend in midair
	65	PERn11		
	66	GND	Ground	Grounding
	67	GND	Ground	Grounding
	68	PERp12		

69	PERn12	Receive Differential Pair PCIe_RX_12	SC5, SC5H connect with BM1684 SC5+ suspend in midair
70	GND	Ground	Grounding
71	GND	Ground	Grounding
72	PERp13	Receive Differential Pair PCIe_RX_13	SC5, SC5H connect with BM1684 SC5+ suspend in midair
73	PERn13		
74	GND	Ground	Grounding
75	GND	Ground	Grounding
76	PERp14	Receive Differential Pair PCIe_RX_14	SC5, SC5H connect with BM1684 SC5+ suspend in midair
77	PERn14		
78	GND	Ground	Grounding
79	GND	Ground	Grounding
80	PERp15	Receive Differential Pair PCIe_RX_15	SC5, SC5H connect with BM1684 SC5+ suspend in midair
81	PERn15		
82	GND	Ground	Grounding

6.2 Definition of Other I/O

6.2.1 RJ45 Interface

There are two groups of RJ-45 interfaces on SC5 accelerator card (demo board), which can be used when connected with SC5-IO extension docker. The definition of IO is as follows:

No.	Pin Name	Net Name	Description	Process Method
1	TX0+	TX0+	The first positive sending differential amplifiers of network signals	Connected to network transformer
2	TX0-	TX0-	The first negative sending differential amplifiers of network signals	Connected to network transformer
3	RX0+	RX0+	The first positive receiving differential amplifiers of network signals	Connected to network transformer
4	TX1+	TX1+	The second positive sending differential amplifiers of network signals	Connected to network transformer
5	TX1-	TX1-	The second negative sending differential amplifiers of network signals	Connected to network transformer
6	RX0-	RX0-	The first negative receiving differential amplifiers of network signals	Connected to network transformer
7	RX1+	RX1+	The second positive receiving differential amplifiers of network signals	Connected to network transformer

8	RX1-	RX1-	The second negative receiving differential amplifiers of network signals	Connected to network transformer
9	LED1-	LED1-	Negative yellow operation indicator light	Connect 470R to ground
10	LED1+	LED1+	Positive yellow operation indicator light	Connect PHY_RTL8211
11	LED2-	LED2-	Negative green operation indicator light	Connect 470R to ground
12	LED2+	LED2+	Positive yellow operation indicator light	Connect PHY_RTL8211
13	SHELD1	GND	Ground	Connected to shield ground
14	SHELD2	GND	Ground	Connected to shield ground

Table 6-6 SC5 RJ45 Interface Definition

6.2.2 SD Card Interface

There is a group of SD card storage interfaces on SC5 accelerator card (demo card), which can be used when connected with SC5-IO extension dock. The definition of interface is as follows:

No.	Pin Name	Net Name	Description	Process Method
1	DAT2	SDIO_DAT2	SD Data Signal	Connect to BM1684 for data transmission
2	CD/DAT3	SDIO_DAT3	SD data signal	Connect to bm1684 for data transmission
3	CMD	SDIO_CMD	Command	CMD signal is pulled up to 18V to connect to BM1684
4	VDD	SD_PWER_3V3	Power	Connected to 3.3V power supply after passing through magnetic beads
5	CLOCK	SDIO_CLK	Clock	Connect to BM1684. The clock is provided by BM1684
6	VSS	GND	Ground	Grounding
7	DAT0	SDIO_DAT0	SD Data Signal	Connect to BM1684 for data transmission
8	DAT1	SDIO_DAT1	SD Data Signal	Connect to BM1684 for data transmission
9	CD	SDIO_CD_X	SD Bit Signal	Connect BM1684 to judge whether the SD card is in place
10	GND1	GND	Ground	Grounding
11	GND2	GND	Ground	Grounding
12	GND3	GND	Ground	Grounding
13	GND4	GND	Ground	Grounding
14	NC1	NC	NC	NC
15	NC2	NC	NC	NC

Table 6-7 SC5 SD Card Interface Definition

6.2.3 RESET Button

On the SC5 accelerator card (demo board), there is a RESET button, which is located at the top of the rear.

No.	Pin Name	Net Name	Description	Process Method
1	RESET	RESET	System RESET button	Press bounce to make a system reset

Table 6-8 SC5 RESET Button Definition

6.3 I/O Transfer Module SC5-IO

SC5-IO expansion board is for SC5 interface expansion. It is connected with SC5 through PCIe x16 slot to provide power for SC5. 100MHz Reference Clock can make SC5 separate from PC to realize SoC working mode. It can add USB3.0, USB 2.0, LTE 4G, SATA, RS-485 and other extended interface to SC5, expanding the application of SC5. The specifications are shown in the table below.

Features	Specification
Form	Side-plug PCIe slot expansion board
Power Supply	DC 12V, Max 6A
Power Consumption	Standby power consumption 3W, maximum working power consumption 30W Provide 35W power supply for SC5 single- processor demo board at the same time
USB 3.0	4 USB 3.1 Gen1 5G A-type socket, used with USB2.0
USB 2.0	4 USB2.0 Hi-speed 480Mb interface, used with USB3.0
LTE Antenna	Standard SMA interface terminal
LTE 4G Module	Support standard mini PCIe LTE 4G module. SC5-IO supports NANO SIM card holder.
SATA	Support SATA 6G and can be externally connected with 2.5-inch standard external machinery or SSD HD.
Connect to SC5	PCI Express 3.0 x16 slot, the actual maximum bandwidth is X2 Gen3 8G, and a 6pin 2.54mm spacing test pin is reserved for conversion to RS-485 interface.
RS-485	Two groups, half duplex, with automatic direction detection and a maximum rate of 500Kbps. If enabled, connect the reserved 6pin 2.54 pin with the corresponding position of SC5.
Heat Dissipation	Passive heat dissipation.

PCB Size	195.87mm X 50.8mm
Status Indication	Three green LED status indicators, one for power supply, one for hard disk status and one for 4G status

Table 6-9 SC5-IO Product specification parameters

As a unified component in one system, SC5-IO and SC5 are powered by 12V power adapter through DC-Jack of SC5-IO extension dock. It is designed to prevent mis-operation and detect PRSNT_IN0 and PRSNT_IN1 through PCIe in-place detection. If SC5-IO and SC5 are not combined together, or the contact between PCIe Golden Finger and PCIe Slot is not in place, the whole system cannot be powered on. The system is equipped with a green power status indicator LED. When the LED is on, it indicates successful power on; when the LED is off, it indicates failed power on.

7 Installation and Deployment

7.1 Hardware Installation

7.1.1 PCIe mode installation steps

SC5 (X) series board is standard PCIe card. Please install and use it according to the following steps.

Please follow the steps below to install the SC5 (X) series PCIe card:

1. Turn off and unplug the computer.
2. Open the computer chassis cover and/or the expansion slot cover.
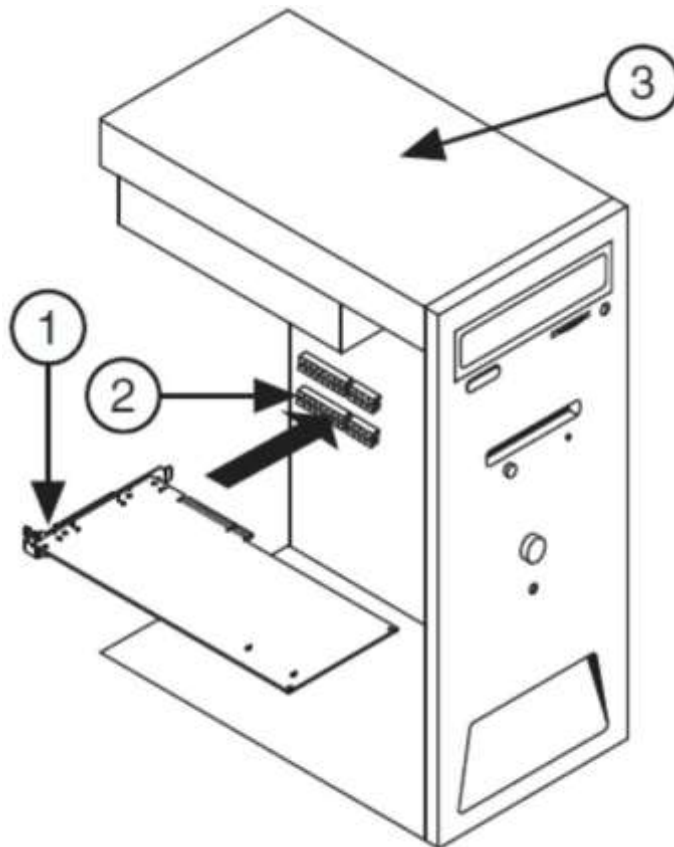3. Touch the metal part of the computer to release static electricity.

Figure 7-1 The Installation of SC5(X) PCIe Board: (1) SC5 (X) series board; (2) PCIe

system slot; (3) Computer with PCIe slot

4. Insert the card into the available PCIe system slot. Please gently shake the device to insert the board into place. Do not install the PCIe card in the PCI slot and vice versa. PCIe card supports up insertion, that is, devices with low width can be inserted into PCIe slots with high width.
5. Fix the mounting bracket on the rear panel guide rail of the computer.

6. For SC5 (X) series PCIe cards, according to the PCIe protocol, the PCIe connector provides 75W power supply without ATX connector.
7. Close the computer chassis cover.
8. Plug in and turn on the computer.

Note: Before installing the board and entering the server, please unplug the 220V power plug of the server. There may be electricity leakage in the PCIe slot when a few servers are connected to the 220V power supply,

7.1.2 Equipment identification confirmation

The demo board has 4 built-in device status indication LEDs. After the equipment is installed, do not cover its back cover. Please power on, and observe the LED status from the upper side. As shown in the figure below, if the four LED lights are blue, blue, green and red respectively, SC5 is powered on successfully.

Figure 7-2 SC5 Demo Board Identification and Confirmation

For SC5 (X) series accelerator card, after installation, you can enter the Linux operating system and enter the command: lspci │ grep 1684.

At this time, you will see the BM1684 device placed in the computer, that is, the corresponding accelerator card.

7.2 Operation System Environment

Various operating system environments (Ubuntu/CentOS/Debian/ NeoKylin) shall meet the following requirements:

7.2.1 Ubuntu

If you need to install Ubuntu, the following version is recommended:

http://old-releases.ubuntu.com/releases/16.04.3/:

– Desktop edition: ubuntu-16.04.3-desktop-amd64.iso

– Server edition: ubuntu-16.04.3-server-amd64.iso

7.2.2 CentOS

If you need to install CentOS, the following version is recommended:

https://vault.centos.org/7.4.1708/isos/x86_64/CentOS-7-x86_64-DVD-1708.iso

7.2.3 Debian

If you need to install Debian, the following version is recommended:

http://cdimage.debian.org/cdimage/archive/9.0.0/amd64/jigdo-dvd/

7.2.4 NeoKylin

If you need to install NeoKylin, the following version is recommended:

http://download.cs2c.com.cn/neokylin/

Note: for all the above operating systems, in the non-virtual machine environment, needs to deactivate the starting parameter intel_iommu

7.3 Driver Installation

For the driver installation on SC5 (X) series board, please refer to the driver installation chapter in the relevant document BMNNSDK2 provided by SOPHGO.

7.4 Software Development Environment

For the development on SC5 (X) series board, please refer to the relevant document BMNNSDK2 provided by SOPHGO.

BMNNSDK2 is an original deep learning development kit provided by SOPHGO. It includes device driver, runtime, header file and corresponding tools. PCIe drivers and BM-SMI tools related to SC5 (X) series accelerator cards are included in BMNNSDK2.

7.5 Installation and Use of SOC Mode (SC5-IO)

7.5.1 SC5-IO Installation Steps

The SOC mode of SC5 single- processor accelerator card (demo board) needs to be used with SC5-IO extension dock. Please install and use it according to the following steps.

Please follow the steps below to install SC5 and SC5-IO boards to realize SOC mode

1. Place SC5 and SC5-IO on a flat workbench.
2. Touch the metal part of the card to release static electricity.
3. Insert the SC5 card into the PCIe slot of SC5-IO. Please gently shake the SC5 board to insert the equipment into place.
4. Use the AC-DC power supply provided to plug into the DC-Jack connector, and connect the other end to AC ~ 220V.
5. The system supports on-site detection mechanism. After correct installation and power on, the on-board green LED will light up. If it does not light up, please disconnect AC ~ 220V and repeat from step 1.

7.5.2 SC5-IO Use Method

The SC5-IO extension dock contains the following functions:

1. USB3.0/2.0

There are four USB interfaces on the SC5-IO extension dock. For convenience of description, they are named as A, B, C and D from the one close to the power socket. Among them, A, B and C are normal USB 3.0/2.0 compatible interfaces. D is the same as A, B and C without 4G module; After the 4G module is selected, only USB 3.0 is available for D, and its USB 2.0 Pin (DP/DM) is switched to 4G module.

The pre-installed kernel version of SC5 supports the following USB devices:

a. USB storage：When the USB flash disk or mobile hard disk is inserted (USB power supply capability shall be considered), the storage device will be recognized as / dev / sdb1 or similar node, which is the same as that in desktop PC Linux environment. The file system supports FAT, FAT32, EXT2/3/4 and NTFS. SC5 does not support automatic mounting, so manual mounting is required: "sudo mount /dev/sdb1 /mnt". When accessing NTFS format storage devices, the pre-installed kernel version only supports reading. If writing is required, the ntfs-3g software package needs to be installed manually. Please refer to https://wiki.debian.org/NTFS. After completing data writing, please use sync or umount operation in time. When turning off

the device, please use "sudo poweroff" command to avoid violent power off and data loss.

    b. USB serial (CP210X/PL2302 series processor, modem): when the USB to serial port device is inserted, it will be recognized as /dev/ttyUSB0 or similar node, which is the same as that in desktop PC Linux environment. You can install minicom and other tools for operation.

    c. USB camera: when the camera is inserted, the device will be recognized as /dev/video0 or similar device, which is the same as that in desktop PC Linux environment. You can use tools such as ffserver to operate.

2. SATA hard disk: hot plug is not supported. Please plug in the hard disk before starting up. The device will also be recognized as /dev/sdb1 or similar nodes, which is the same as the above USB storage device.

3. 4G module (optional): please refer to the official documents of the module manufacturer for dialing, routing table configuration and other operations, which are the same as those under desktop PC Linux environment.

4. RS485 terminal (reserved position): the IO extension board provides two groups of reserved RS485 terminals, which are not wired at present. When the hardware connection is ready, it is expressed as /dev/ttyS1 and /dev/ttyS2 in the system. It can be used in the normal serial port access mode and does not support hardware flow control.

7.6 Use of Device in Docker Environment

1. Ensure that docker has been installed on the host and SC5 (X) driver installation has been completed. Bmdev-ctl, bm-sophonx and other devices can be found under /dev;

2. The docker image for development can be obtained in bmnnsdk2-bm1684 from the development file of sophon_vx.x.x.docker, docker load - i load image

3. When creating a docker container, you need to mount the corresponding device in the docker environment, --device=/dev/bmdev-ctl:/dev/bmdev-ctl, --device=/dev/bm-sophon0:/dev/bm-sophon0

7.7 Kubernetes（K8S）Device Plugin Installation and Application

7.7.1 Access to Device Plugin

The released software package contains the device plugin image file in the following format:

bitmain-tpu-plugin_ v1.x.x.tar

(The arm architecture is bitmain-tpu-plugin-arm64_v1.x.x.tar)

Of which 1 x.x stands for version number. The specific version number is related to the release.

7.7.2 Device Plugin Deployment

First, put the device plugin image into the docker image warehouse of the corresponding node, and then execute the command on the K8S master,

kubectl apply -f tpu_plugin_x86.yaml

to deploy.

The contents of tpu_ plugin. yaml refer to the following:

apiVersion: apps/v1

kind: DaemonSet

metadata:

  name: bitmain-tpu-plugin

  namespace: kube-system

  labels:

   app: bitmain-tpu-plugin

spec:

  selector:

   matchLabels:

    app: bitmain-tpu-plugin

  template:

   metadata:

    labels:

     app: bitmain-tpu-plugin

   spec:

    containers:

    - name: bitmain-tpu-plugin

```
                image: bitmain-tpu-plugin:1.0.8
                imagePullPolicy: IfNotPresent
                volumeMounts:
                - name: devfs
                  mountPath: /dev
                - name: sysfs
                  mountPath: /sys
                - name: kubeletsockets
                  mountPath: /var/lib/kubelet/device-plugins
              volumes:
              - name: devfs
                hostPath:
                  path: /dev
              - name: sysfs
                hostPath:
                  path: /sys
        - name: kubeletsockets
          hostPath:
            path: /var/lib/kubelet/device-plugins
```

Note: image users should modify the image manually according to their own image name.

8 Software Command Reference

8.1 Functions Provided by BM-SMI

The main functions of BM-SMI are as follows:

1) View device parameters and runtime status

    l View device operating mode (PCIE / SOC)

    l Check the device processor ID and the PCIE bus ID

    l View device temperature and power consumption

    l Check whether the ECC of the equipment is enabled and the number of corrections

    l View gmem totals and utilization

    l View TPU utilization rate

    l View equipment operating frequency information

    l View the gmem size of each process when running

2) Modify board parameters

    l Prohibit and enable ECC

    l Turn on and off the LED indicator on the board

3) Perform the recovery operation of the failed device

8.2 BM-SMI Basic Usage and Parameter Description

Figure 8-1 BM-SMI Display Interface

The above figure shows the display status of SC5/SC5+ (three- processor card). Each card is separated by = = = = = =. The board level attribute is displayed on the far left, and the status of single processor is displayed on the right.

Bm-smi is an executable file that does not depend on other dynamic libraries. It is located in the bin directory of the release package. The figure above is a schematic diagram of executing bm-smi.

Running bm-smi under Linux depends on the environment variable TERMINFO:

For CentOS, TERMINFO=/usr/share/terminfo; export TERMINFO is required;

For Ubuntu/debian: TERMINFO=/lib/terminfo; export TERMINFO is required.

The displayed parameters are described as follows:

- Mon Jul 8 19:57:32 2019: The time and date when bm-smi is executed (here is only an example of the format, which may be inconsistent with the image in the above example)

- SDK Version: 2.6.0: SDK version number (this is just an example and will be updated with the version)

- Driver Version: 2.6.0: the version number of the driver (this is just an example and will be updated with the version)

- Name: board name, for example 1684-SC5+

- Status: board status, Active for active status and Fault for fault status

- BoardT: board-level temperature

- Processor T: processor temperature

- BoardP: board-level power consumption

- TPU_ P: TPU module power consumption

- TPU_ V: TPU module voltage

- 12V_ ATX: board-level 12V power supply current

- ECC: whether DDR ECC is enabled

- CorrectNum: the number of times to correct errors if DDR is enabled

- SN: board serial number (17 digits in total)

- Bus-ID: PCIe mode domain:b:d.f

- Mode:PCIE or SOC mode

- Minclk: TPU minimum operating frequency

- Maxclk: TPU maximum operating frequency

- Curclk: TPU current operating frequency

- MaxP: maximum board power consumption

- TPU_ C: Working current of TPU module

- Memory-Usage: gmem total and used quantity; The memory on the board may be distributed in different address spaces. The memory we allocate is memory with continuous addresses. Moreover, due to the different size of each allocation, the

　　　memory will be fragmented, so it is likely that the utilization rate will not reach
　　　100%.

- ● TPU-Util: instantaneous utilization of TPU

- ● Fan: fan speed

The following shows the number of gmems occupied by each process (or thread) on
each device.

Processes:　　　　　　　　　　　　　TPU Memory

TPU-ID　PID　Process name　　　　　Usage

xxx　　　xxx　xxx　　　　　　　　　xxM

xxx　　　xxx　xxx　　　　　　　　　xxM

Attention:

1. Because our board supports simultaneous use of multiple tasks and users, in theory, there
can be unlimited processes to create unlimited handles to apply for global memory.
Therefore, the information of gmem occupied by all processes is not displayed in the area,
but if saved as a file, all process information will be included.

2. The gmem information occupied by the process. Each line displays the gmem
corresponding to a handle created by the process. If the process creates multiple handles,
the gmem information occupied by each handle is displayed in a separate line.

8.3 BM-SMI Command Line Parameter Description

bm-smi supports the following parameters:

　　　dev (which dev is selected to query, 0xff is for all.) type: int32，default: 255

It is used to select, query or modify the parameters of that device. All devices are selected
by default.

　　　ecc (ECC on DDR is on or off.) type: string default: ""

It is used to configure the enabling and closing of DDR ECC. An example is as follows:

bm-smi --dev=0x0 --ecc=on

bm-smi --dev=0x0 --ecc=off

When executing this command, do not let any process use this device. After setting, restart the server and PC to take effect.

- file (target file to save smi log.) type: string default: ""
  You can redirect the status of the device to a text document as follows:

  bm-smi --dev=0x0 --

  file=/home/zhangsan/bm-smi.log

- led (pcie card LED status: on/off/blink) type: string default: "on"

  It is used to configure the on and off of the board LED, as shown in the following example:

  bm-smi --dev=0x0 --led=on

  bm-smi --dev=0x0 --led=off

  Note: this function is only supported on SC5 +, SC5H and SC5P

- lms (sample interval in loop mode.) type: int32 default: "500"

  It is used to set the time interval for querying the equipment status when running bm-smi. The default is 500 ms. the minimum value of this parameter is 300 ms.

- loop (true is for loop mode, false is for only once mode.) type: bool, default: "true"

  It is used to set whether bm-smi is run in single mode or periodic mode. The default is periodic mode. Bm-smi exits after querying the equipment status once in single mode; In the cycle mode, query the equipment status repeatedly according to the lms cycle.

- recovery:

  The x function of a device is found to be faulty.

  The user removes all services from the device to the state that no upper layer services and applications use the board, and executes bm-smi -- dev = x (0/1/2/3. ....) -- recovery

  → Do not perform this operation when the board is working normally.

  → There are three BM1684 processor on SC5+. If the device mapped to any one processor is wrong, the service on the corresponding SC5+ needs to be removed. Recover any device on the board, and the whole multi processor card will recovery.

9 Equipment Status Query Interface

After the successful installation of the driver, there will be a file node to view the device status under proc and sys, as follows (take sophon0 device as an example):

9.1 sysfs Properties File Node

The following property files exist under /sys/class/bm-sophon/bm-sophon0/device:

- npu_usage, Tpu (npu), the percentage of time (window width) in operation.
- npu_usage_enable, whether to enable statistics of NPU utilization. It is enabled by default.
- npu_usage_interval, the time window width of NPU utilization, unit: ms, default: 500ms. Value range [200,2000].

Examples are as follows:

Change the time window width (only by super user):

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# cat npu_usage_interval

"interval": 600

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# echo 500 > npu_usage_interval

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# cat npu_usage_interval

"interval": 500

Enable to turn off statistics of NPU utilization:

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# cat npu_usage_enable

"enable": 1

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# echo 0 > npu_usage_enable

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# cat npu_usage_enable

"enable": 0

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# cat npu_usage

Please, set [Usage enable] to 1

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# echo 1 > npu_usage_enable

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# cat npu_usage_enable

"enable": 1

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# cat npu_usage

"usage": 0, "avusage": 0

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device#

To view NPU utilization:

root@bitmain:/sys/class/bm-sophon/bm-sophon0/device# cat npu_usage

"usage": 0, "avusage": 0

Usage indicates NPU utilization in the past time window.
Avusage indicates NPU utilization since the driver is installed.

9.2 proc fs Properties File Node

ls /proc/bmsophon/bmsophon0/

| board_power | boot_loader_version | cdma_out_time | completed_api_counter | ecc |
| pcie_cap_speed | pcie_region | tpu_maxclk | versions |

| board_temp | cdma_in_counter | processoid | dbdf | maxboardp | pcie_cap_width |
| sent_api_counter | tpu_minclk |

| board_type | cdma_in_time | processor_temp | driver_version | mcu_version |
| pcie_link_speed | sn | tpu_power |

| board_version | cdma_out_counter | clk | dynfreq | mode | pcie_link_width |
| tpuid | tpu_process_time |

The parameters are described as follows:

- board_power

    Attribute: read only;

    Meaning: board level power consumption

- board_temp

    Attribute: read only;

    Meaning: board level temperature

- Processor id

    Attribute: read only;

    Meaning: Processor ID (0x1684 / 0x1682)

- Processor _temp

    Attribute: read only

    Meaning: processor temperature

- dbdf

    Attribute: read only

    Meaning: domain: bus: dev.function

- dynfreq

    Attribute: read and write

    Meaning: enable or disenable the dynamic tpu frequency modulation function; 0/1
    is valid, other values are invalid

- ecc

    Attribute: read only

    Meaning: turn on or off ECC function

- maxboardp

    Attribute: read only

    Meaning: maximum board-level power consumption

- mode

    Attribute: read only

    Meaning: working mode, PCIe / SOC

- pcie_cap_speed

Attribute: read only

Meaning: the maximum speed supported by the device

- pcie_cap_width

  Attribute: read only

  Meaning: the maximum width of the lane supported by the device

- pcie_link_speed

  Attribute: read only

  Meaning: link speed of equipment

- pcie_link_width

  Attribute: read only

  Meaning: link width of equipment

- pcie_region

  Attribute: read only

  Meaning: the size of the device bar

- tpuid

  Attribute: read only

  Meaning: ID of TPU (0/1/2/3...)

- tpu_maxclk

  Attribute: read only

  Meaning: maximum working frequency of tpu

- tpu_minclk

  Attribute: read only

  Meaning: minimum working frequency of tpu

- tpu_power

  Attribute: read only

  Meaning: instantaneous power of tpu

- sn

  Attribute: read only

  Meaning: board serial number

- boot_loader_version

  Attribute: read only

  Meaning: bootloader version number in spi flash

- board_type

  Attribute: read only

Meaning: board type, format: product type_ PCB version number_ BOM version number_ MCU software version number

- driver_version

  Attribute: read only

  Meaning: drive version number

- board_version

  Attribute: read only

  Meaning: board hardware version number

- mcu_version

  Attribute: read only

  Meaning: MCU software version number

- versions

  Attribute: read only

  Meaning: collective display of board software and hardware versions

- cdma_in_time

  Attribute: read only

  Meaning: the total time taken by CDMA to move data from the host to the board

- cdma_in_counter

  Attribute: read only

  Meaning: the total number of times CDMA moves data from the host to the board

- cdma_out_time

  Attribute: read only

  Meaning: the total time taken by CDMA to move data from the board to the host

- cdma_out_counter

  Attribute: read only

  Meaning: the total number of times CDMA moves data from the board to the host

- tpu_process_time

  Attribute: read only

  Meaning: the time consumed by tpu during processing

- completed_api_counter

  Attribute: read only

  Meaning: the number of times the api has been completed

- sent_api_counterr

  Attribute: read only

  Meaning: the number of api that have been sent

10 Maintenance and Management

10.1 In-Band Management

SOPHGO provides tools for upgrading board firmware, which are used as follows (this function needs to be carried out under the guidance of our FAE):

a) Update MCU application only

./bm_firmware_update –file=bm1684evb-mcu.bin --target=mcu

./bm_firmware_update

device id: 0

bin file: bm1684evb-mcu.bin

target: mcu

bin file size 65536

EFIE: offset 0x00008000 length 0x00005208 checksum 054dc3e5b64ac636922e423f6c7eba6f

Programming MCU APP firmware ...

program efie succeeds.

program app succeeds.

checksum compare succeeds.

BM1684 mcu firmware update processor _id = 0 completed

Note: The default of dev_id is 0. When you need to upgrade multiple cards, you need to specify dev_ id; When you need to upgrade multiple cards, dev_ id=0xff. Path of files required for mcu firmware upgrade:

bmnnsdk2-bm1684_v2.5.0/bin/firmware/mcu

The firmware of different boards is different. Upgrade after confirmation.

b) Update all MCU firmware

./bm_firmware_update –file=bm1684evb-mcu.bin --target=mcu --full

./bm_firmware_update

device id: 0

bin file: bm1684evb-mcu.bin

target: mcu

bin file size 65536

Programming MCU FULL firmware ...

BM1684 mcu firmware update processor _id = 0 completed.

Note: The default of dev_id is 0. When you need to upgrade multiple cards, you need to specify dev_ id; When you need to upgrade multiple cards, dev_ id=0xff. Path of files required for mcu firmware upgrade:

bmnnsdk2-bm1684_v2.5.0/bin/firmware/mcu

The firmware of different boards is different. Upgrade after confirmation.

Attention:

a) MCU app burning takes about 15 seconds, and all burning takes about 30 seconds.

b) After the burning is completed, the MCU needs to be powered off and restarted before loading a new firmware.

c) You can view the MCU version number through /proc interface.

10.2 Out-of-Band Management

The SC5 (X) accelerator card supports docking with the host server through SMBUS on the board side. The server manufacturer bmc can read the reg of the board and adjust the heat dissipation conditions of the server in real time according to the working conditions. Please consult FAE engineer or product manager for details.

10.3 LED Status

The LED indicator on the SC5 board flashes in the following way after the driver is installed:

After the board is installed and driven:

When TPU works at the maximum frequency, it flashes once in 2 seconds;

When TPU works at intermediate frequency, it flashes once a second;

When TPU works at the minimum frequency, it flashes three times in one second;

The bm-smi tool can control the on and off of the LED light. The LED light is only bound to the device corresponding to function 0 on the board. For specific control methods, please refer to the bm-smi chapter.

SOPHGO 算能

11 Compatibility List

11.1 Server brand, Model and Core Configuration

Brand	Series	Basic Configuration	PCIE Slots	Height
Dell	R740	Intel® Xeon® Gold 5117 PROCESSOR @ 2.00GHz	5	2U
Dell	R940	4x Xeon Gold 5118 @2.3G	13*	3U
SuperMicro	4028GR-TR2	Intel® Xeon® PROCESSOR E5-2640 v4 @ 2.40GHz	9	4U
SuperMicro	6049GP-TRT	Intel® Xeon® Gold 5120 PROCESSOR @ 2.20GHz	16	4U
PowerLeader	PR4764GW	Intel® Xeon® PROCESSOR E5-2650 v4 @ 2.20GHz	5	4U
Inspur	NF5468M5	2* Intel Xeon Skylake 4210 Processors	20	4U
Sugon	X785-G30	Intel® Xeon® Gold 5120 PROCESSOR @ 2.20GHz	16	4U
H3C	R5300 G3	2*Intel® Xeon® Scalable Processors	20	4U
Phytium		Phytium 2000+ /64 Processor	2/4	2U
Inspur	NF5468M6	Intel 3rd generation processor/fully equipped with 16 SC5+	20	4U
Inspur	CS5280H2	3*Hygon 7000 Processors fully equipped with 7 SC5+	10	2U
H3C	R5300 G5	Intel 3rd generation processor/fully equipped with 16 SC5+	20	4U
Nettrix	X620 G40	Intel 3rd generation processor/fully equipped with 16 SC5+	9	2U
Nettrix	X620 G30	Intel 2rd generation processor/fully equipped with 6 SC5+	8	2U
HUAQIN	P6210	Intel 2rd generation processor/fully equipped with 6 SC5+	8	2U
HUAQIN	P6220	Intel 3rd generation processor/fully equipped with 6 SC5+	8	2U
ENGINETECH	EG520-G20	Intel 2rd generation processor/fully equipped with 6 SC5+	6	2U
ENGINETECH	EG520-G30	Intel 3rd generation processor/fully equipped with 10 SC5+	6	2U
ENGINETECH	EG620-G20	Intel 3rd generation processor/fully equipped with 6 SC5+	6	2U
SUPER CLOUD	RS216	HYGON Dhyana 72xx/52xx/full range of processors/fully equipped with 6 SC5+	10	2U
GOOXI	SL201-D12RE	Intel 2rd generation processor/fully equipped with 8 SC5+	11	2U

SITONHOLY	IW2221-2GR	Intel 3rd generation processor/fully equipped with 4 SC5+	10	2U
HANGTIANLIAN ZHI	28121TR-IC	Intel 3rd generation processor/fully equipped with 6 SC5+	10	2U
HANGTIANLIAN ZHI	26082R-IB	Intel 2rd generation processor/fully equipped with 1 SC5H	6	2U
HANGTIANLIAN ZHI	28121TR-HC	HYGON 2, 3processor/fully equipped with 4*SC5+	8	2U
OCLOUDWARE	YJ-5210	Intel 2rd generation processor/fully equipped with 6 SC5+	6	2U
LOONGSON	T22C08B1	LS3C5000 processor / fully equipped with 6 SC5+	6	2U
JINPIN	KU2208	Intel 2rd generation processor/fully equipped with 6 SC5+/6 SC7 HP75	10	2U
DAYINMAO	SR659FV2	Phytium S2500 processor/fully equipped with 6 SC5+	8	2U
This form will be continuously updated					

Table 11-1 SC5（X）List of Compatible Servers

*13 x Gen3 PCIE slots (3 x 8 slots + 10x 16 slots)

11.2 Operating Systems

The following table is continuously updated. If your system is not included in the table, please consult our technicians for details.

Operation System	Recommended major version	Recommended kernel version
Ubuntu	16.04.3	4.15.0-88-generic
CentOS	7.4.1708	3.10.0-693.el7.x86_64
Debian	9.0.0	4.9.0-11-amd64
Kylinos	5.0.0	

Table 11-2 SC5（X）List of compatible operating systems

11.3 Supported Algorithm Frameworks and Operators

The algorithm model framework supported by SC5 (X) series accelerator card includes: Caffe, DarkNet, Pytorch, TensorFlow, Mxnet, Paddle, etc.

The series of SOPHON BM1684 products all use the unified tool chain program, algorithm migration, deployment tool and SDK package. For the list of supported algorithm frameworks and operators, please refer to the *NNToolchain User Guide*.

Algorithm Framework	Version Supported
Caffe	Latest
DarkNet	Latest
Pytorch	$\leqq 1.3.0$
TendorFlow	$\leqq 1.13.1$
MxNet	$\leqq 1.15.1$
PaddlePaddle	Paddle Lite version 1.6 and above

Table 11-3 Algorithm framework and version supported by SC5 (x)

12 Common Problems and Solutions

Problem: PCIe device can be viewed through lspci, but the program fails to open the device.

Solution: check whether/dev/bm-sophon0 and other device nodes are generated. If not, install the PCIe driver according to Chapter 7 of this manual. If it still fails after driver installation, please provide the kernel log to our technical support personnel. Please obtain the kernel log in the following way: sudo cat/dev/kmsg.

Problem: PCIe link width may be recognized as x1.

Solution: power off the host, pull out the SC5 board and re-insert it, and confirm that the gold finger at the bottom of the board is fully embedded in the PCIe card slot (without unilateral warping), restart the host and check the link width again.

14 Certificates

Relevant domestic and international certifications of SC5 (x) series accelerator cards that have been and will be obtained include: CE, FCC, ROHS, IC, reach, WEEE, etc.

No.	Country-Region	Certificate	Standard
1	Europe	CE	Safety: / EMC: EN 55032:2015 EN 55035:2017 EN 61000-3-2:2014 EN 61000-3-3:2013 IEC 61000-4-2:2008 IEC 61000-4-3:2006+A1:2007+A2:2010 IEC 61000-4-4:2012 IEC 61000-4-5:2014 IEC 61000-4-6:2013 IEC 61000-4-8:2009 IEC 61000-4-11:2004
2	U.S.A.	FCC	47 CFR FCC Part 15 Subpart B
3	Canada	ISED(IC)	ICES-003 Issue 6, 2016(updated 2019)
4	Europe	RoHS	IEC 62321-3-1:2013 Ed.1.0 IEC 62321-4:2013+AMD1:2017 IEC 62321-5:2013 Ed.1.0 IEC 62321-6.2015 Ed.1.0 IEC 62321-7-1:2015 Ed.1.0 IEC 62321-7-1:2017 Ed.1.0
5	Europe	REACH	REACH：Regulation (EC) No 1907/2006
6	Europe	WEEE	2002/96/EC

Table 14-1 SC5 (x) Certificate List

15 Acronyms

ARM	Advanced RISC Machine in this document refers to the main control PROCESSOR unit used in bm1684 processor
BM168x	The X-generation cloud TPU launched for the field of deep learning by SOPHGO
BMNNSDK	SOPHGO original deep learning development kit
BM-SMI	Software management tool for board status and information on the host side when running in PCIe mode
CFM	Cubic Feet Per Minute
ECC	Error Checking and Correction
inch H2O	inch water column
JPU	JPEG Process Unit
PCIe	Peripheral Component Interconnect Express
PCIe-Mode	A product form. The SDK runs on the X86 platform. The bm168x exists as a deep learning computing accelerator card with PCIe interface
SAIL	SOPHON artistic intelligent library, also known as SOPHON Inference, provides some high-level interfaces, mainly encapsulating runtime modules such as BMRuntime, BMCV and BMdecoder.
SMBUS	System Management Bus
SOC-Mode	A product form. The SDK runs independently on the bm168x platform and supports interconnection with other devices through Gigabit Ethernet
TPU	Tensor process unit, a customized ASIC processor pioneered by Google, is widely recognized internationally. SOPHON BM168x series deep learning processor is completely independently developed, and the TPU part also adopts similar processor design ideas, which also belongs to the type of TPU processor
VPU	Video Process Unit

16 Appendix

You can find our common files in the following websites:

（1） official website：English https://en.sophgo.com Chinese https://www.sophgo.com

（2） SDK related information： https://developer.sophgo.com/site/index/material/all/all.html

（3） Official Forum： https://developer.sophgo.com/forum/index.html

（4） Video tutorial： https://developer.sophgo.com/site/index/course/all/all.html

（5） Open source community： https://github.com/sophgo

Note: in case of document update, please contact the technical personnel of SOPHGO in time to obtain the

latest version.

16 Appendix

17 Index