What is SKARN?

The project began in early March of 2024, as...

“I will make a basic AI workstation to learn stuff about this new AI Hype thing”

What proceeded was a yearlong endeavor to acquire the most powerful hardware possible at the lowest cost. From LGA-3647 dual socket to LGA-4189, and LGA-4677 based systems to dual AMD Mi50, P100, Dual 2080ti 22gb SLI. All of which fall into the same trap as the “Quad RTX 4090 AI 9000 MONSTER”.

Systems that are lack NVLink or InfiniBand cannot effectively pool their video memory. The Quad RTX 4090 system essentially has four separate VRAM domains that rely on the following communication path; when running AI models that spread layers across multiple PCIe GPU’s.

The comparison between a commercial AI workstation communication to Enterprise NVLink

GPU1 → PCIe Bus → CPU (RAM Pool/PCIe NVME-M.2) → PCIe Bus → GPU2

vs

(GPU1 → GPU2)=>PCIe Bus → CPU (RAM Pool/PCIe NVME-M.2)

Usually developers will opt to run separate models on individual GPU’s which are often limited in VRAM capacity. For example our RTX 4090 only has 24gb of GDDR6X at a premium price and

limited quantity. The newest RTX Pro Blackwell 6000 features 96gb of GDDR7 at an eye-watering price-point.

Don’t even think about buying a A100 PCIe or even any of the Hopper PCIe model GPU’s for your LLM’s or scientific compute projects. You will spend $8k for a used datacenter A100 that has been converted from SXM4 to PCIe as the cheapest option.

The more cost effective option is to run some layers on a commercial GPU while offloading all other layers to the CPU. Yes the system RAM pool can be absolutely massive based on CPU Architecture but CPU offload has a drastic drop in overall throughput performance compared to GPU’s.

There has to be a better way.

Today’s AI data-center boom is a treasure trove of SXM based systems and complementary components. Enterprise processors and motherboards are arriving on popular resale sites such as Ebay, at very reasonable prices. In addition to this, server power supplies are dirt cheap especially for blade redundant server PSU’s.

The destiny for the majority of this enterprise hardware is either a landfill or to run Arch Linux in a basement somewhere, developing the next 90% discounted Indie game on Steam.

What is keeping us from using enterprise hardware?

Currently enterprise hardware is trapped behind a wall of enterprise gate keeping. For example the Super-micro AOM-SXMV SXM2 NVLink host board is an incredible piece of engineering. This board originally was part of the Supermicro 1028GQ-TXR(T) Super-server system. A dual socket Xeon system that either used V100’s or P100’s.
AOM-SXM2 = P100 Only
AOM-SXMV =V100 Definitely...maybe the P100..?

The issue arises when trying to interface with the proprietary JPCIe connector on the AOM boards. This is a Super-micro proprietary connector that requires the “AOM-SXM-BRG”.

According to page 1-5 of Chapter 1 of Super-Micro’s Superserver 1028GQ-TXR(T) 1028GQ-TVR(T) manual. **Found here (https://www.supermicro.com/manuals/superserver/1U/MNL-1931.pdf)**

The 1028GQ-TVR(T) system supports four Volta SXM V100 GPUs installed on the AOM-SXMV add-on module which is connected to the motherboard by two bridges.

A direct connection between all GPUs is a double NVlink connection (2x25 GB/s). Fastest connections are afforded when GPUs are added in pairs.

A direct connection from the GPUs to the network is provided using the OCuLink cable from the add-on module connector to the riser card connector. A fast network (expansion) card installed on the riser card affords very high speeds. With OCuLink cables connected, data can go from GPU to the PLX to the NIC, bypassing the CPU. For the NIC, the system can support both FDR and EDR (in x16 slots).

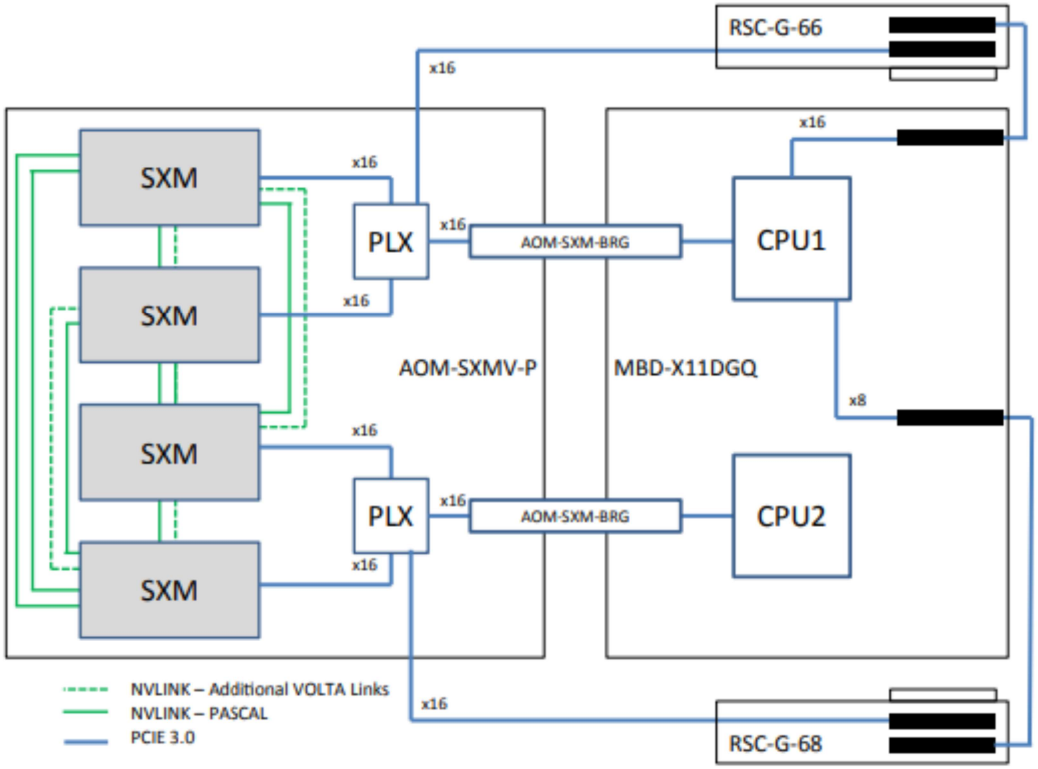
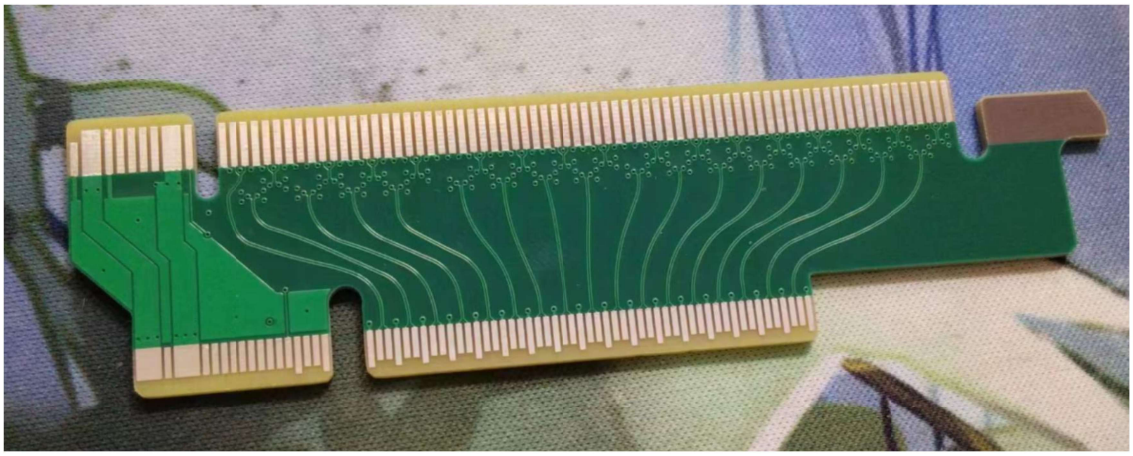


Figure 1-3. SXMV Add-on Module Block Diagram

In reality this simple JPCIe to PCIe bridge connectors that spans roughly five or so inches in length has almost no pictures of the part let alone parts for sale. I found some of these bridge boards sold by an Enterprise hardware re-selling company fors $900 each.

So a Chinese developer by the name
地摊垃圾佬-锂离子 or in English **“Street vendor garbage collector - lithium ion”**
made the following adapter. Now a PCIe3.0x16 Riser Cable can interface between two of these adapters and the almost any motherboard someone can purchase even X99 or below.

Original Link found here (https://oshwhub.com/keiskeis/pcie-to-aom-sxmv-adapter-board)
The original page contains Gerber files, BOM, PCB layouts that can be downloaded without an account.



This project intends to find online open source solutions that already exist in the case of the AOM-SXMV and also make adaptations or changes to other obsolete and discarded enterprise SXM trays.

The Four Pillars of SKARN

- **Hardware flexibility and aggressive cost effective solutions.**
- **Provide developers and educators with information on building advanced research compute nodes.**
- **Community of openly shared developer contributions for project progression.**
- **Reducing E-Waste by re-configuring obsolete or proprietary enterprise hardware.**

The printing press, metal lathe, voltaic pile battery, or even the typewriter were transformative. When these inventions became affordable or could be replicated, the average citizen of the world could leverage capabilities only Universities or Government Institutions had prior.

AI and Scientific Computing is not an insurmountable barrier for the average person let alone educators and professors in developing countries. These hardware limitations can be overcome with the builder community and educators openly sharing information.

“Great innovations, whether in art or literature, in science or in nature, seldom take the world by storm. They must be understood before they can be estimated, and must be cultivated before they can be understood.”

CLARENCE EDWARD DUTTON (winter-1880)

License CERN

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.