Cambricon

寒　　武　　纪

MLU270-F Series Intelligent Processing Card User Manual

V0.9

Revision History

Document Name	MLU270-F Series Intelligent Processing Card User Manual		
Revision	V0.9		
Created Date	2019.07.03		
Revision History			
No.	Date	Version	Notes
0	2019.07.03	V0.9	First Release
1			
2			
3			
4			
5			

Technical supports contact: service@cambricon.com

www.cambricon.com

Content

1.   Product Brief
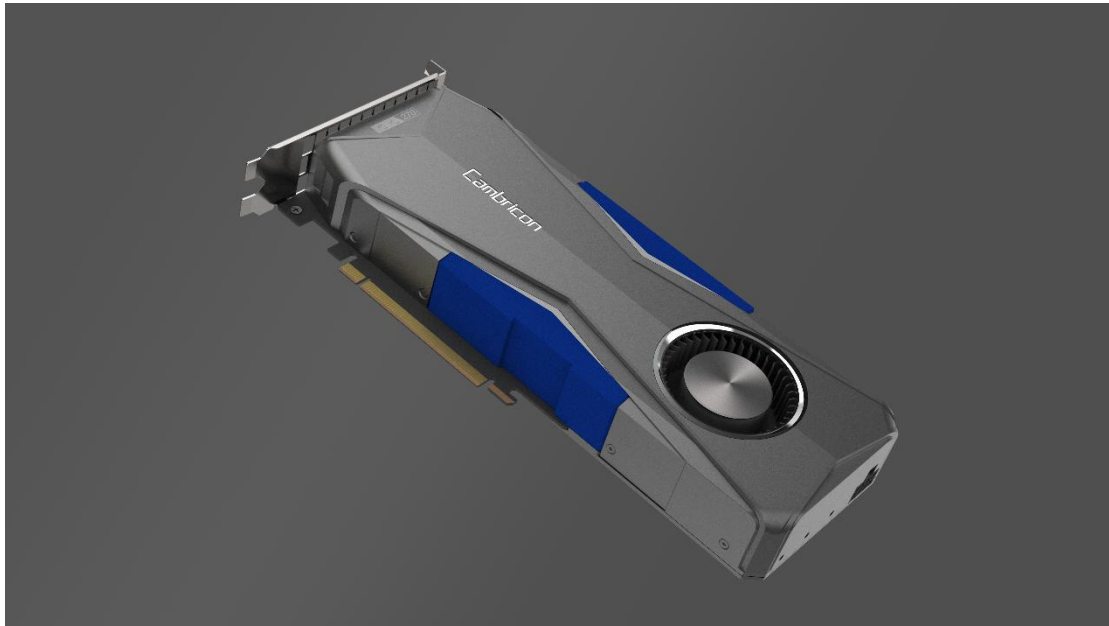

Figure.1 MLU270-F Series Intelligent Processing Card appearance

Specially designed MLU270-F series dedicated to AI inferences and data center accelerations with high EER (Energy Efficiency Rate)

The SIYUAN 270 ASIC has been designed based on several innovative technologies in the architecture of processors owned by Cambricon. Along with it, many up-to-date features have been integrated into a standard FHFL PCIe card, which can be inserted into a modern AI PC/server so as to provide a huge extension of calculation power of AI inferences. MLU270-F series has a moderate TDP of 150W and it can provide a calculation power as high as 4 times of the previous generation. It can be widely used in applications such as vision, voice, natural language processing, legacy machine learning, and many other AI scenarios, and it can be used in an AI inferences platform to make it work with even higher EER.

Brand-new Cambricon MLUv02 Architecture

The MLUv02 architecture is not just a simple update from previous generation, but a brand new design based on NOC (Network on Chip), which will guarantee the parallel efficiency of execution of 16 NPU cluster within the SIYUN 270 ASIC. The dataflow will be compressed within the chip by dedicated hardware engine and this will increase the volume and bandwidth effectively. The new architecture can fully support all the AI accuracies such as INT16, INT8, INT4, FP32, FP16, and provide necessary calculation power to many kinds of Neural Network. In a word, the new architecture can provide both good universality and best performances to customers at the same time.

One more step on the performances of inferences

When using INT8 accuracy for AI inference calculations, the performance of non-sparse networks is improved up to 4 times as much as that of the previous generation. MLU270-F series can provide a great EER as high as 40 times of a normal CPU. There are embedded newly designed code/decode hardware for videos and pictures, so when the system have to

handle this kind of tasks, MLU270-F series will decrease the load of CPU and the occupation for PCIe bandwidth, so as to increase the performances even more.

Calculating flexibility and programmability

SIYUAN 270 ASIC can provide supports to many kinds of Neural Networks. Cambricon NeuWare software stack can easily deploy inference environment.BANG Lang.CHENG environment can be directly customized for computing resources ，meeting all kinds of AI customization requirements,professional but not specialized.

2.   Product specifications

2.1   Performance specifications

Table 1 MLU270-F Series Intelligent Processing Card Hardware Specifications

Card Type	MLU270-F4/F5/F4K/F5K
Processor Architecture	Cambricon MLUv02
Core Clock	MLU270-F4/F5:1 GHz, MLU270-F4K/F5K:1.2 GHz
Peak INT8	MLU270-F4/F5:128 TOPS （Dense）， MLU270-F4K/F5K:160 TOPS （Dense）
Calculation accuracy	INT16, INT8, INT4, FP32, FP16
Video Decoding Support	Yes
Total Memory Size	MLU270-F4/F4K:16GB, MLU270-F5/F5K:32GB
Memory Bus Width	256-bit
Memory Bandwidth	102.4 GB/s
System Interfaces	PCI Express Gen3, x16, Supporting Lane Reversal
PCIe ID	PCIe Vendor ID 0xCABC PCIe Device ID 0x0270 PCIe Sub-Vendor ID 0xCABC PCIe Sub-System ID 0x0018/0x0019
Outline	267mm*111.15mm, Double slots
TDP	150W
ECC Supporting	Yes
Heat Dissipation	Passive

2.2   Software specifications

Table 2 MLU270-F Series Intelligent Processing Card Software Specifications

PCIe Base address	PF （1，64bit）： BAR0: 256MB prefetchable

	BAR2: 64MB prefetchable
	BAR4: 64MB prefetchable
	VF（4，64bit）：
	BAR0: 256MB prefetchable
	BAR2: 64MB prefetchable
	BAR4: 64MB prefetchable
SMBus（8bit Address）	0x8E(write) 0x8F（read）

SMBUS Register is 32-bit wide, and below describes how to read a register（S:Slave，M:Master）：

Table 3 SMBUS Registers Reading

Direction	Bits	Content
M → S	1	START
M → S	8	SLAVE ADDRESS（Write）
S → M	1	ACK
M → S	8	REGISTER ADDRESS
S → M	1	ACK
M → S	1	RE START
M → S	8	SLAVE ADDRESS（Read）
S → M	1	ACK
S → M	8	DATA[7:0]
M → S	1	ACK
S → M	8	DATA[15:8]
M → S	1	ACK
S → M	8	DATA[23:16]
M → S	1	ACK
S → M	8	DATA[31:24]
M → S	1	NACK
M → S	1	STOP

Table 4 SMBUS Registers Description

Registers	Address	ACCESS	Description
Total Card Power	0x01	RO	Card Power consumption, Float Data, Unit W
Card Temperature	0x02	RO	Card Temperature, Float Data, Unit ℃
Chip Temperature	0x03	RO	Chip Temperature, Float Data, Unit ℃
power brake	0x05	WO	The main frequency is reduced to 25% of current frequency when write 0x04,and restore to pre-frequency when write 0x01
PCIE Vendor ID and Device ID	0xA0	RO	[15:0] Vendor ID：0xCABC [31:16] Device ID：0x0270
PCIE Sub-Vendor ID and Sub-System	0xA1	RO	[15:0] Sub-Vendor ID：0xCABC [31:16] Sub-System ID：for example 0x0018

ID			
Device Name	0xF0	RO	Display device type，e.g 0X18 means device type is F4
Vendor Name	0xF1	RO	Display vendor name
Hardware vision	0xF2	RO	Display hardware vision
Firmware Revision	0xF3	RO	Display Firmware Revision，e.g 0x02180016 means the main chip version is 0x02,device type is 0x18,main version is 0x00,subversion is 0x1,patch number is 0x6
Manufacturing Time	0xF4	RO	Dispaly manufacturing time,e.g 0x1811 means manufactured in 2018.11
Device ID	0xF5	RO	Display Device ID，e.g 0x20023 means device id is 20023

2.3　Operating Environment

Table 5 MLU270-F Series Intelligent Processing Card Working Environment

Operating Temperature	0℃ ~ 45℃
Storage Temperature	-40℃ ~ 75℃
Operating Humidity	5%—95% Relative Humidity
Storage Humidity	5%—95% Relative Humidity

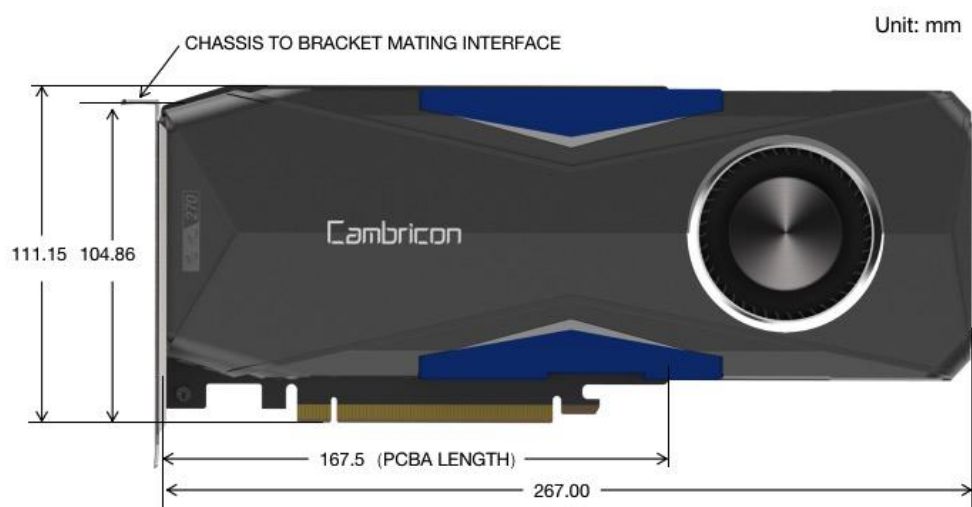## 2.4　Mechanical Specifications

Figure. 2 MLU270-F Series Intelligent Processing Card Dimensions

2.5 Package Size and Weight

Table 6 MLU270-F Series Package Size and Weight

Type	Weight	Size	Notes
Single Card	874g	267mm*111.15mm*37.10mm	NA
Gift package with single card	1.4kg	385mm*230mm*97mm	Including gift package,MLU270-F series card,expand stent,8-pin dongle cable,PE bag,desiccant
Gift Box	9.3kg	600mm*400mm*253mm	6 gift packages each box

Notes：The weight is a measured value with a tolerance of +-10%

2.6 Thermal Specifications

2.6.1 MLU270-F Series Power and Temperature specifications

Table 7 MLU270-F Series Thermal Design specifications

Item	Parameters
TDP(Thermal Design Power)	150W
Recommended Operating Tj	0-85℃
Slowdown Tj	87℃
Slowdown Rate	50%
Shutdown Tj	95℃

2.6.2 MLU270-F Series Airflow Directions Requirement

MLU270-F Series use a blower to cool the card, the airflow direction is shown below：

F

Figure.3   MLU270-F Series Airflow direction

2.7   Power Supplies Specifications

MLU270-F Series Intelligent Processing Card provides a 8-Pin Power Connector as below：
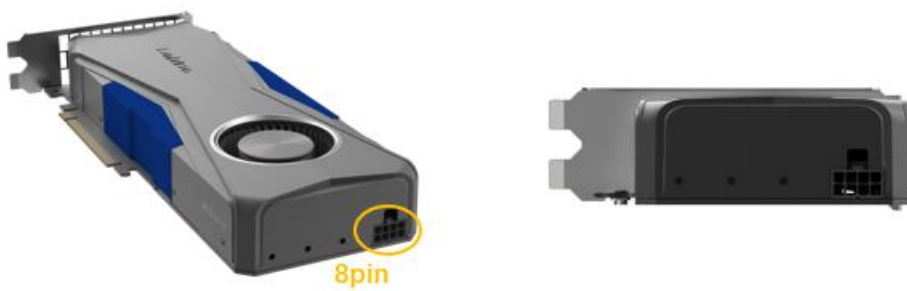

8pin

Figure.4 8-Pin Power supply Connector

Table8 Supported Auxiliary Power Connections

Board Connector	Cable used
CPU 8-Pin	1× CPU 8-pin cable
CPU to PCIe 8-Pin Dongle Cable	2× PCIe 8-pin cable
	2× PCIe 6-pin cable
	1× PCIe 8-pin cable and 1× PCIe 6-pin cable

Users can use the CPU 8-Pin Power Socket within a Server directly, or use two 8-pin PCIe Power Sockets, along with a One-To-Two Adapter Cable.

Table9 Adapter Cable Connections

	8-Pin CPU Power connector	PCIe 8-Pin Socket 1	PCIe 8-Pin Socket 2	Colour
12V	5	3		Yellow
12V	6	1,2		Yellow
12V	7		3	Yellow
12V	8		1,2	Yellow
GND	1	7,8		Black
GND	2	5,6		Black
GND	3		7,8	Black
GND	4		5,6	Black

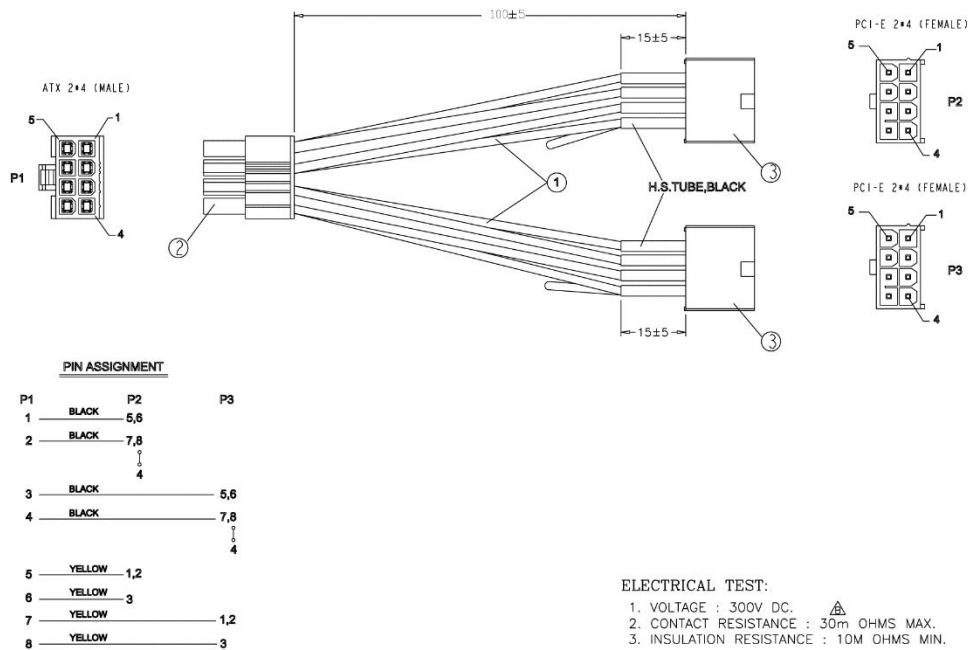Figure.5 Adapter Cable

Note: The first pin (Pin No.1) of every connector/socket should have obviously distinguishable mark at the side.

Table 10Input Voltage Requirement

Power Supply	Min	Normal	Max
PCIe edge connector（12V）	11.04V	12V	12.96V
8pin connector（12V）	11.04V	12V	12.96V
PCIe edge connector（3V3）	3.0V	3.3V	3.63V

Table 11 Current Requirement

Power Supply	Peak Current	Moving Average
8pin connector（12V）	20A	200us
	17A	1ms
	13A	5ms

Table12 Power Capping

Power Capping Threshold	150W
Power Capping Response time （typical）	50ms
Power Capping Response time （max）	100ms

Table 13 Power Break

PB# PCIe pin assignment	B30
Power Brake response time（typical）	150us
PB# input insertion low time （min）	250ms
Power brake hardware slowdown factor	4x

3.　Cambricon NeuWare Development Environment

NeuWare can fully support all kinds of mainstream programming framework, such as TensorFlow, Caffe, PyTorch, MXNet and so on.With above mentioned Programming Frameworks, users can easily and conveniently develop and deploy their Deep Learning Applications on Cambricon MLU270-F Series Intelligent Processing Cards.At the same time, NeuWare provides complete runtime system and driver software to speed up the system integration procedure.

NeuWare also provides a full set of software tools such as Application Development, Function Debugging and Performance Optimization. The Application Development Tools include Machine Learning Library, Runtime Library, Compiler, Model retraining tools and Domain-Specific (for example Video Analysis) SDK; The Function Debugging Tools can fulfill all the requirements from different levels of Programming Framework and Function Library; The Performance Optimization Tools include tools for performances analysis and system monitoring.
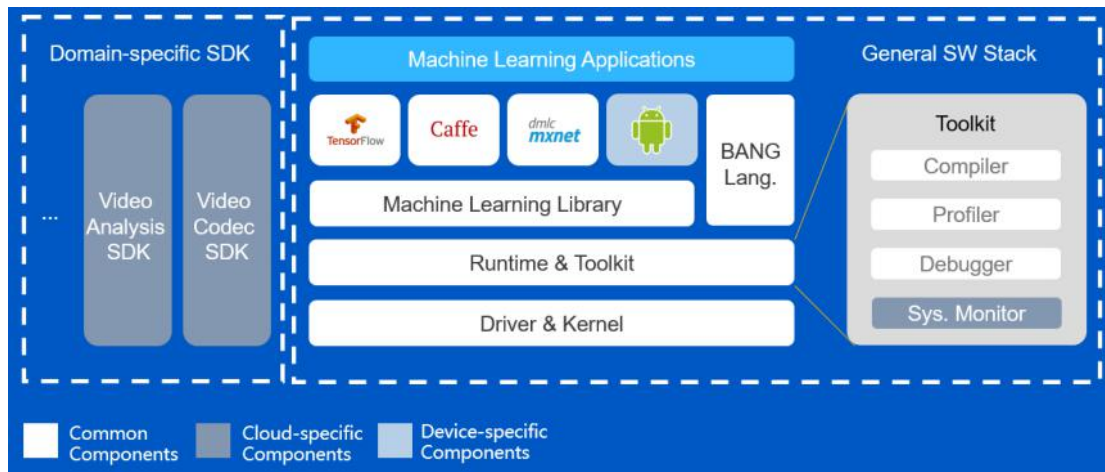
Figure.6 Cambricon NeuWare

For more information,please visit www.cambricon.com
Tel: 86-10-83030003
E-mail: business@cambricon.com
Address: 11/F, Block D, Zhizhen Tower, No.7 Zhichun Road, Haidian District, Beijing, China

4.   Compliance

The MLU270-F Series is compliant with the regulations listed in this section. Compliance marks, including the FCC ID numbers, can be found on the labels of each devices.

United States

Federal Communications Commission (FCC)

This device complies with Part 15 of the FCC Rules.
Operation is subject to the following two conditions: (1) This device may not cause harmful interference, and (2) this device must accept any interference received, including interference that may cause undesired operation.

This equipment has been tested and found to comply with the limits for a Class B digital device, pursuant to Part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference in a residential installation. This equipment generates, uses and can radiate radio frequency energy and, if not installed in accordance with the instructions, may cause harmful interference to radio communications. However, there is no guarantee that interference will not occur in a particular installation.

If this equipment does cause interference to radio or television reception, which can be determined by turning the equipment off and on, the user is encouraged to try to correct the interference by one or more of the following measures:

- Reorient or relocate the receiving antenna
- Increase the separation between the equipment and receiver
- Connect the equipment into an outlet on a circuit different from that to which the receiver is connected
- Consult the dealer or an experienced radio/TV technician for help

Underwriters Laboratories (UL)

UL Listed Product Logo for MLU270-F Series Intelligent Processing Cards，model name

MLU270-F.