Current Solutions from China and Enterprise

Introduction:

The cryptocurrency mining boom of the mid-2010's had a profound impact on the Chinese domestic market. Overnight small businesses sprung up to meet the demand for large ASIC and GPU mining farms that were increasingly more popular. This became an industrial undertaking with the development of custom power solutions with discontinued server power supplies and the modification of commercial video cards for optimal Hash Rate.

This industry has since pivoted to develop custom solutions for the Artificial Intelligence boom today. New innovative products like the 2080ti 22Gb and the 2080ti 44Gb were the initial push forwards in terms of hosting LLM's on home hardware. Later as V100's, Titan V's, and all the flavors of the Volta architecture began appearing on the market at reasonable prices. With the appearance of Ampere, Hopper, Ada, and now Blackwell architectures; small Chinese developers followed with new SXM adaptations.

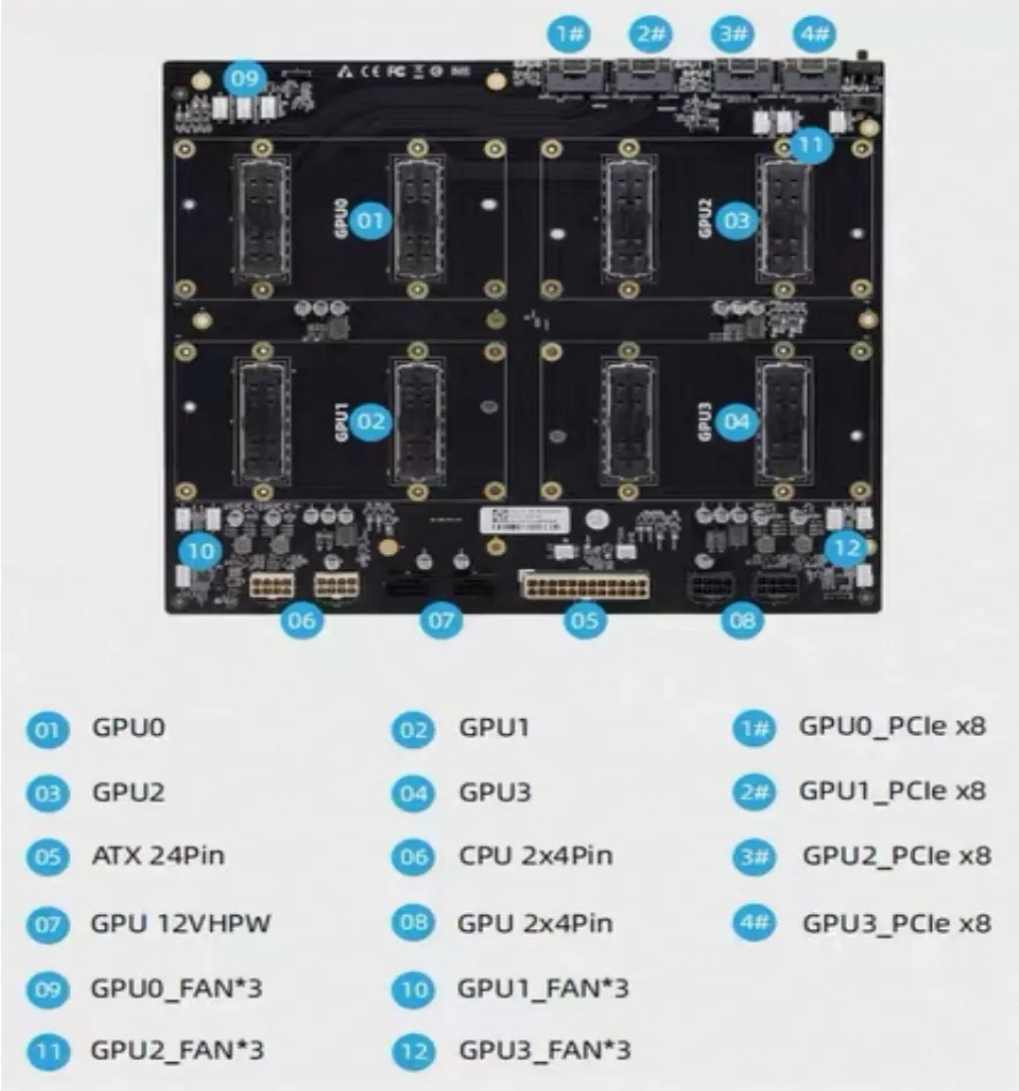
This document will explore the current offerings on the Chinese market, that can still be ordered from a variety of websites. Additionally, we will explore a more sustainable and stable source of NVLink capable hardware that is reasonably priced and is independent of future tariffs for U.S. consumers.

SECTIONS

1-Current Chinese Offerings
1.1- Quad SXM2 Chinese Board pg.3
1.2- Dual SXM2 Chinese Board pg.4-5
1.3- PCIe SFF-8654 SlimSaS Adapters pg.6-8
1.4- PCIe Single SXM Module Adapters pg.8-9
2-Enterprise Solutions
2.1- Cheap Server Platform pg.9-9
2.2- Quad and Octa Server Trays SXM2 pg.11-14
2.3- Nvidia DGX/HGX Trays SXM4 and SXM5 (Holy Grail) pg.14-19
3- CERN-P v2.0 License pg.20

Current Chinese Offerings:

Quad SXM2 NVLink Adapter
Status: **Currently not on U.S. Market or Chinese Sites**
Article (https://blog.rexyuan.com/the-most-esoteric-egpu-dual-nvidia-tesla-v100-64g-for-ai-llm-41a3166dc2ac)



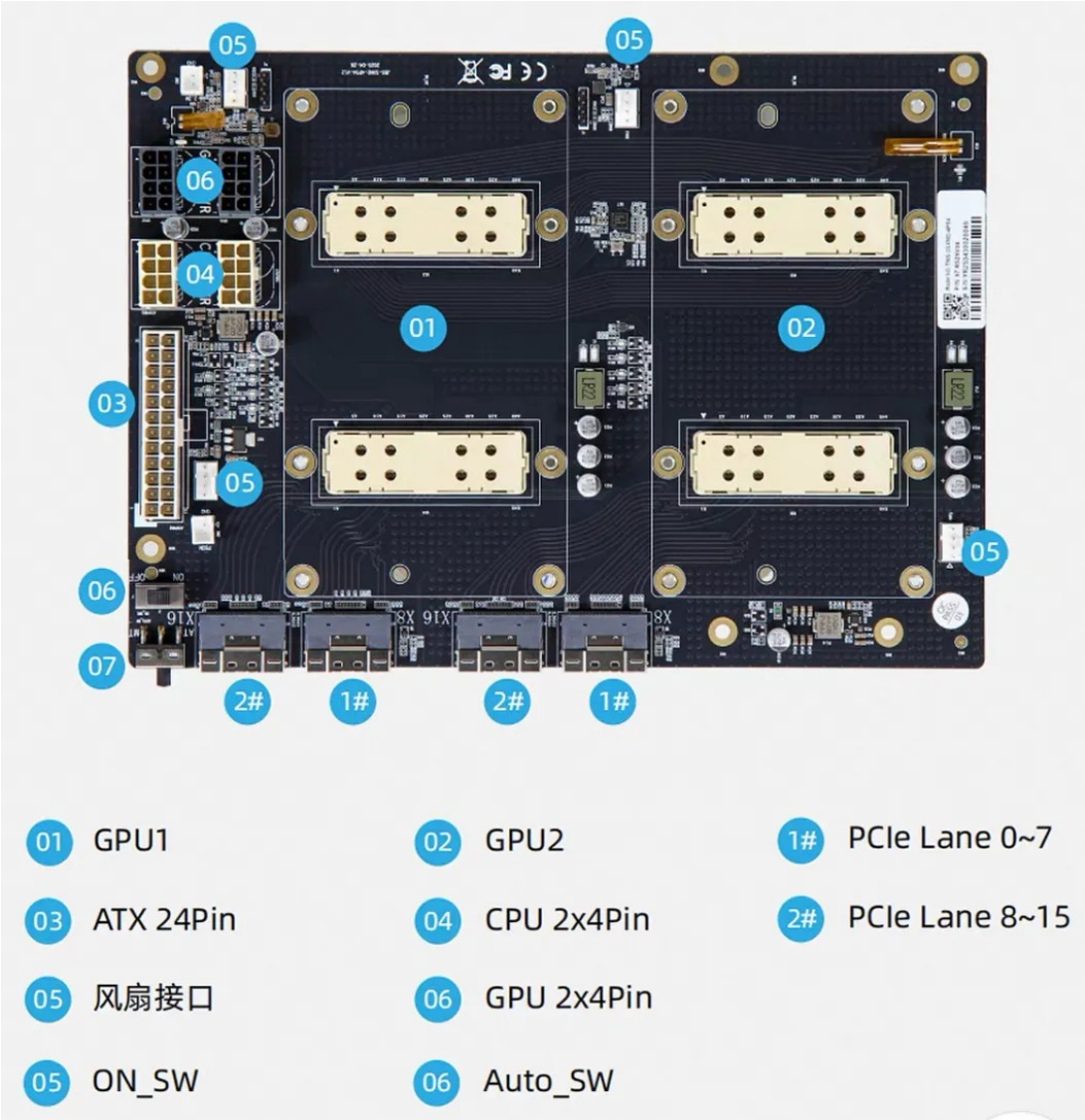
Dual SXM2 NVLink Adapter
Status: Currently on the U.S. Market $250-$400

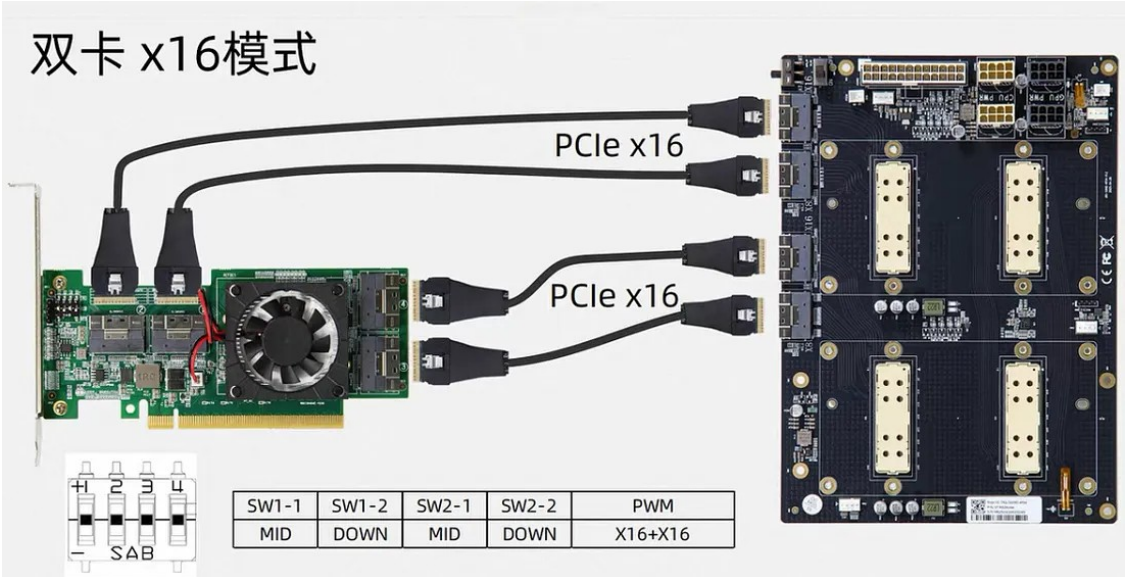
There are excellent and very affordable NVLink capable boards. This allows developers to utilize a part of P100 16 Gb, V100 16Gb or V100 32Gb SXM2 Modules. Fielding a combined **32gb** or **64gb** of NVLinked HBM2 memory with a cumulative bandwidth of **1800Gb/s**! These adapters allow the full power of the SXM2 V100 to be utilized. These SXM2 modules offer superior performance in every parameter to the subsequent Turing architecture. These boards use 2 PCIe power 12vdc and 2 EPS/CPU 12vdc Molex ports for maximum SXM2 power delivery.

This adapter is a simple dual SXM2 NVLink capable adapter. The output is 4 SFF-8654 ports, this allows for each SXM2 modules to access full PCIe 3.0X16 on the host motherboards. The module can be powered via the same PSU as the host system by engaging the **AUTO Switch**. **MANUAL Switch** allows for the addition of a second power supply via the 24-pin Molex connector.
Overall Awesome Entry to the AI and Scientific Compute market!

GitHub Link to the original Chinese Manual: (https://github.com/SKARN-eng/SKARN-Open-Source-Project-/blob/main/Chinese%20Dual%20SXM2%20NVLink%20Adapter%20Manual.pdf)

$250 Board no PLX-8749 PCIe Switch IC





$350-400 Board with PLX-8749 PCIe Switch IC

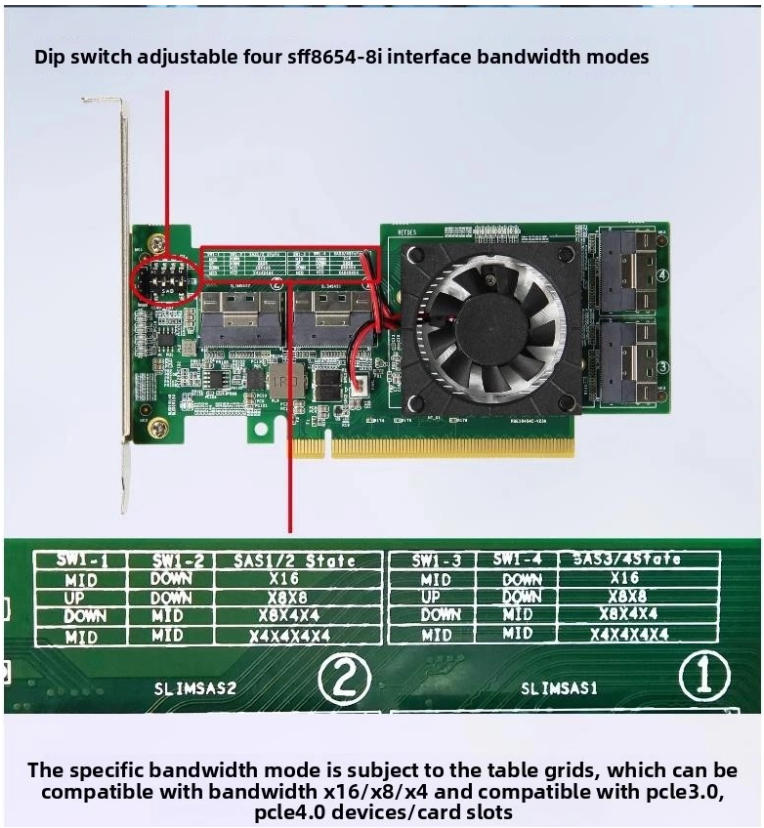
PCIe SFF-8654 Adapters for Chinese NVLink Adapter Boards
Status: Plentiful and Cheap $15-$130 Enterprise Solutions $100-$1000

The enterprise PCIe SFF adapters that use various PEX/PLX Switches for storage and PCIe. These are a bit more expensive but you will know 100% you are getting the best PCIe signal retention and switching speeds.
Enterprise 10GTek Link (https://www.10gtek.com/pcleswitchadapterforu.2ssd%28pexontroller%29)

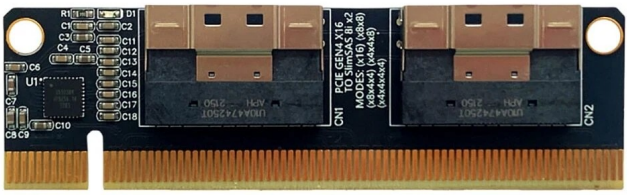
Chinese Adapters

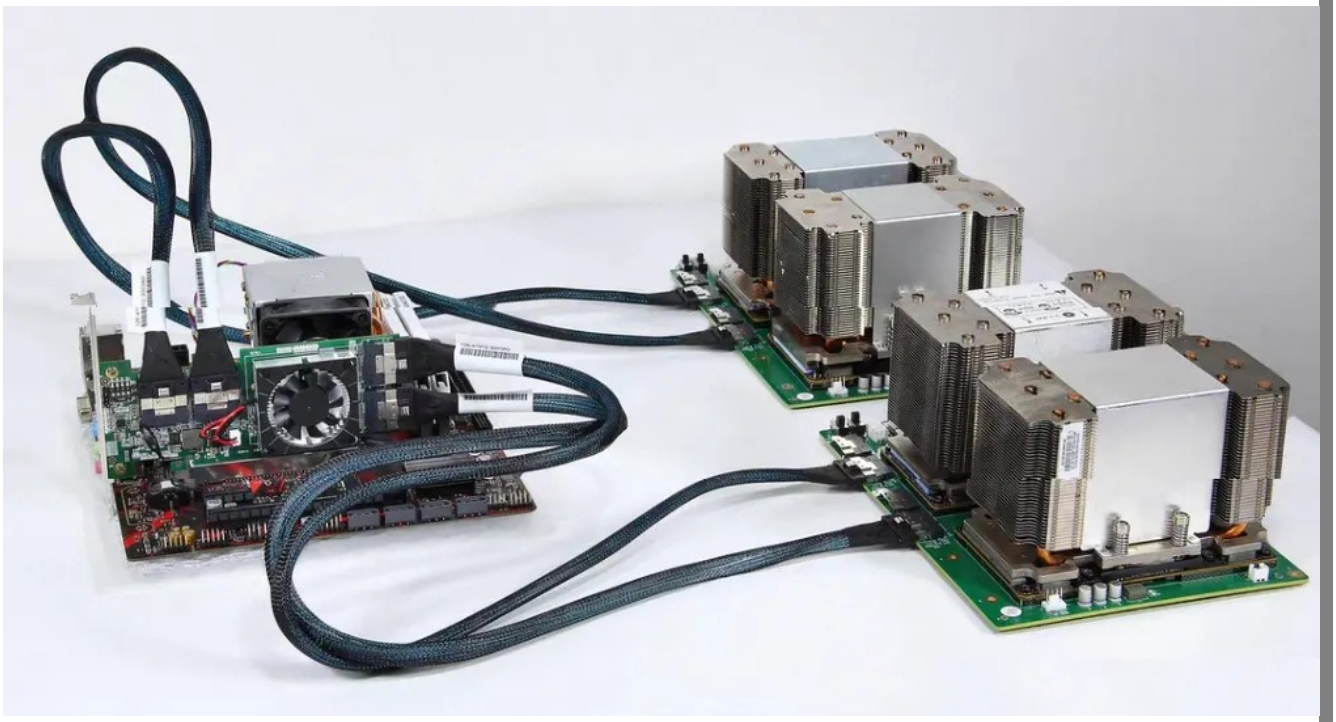
$120 Single PLX-8749 PCIe 3.0x16 to 4x SFF-8654 SlimSaS

Now on Amazon!
(https://www.amazon.com/PCIe-SFF-8654-Adapter-NVMe-PCIe4-0/dp/B0FGJ923MC?th=1)



$50 PCIe 3.0x16 2x SFF-8654 SlimSaS
Also on Amazon





Single PCIe 3.0, 4.0, and 5.0 to SXM2,4,5 Adapter Boards
Status: Plentiful and Cheap $20-$250 SXM2 $500-$1200 More Rare SXM4 and SXM5

These boards are interesting since they were the first Chinese adaptations early on during the Crypto boom and early AI innovations. These adapters comes from the simplest SXM2 adapter with PCIe dual 8 pin power 12vdc to SXM5 advanced PCB adapters.

$20-$100 Simple SXM2 Adapters

$500-$1200 SXM4 and SXM5 Single PCIe Adapter Boards

SXM4: Ampere Generation A100

SXM5: Hopper Generation H100.H200

Enterprise Solutions

New Old Stock Server
Status: Plentiful $800-$3000 SXM2

GIGABYTE T181-G20 1U DUAL SCALABLE 4x V100 SXM2 SOCKET SERVER
Ebay Link (https://ebay.us/m/MeHEK6)

Currently only $1200 for a whole server still in the box and arguably one of the best deals currently. I believe these are missing storage, cpu, and SXM2 V100's only. These SXM2 Trays are also cheap whenever they are found on Ebay or Amazon sub-$400. Ideally if this can be reconfigured to have all the GPU's on a single CPU, this would eliminate NUMA penalties if using CPU layer offloading.

Ewaste Tier Server Hardware

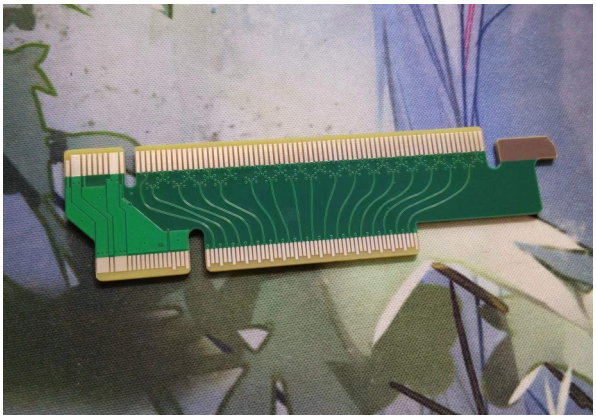
SXM2 Quad and Octa Trays
Status: Plentiful and Cheap $300-$2000
NO Brand Lock

$300-$800 Supermicro AOM-SXMV SXM2 Board (My Current Favorite)
This is an absolutely amazing board featuring No Brand Lock and also a simple CPU power solution. One limitation is that these boards have a Supermicro JPCie connection rather than simple PCIe 3.0x16.

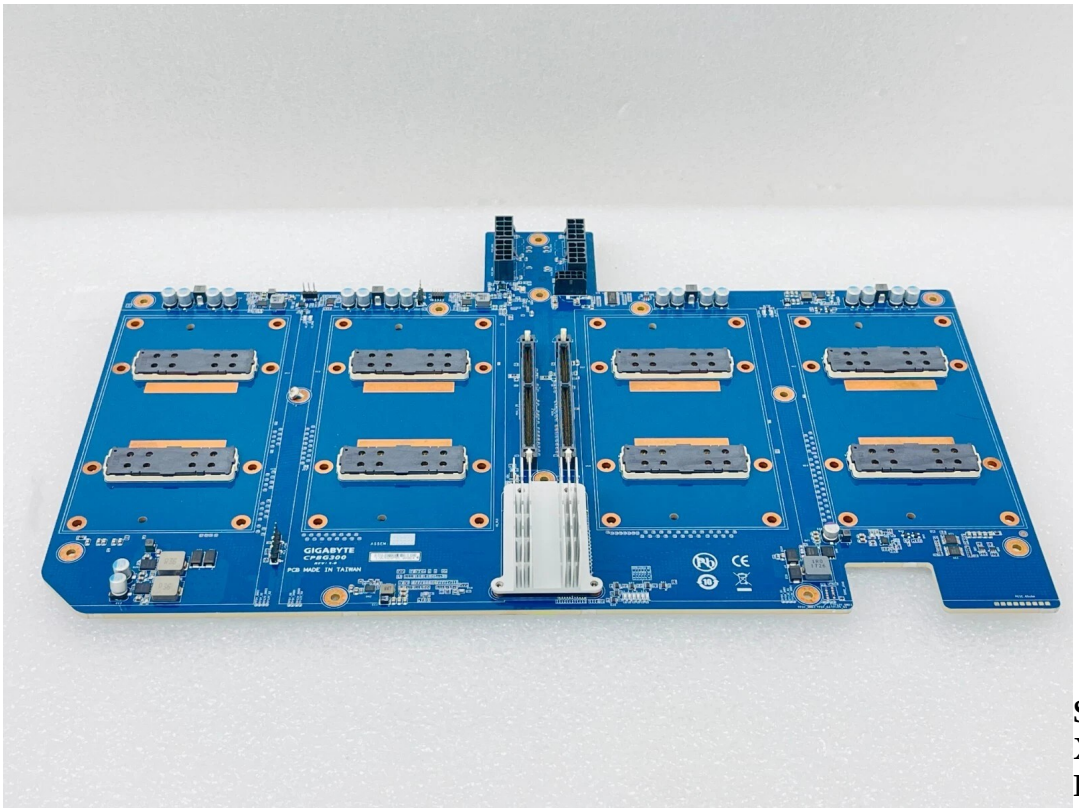
No Matter I have a solution for everyone.

You don’t need to make an account to view the drawings, download the BOM, Gerber File, Pinout Diagrams! This location also has an alternative version of this adapter that adapts JPCie to 2x SFF-8654.

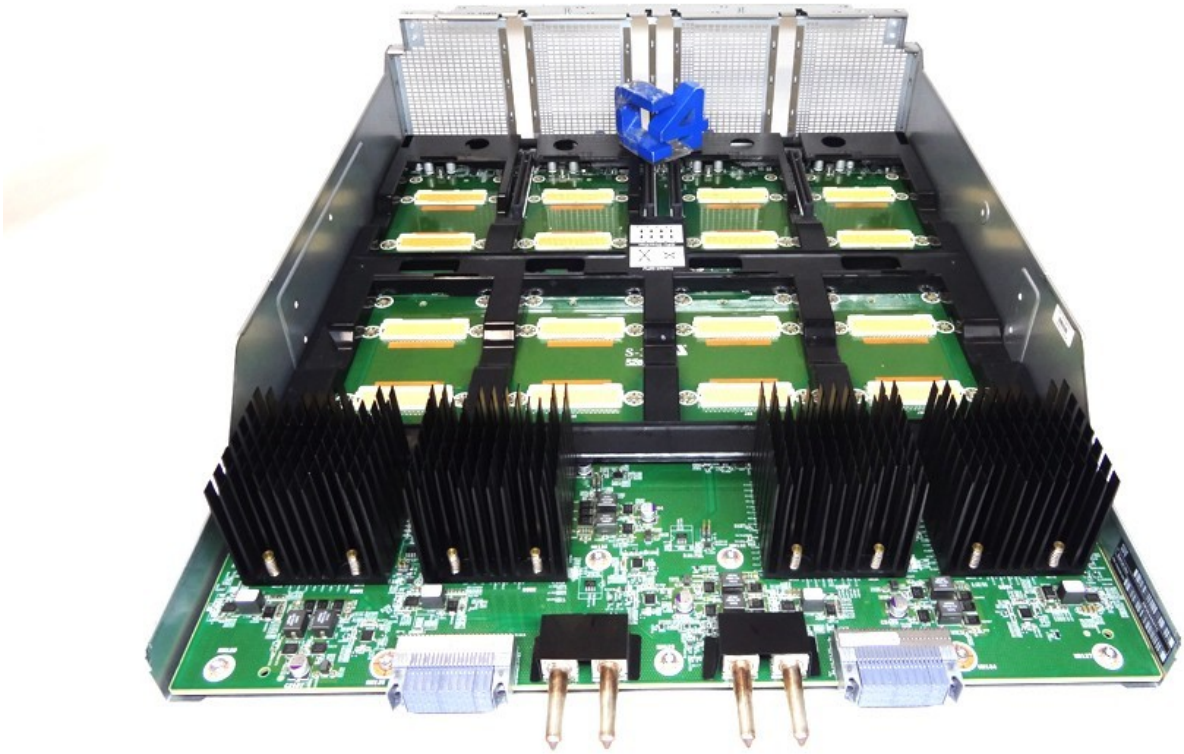
AOM-SXMV Adapter Chinese Gerber File (https://oshwhub.com/keiskeis/pcie-to-aom-sxmv-adapter-board)



$250-$500 GIGABYTE CPBG300 4x SXM2 SOCKET SERVER BOARD
Would be a good project for someone to get these boards figured out and adapters created. Would be a great AOM-SXMV alternative for a quad system. Likely Brand Locked.



$150-$600 XL270d HPE6500 Apollo 8x SXM2 Server Tray
Amazing value! I just purchased one on Ebay for $199 and will see if I can get this figured out for the community.



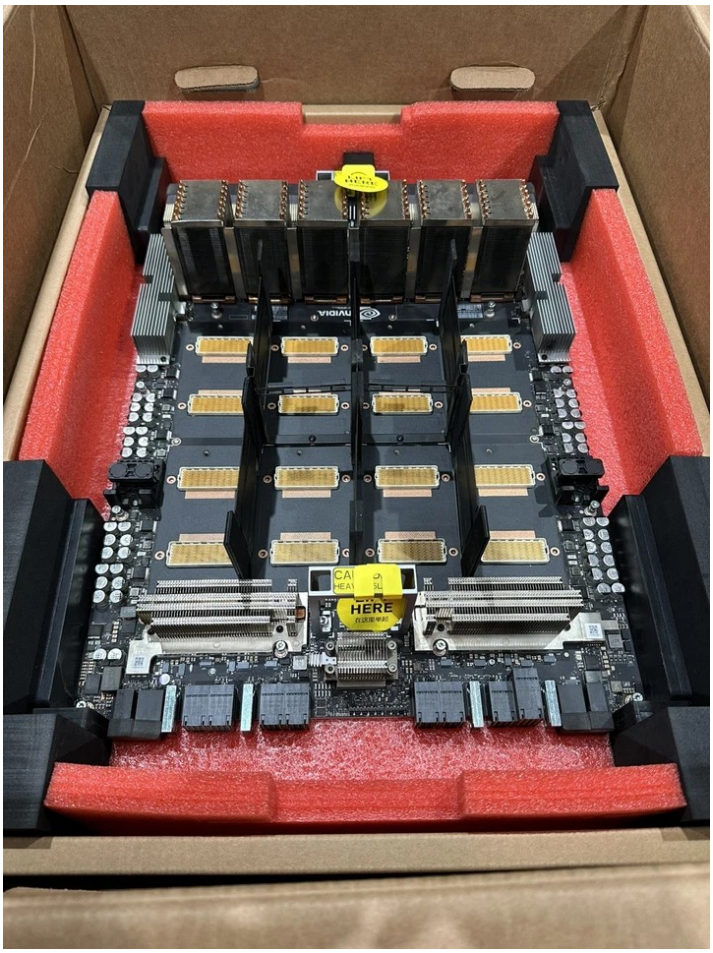
$3,000-$5,000 Dell NVLink SXM2 Board for C4140 C4130 96PWG SXM

I have heard of people using these in the past but only with Dell systems since there is likely a brand lock. Additionally these boards are still incredibly expensive, they will be a good project in the future if prices come down.

Nvidia Quad and Octa SXM4 and SXM5 Trays
Status: **Rare and Expensive $1,200(Broken)-$8,000**

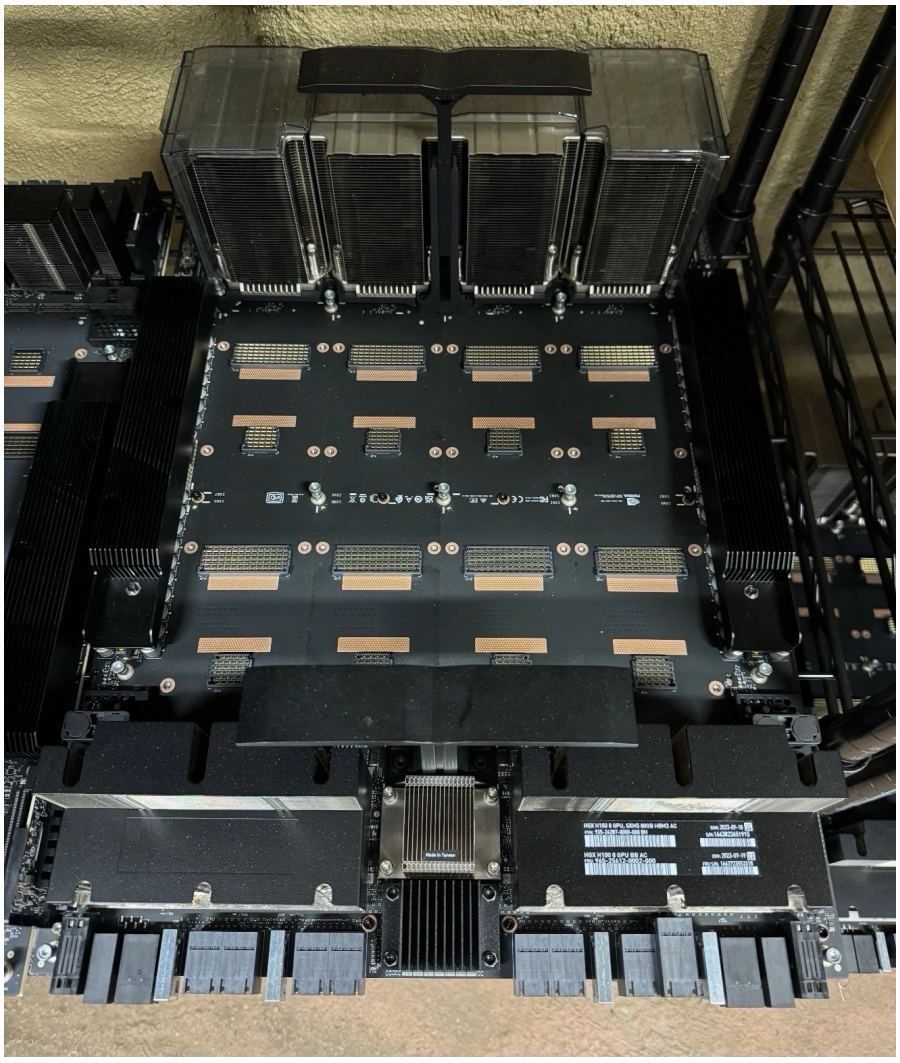
These boards are advanced Nvidia solutions for the high speed NVLink generations beyond the SXM3 and SXM2 socket. Starting with the Ampere architecture, the NVLink topology depends on NVSwitches that are a Nvidia secret. Thus stricter control over the manufacturing process of these SXM and SXM5 trays from the start. Hmm… if only someone could solder PCIe 5.0x16 pcb's to the underside of those massive ExaMAX connectors/ I hear these are 48VDC powered boards, not sure about a brand lock.

$8000 The Octa SXM4 A100 HGX Host Board

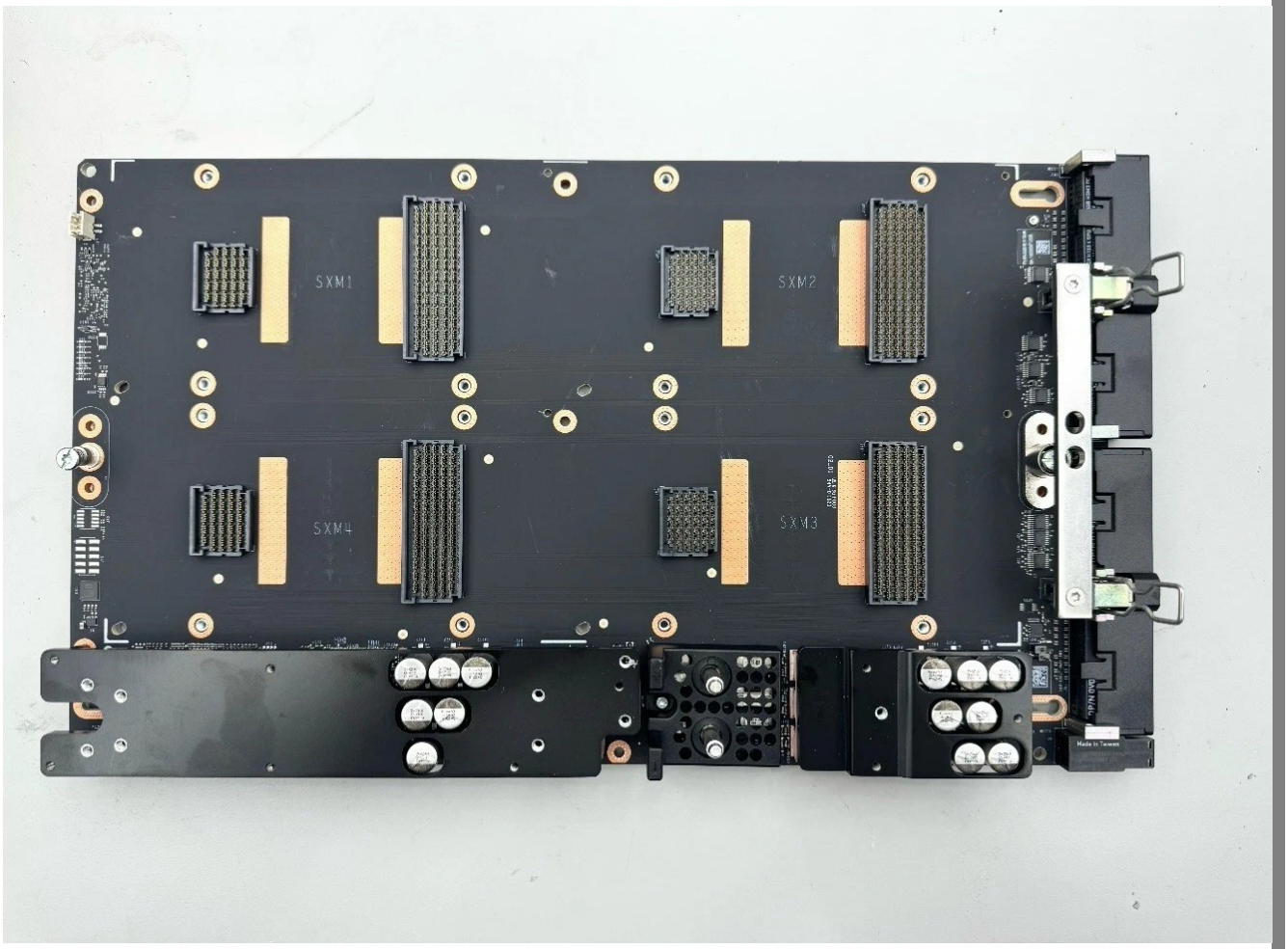


$800-$2000 Quad SXM4 Tray (Slight Damaged from Data-center Installs)

$1400 SXM5 Hopper Architecture HGX Tray (Broken)



$10,000-$20,000 Dell XE8640 4X GPU H100 80GB SXM5 Baseboard V0W6P PN 935-23087-0101-000 (Used)



License CERN

SPDX-License-Identifier: CERN-OHL-P
(c) Project SKARN- Shared Knowledge of Adaptable Research Nodes

1 Definitions

1.1 'Licence' means this CERN-OHL-P.

1.2 'Source' means information such as design materials or digital code which can be applied to Make or test a Product or to prepare a Product for use, Conveyance or sale, regardless of its medium or how it is expressed. It may include Notices.

1.3 'Covered Source' means Source that is explicitly made available under this Licence.

1.4 'Product' means any device, component, work or physical object, whether in finished or intermediate form, arising from the use, application or processing of Covered Source.

1.5 'Make' means to create or configure something, whether by manufacture, assembly, compiling, loading or applying Covered Source or another Product or otherwise.

1.6 'Notice' means copyright, acknowledgement and trademark notices, references to the location of any Notices, modification notices (subsection 3.3(b)) and all notices that refer to this Licence and to the disclaimer of warranties that are included in the Covered Source.

1.7 'Licensee' or 'You' means any person exercising rights under this Licence.

1.8 'Licensor' means a person who creates Source or modifies Covered Source and subsequently Conveys the resulting Covered Source under the terms and conditions of this Licence. A person may be a Licensee and a Licensor at the same time.

1.9 'Convey' means to communicate to the public or distribute.

2 Applicability

2.1 This Licence governs the use, copying, modification, Conveying of Covered Source and Products, and the Making of Products. By exercising any right granted under this Licence, You irrevocably accept these terms and conditions.

2.2 This Licence is granted by the Licensor directly to You, and shall apply worldwide and without limitation in time.

2.3 You shall not attempt to restrict by contract or otherwise the

rights granted under this Licence to other Licensees.

2.4 This Licence is not intended to restrict fair use, fair dealing, or any other similar right.

3 Copying, Modifying and Conveying Covered Source

3.1 You may copy and Convey verbatim copies of Covered Source, in any medium, provided You retain all Notices.

3.2 You may modify Covered Source, other than Notices.

You may only delete Notices if they are no longer applicable to the corresponding Covered Source as modified by You and You may add additional Notices applicable to Your modifications.

3.3 You may Convey modified Covered Source (with the effect that You shall also become a Licensor) provided that You:

a) retain Notices as required in subsection 3.2; and

b) add a Notice to the modified Covered Source stating that You have modified it, with the date and brief description of how You have modified it.

3.4 You may Convey Covered Source or modified Covered Source under licence terms which differ from the terms of this Licence provided that You:

a) comply at all times with subsection 3.3; and

b) provide a copy of this Licence to anyone to whom You Convey Covered Source or modified Covered Source.

4 Making and Conveying Products

You may Make Products, and/or Convey them, provided that You ensure that the recipient of the Product has access to any Notices applicable to the Product.

5 DISCLAIMER AND LIABILITY

5.1 DISCLAIMER OF WARRANTY -- The Covered Source and any Products are provided 'as is' and any express or implied warranties, including, but not limited to, implied warranties of merchantability, of satisfactory quality, non-infringement of third party rights, and fitness for a particular purpose or use are disclaimed in respect of any Source or Product to the maximum extent permitted by law. The Licensor makes no representation that any Source or Product does not or will not infringe any patent, copyright, trade secret or other

proprietary right. The entire risk as to the use, quality, and performance of any Source or Product shall be with You and not the Licensor. This disclaimer of warranty is an essential part of this Licence and a condition for the grant of any rights granted under this Licence.

5.2 EXCLUSION AND LIMITATION OF LIABILITY -- The Licensor shall, to the maximum extent permitted by law, have no liability for direct, indirect, special, incidental, consequential, exemplary, punitive or other damages of any character including, without limitation, procurement of substitute goods or services, loss of use, data or profits, or business interruption, however caused and on any theory of contract, warranty, tort (including negligence), product liability or otherwise, arising in any way in relation to the Covered Source, modified Covered Source and/or the Making or Conveyance of a Product, even if advised of the possibility of such damages, and You shall hold the Licensor(s) free and harmless from any liability, costs, damages, fees and expenses, including claims by third parties, in relation to such use.

6 Patents

6.1 Subject to the terms and conditions of this Licence, each Licensor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section 6, or where terminated by the Licensor for cause) patent licence to Make, have Made, use, offer to sell, sell, import, and otherwise transfer the Covered Source and Products, where such licence applies only to those patent claims licensable by such Licensor that are necessarily infringed by exercising rights under the Covered Source as Conveyed by that Licensor.

6.2 If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Covered Source or a Product constitutes direct or contributory patent infringement, or You seek any declaration that a patent licensed to You under this Licence is invalid or unenforceable then any rights granted to You under this Licence shall terminate as of the date such process is initiated.

7 General

7.1 If any provisions of this Licence are or subsequently become invalid or unenforceable for any reason, the remaining provisions shall remain effective.

7.2 You shall not use any of the name (including acronyms and abbreviations), image, or logo by which the Licensor or CERN is known, except where needed to comply with section 3, or where the use is otherwise allowed by law. Any such permitted use

shall be factual and shall not be made so as to suggest any kind of endorsement or implication of involvement by the Licensor or its personnel.

7.3 CERN may publish updated versions and variants of this Licence which it considers to be in the spirit of this version, but may differ in detail to address new problems or concerns. New versions will be published with a unique version number and a variant identifier specifying the variant. If the Licensor has specified that a given variant applies to the Covered Source without specifying a version, You may treat that Covered Source as being released under any version of the CERN-OHL with that variant. If no variant is specified, the Covered Source shall be treated as being released under CERN-OHL-S. The Licensor may also specify that the Covered Source is subject to a specific version of the CERN-OHL or any later version in which case You may apply this or any later version of CERN-OHL with the same variant identifier published by CERN.

7.4 This Licence shall not be enforceable except by a Licensor acting as such, and third party beneficiary rights are specifically excluded.