



Sophon SC1 White Paper

V1.0

Version	Update Content	Release Date
V1.0	-	2017/10/25

CONTENT

1. Overview	4
1.1 Specifications.....	4
1.2 Functional List.....	5
1.3 Appearance.....	5
1.4 System Block Diagram.....	6
2. Interfaces & Electrical Characteristics	7
2.1 Interface.....	7
2.2 PCIE Interface.....	7
2.3 Power.....	7
3. Design for Stability & Reliability	8

1. Overview

Sophon SC1 is the latest deep learning accelerated card, an advanced fan-cooled module that combines the BM1680 into a compact, easy to integrate package. It is designed for the inference of multiple deep learning applications, allows users to calmly deal with the large-scale data.

There is one BM1680 on Sophon SC1, which contains 64 NPUs, each NPU contains the following configuration:

64 single precision (FP32) EU operation cores
512KB SRAM (program visible)

The single-precision peak performance of SC1 is 2TFlops/s, and SC1+ card which equipped with two units of BM1680 reach the single-precision peak performance to 4TFlops/s.

Sophon SC1 card equipped with 16GB of DDR4 memory, bandwidth up to 83.2GB/s. With the provided toolkit and APIs from us, you can further simplify the buffer management of DL workload to achieve better load balancing.

Each Sophon SC1 card has a PCIE3.0x8 interface, can access large-capacity system DRAM memory, to achieve two-way data exchange and stream between the card and the system.

Sophon SC1 card max power consumption is 85W, Sophon SC1+ card max power consumption is 150W, the actual power consumption according to the workload was dynamic. For the convenience of deployment, the card were carried out active cooling design, to ensure that it can work stably under 0 °C ~ 50 °C ambient temperature.

1.1 Specifications

Sophon SC1/SC1+ Specifications:

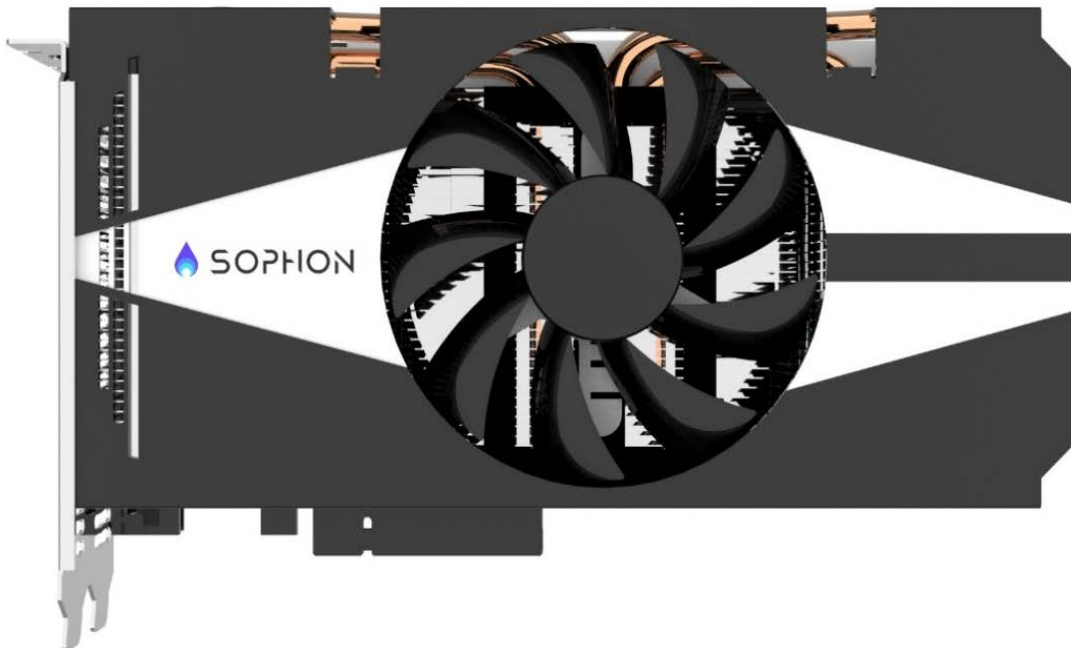
	Sophon SC1	Sophon SC1+
Number of BM1680	1	2
Single-precision (FP32) Peak Performance	2T	4T
On Chip SRAM	32 MB	64 MB
Memory	16 GB	32 GB
Chip Link Bandwidth	50 GB/s	50 GB/s
System Interface	PCIE3.0x8	PCIE3.0x8
Thermal Solution	Positive	Positive
Form Factor (mm) (Length*Height*Thickness)	216.6*128.9 *39.8	324.8*125.9*39.8

1.2 Functional List

Support the motherboard temperature detection and automatic fan speed control
Support manual reset function
Widely used for deep learning Inference acceleration
Applicable to the CNN/DNN/RNN and other deep learning network acceleration

1.3 Appearance

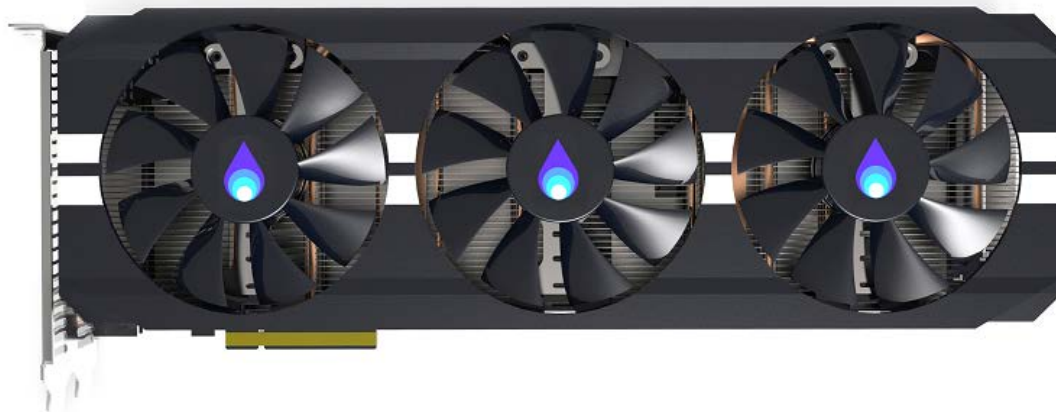
Sophon SC1 appearance:



Sophon SC1 PCB:



Sophon SC1+ appearance:

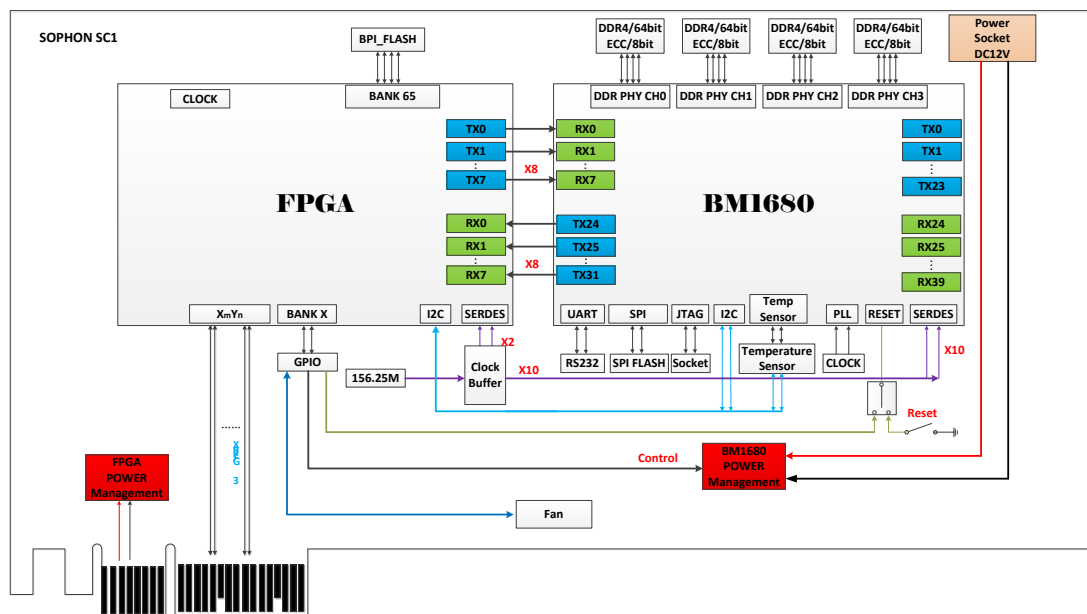


Sophon SC1+ PCB:

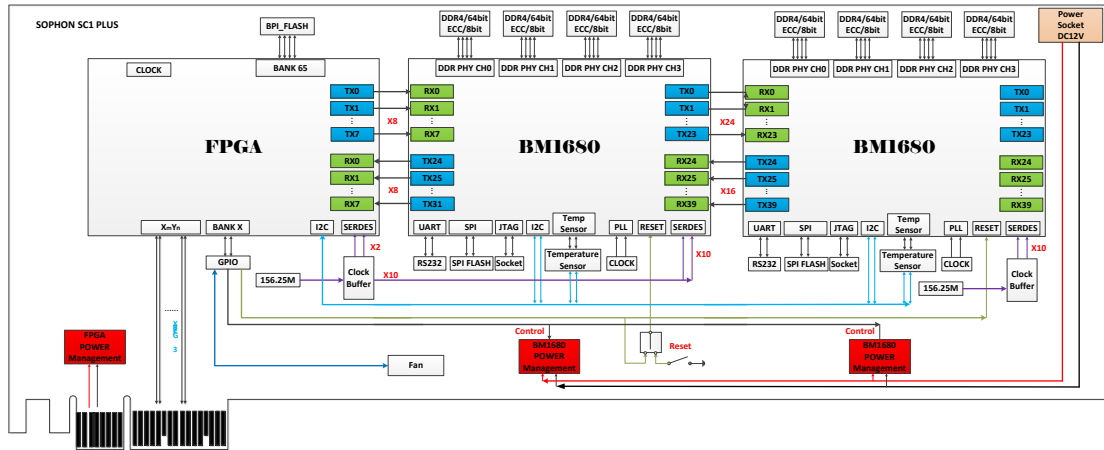


1.4 System Block Diagram

Sophon SC1 system block diagram:



Sophon SC1+ system block diagram:



2. Interfaces & Electrical Characteristics

2.1 Interface

Support PCIE 3.0 transmission interface, backward compatible with PCIE 2.x / 1.x
Support program downloading by JTAG interface
Support system software debugging through RS232 serial port

2.2 PCIE Interface

PCIE bus is high-speed differential bus, the link can be composed of multiple lanes. The PCIE link can support 1,2,4,8,12,16 and 32 Lanes, namely x1, x2, x4, x8, x12, x16 and x32 width of the PCIE link. Sophon SC1 uses x8-wide links. Its related characteristics are as follows:

PCIE Version	Single Lane Peak Bandwidth	Coding
1.0	2.5GT/s	8/10b
2.0	5GT/s	8/10b
3.0	8GT/s	128/130b

2.3 Power

Sophon SC1 card has two power supply chains: the BM1680 chip and related peripherals get 12V power which directly comes from ATX, and the FPGA and related peripherals get 12V and 3.3V power from the PCIE interface.

The electrical characteristics of the two cards are as follows:

Power Domain	Power Supply	Voltage	Sophon SC1		Sophon SC1+	
			Current	Power	Current	Power
FPGA domain	PCIE	12V	1.1A	13.2W	1.1A	13.2W
	PCIE	3.3V	0.5A	1.65W	0.5A	1.65W
BM1680 domain	ATX	12V	5.8A	70W	11.3A	135W

3. Design for Stability & Reliability

In order to ensure the stability & reliability of the card, the following design requirements have been done:

Independent power supply for FPGA and BM1680, to minimize the mutual interference between them
Power protection has been supported in power supply module, in order to avoid power supply failure, such as overvoltage, overcurrent and short circuit protection
14-layer PCB card design, which contains multiple isolation layers, the key signals are fully isolated to reduce the interference between the signals
Choose a high precision, high stability outside the modulation clock
Use both heat sink and fan in thermal solution, to ensure there is no heat dissipation occurring and the card won't crash during long-time running