Open. Together.

OCP
SUMMIT

SERVER

Datacenter-ready Secure Control Module and Interface

for Modular Building Block Architecture (MBA, *The Catalyst*)

Siamak Tavallaei, Principal Architect
Microsoft, Azure

OPEN
PLATINUM™

Outline

- Motivation
- Approach
- Examples
- Requesting Participation and Feedback

SERVER

Specifications

TTM

- You have developed a production-ready server with a PCIe Slot.
- A solution provider has built a fully-compatible PCIe Add-in-Card with a unique feature.
- How long will it take you to integrate that PCIe AIC into your Server?

TTM

- You have developed a solution (e.g., a new server)
- Your CSP customer learns about it and likes it so much that wants to deploy it into the Datacenter in 3 months?
- How would you do that?

Modular Building Block Architecture (MBA)

- Small building blocks allow flexible and agile system integration to meet the needs of various use cases

- Clearly defined input/output ports for interoperability with CPU boards from various suppliers

- Connector/Cable-based IO Slots offer flexible choice of high-speed IO for AICs

DC-SCM Facilitates MBA

MBA is a *Catalyst* for interoperable *Innovation*!

A commonly designed secure control module for datacenters, ***DC-SCM***, enables the design and deployment of CPU/Memory Complexes and Expansion Chassis based on guidelines from CPU and SoC suppliers to become simply a **routine exercise**!

Modular in every way!

“The conductor and the musician interpreting the composer’s work”
Vincent Van Gogh

Architectural Specification
Detailed Design
Product Manufacturing

Modular in every way!

“The conductor and the musician interpreting the composer’s work”
Vincent Van Gogh

Mechanical Design
Electrical Design
Software/Firmware Design
Security and Management
Debug

Examples of

Modular Building Block Architecture

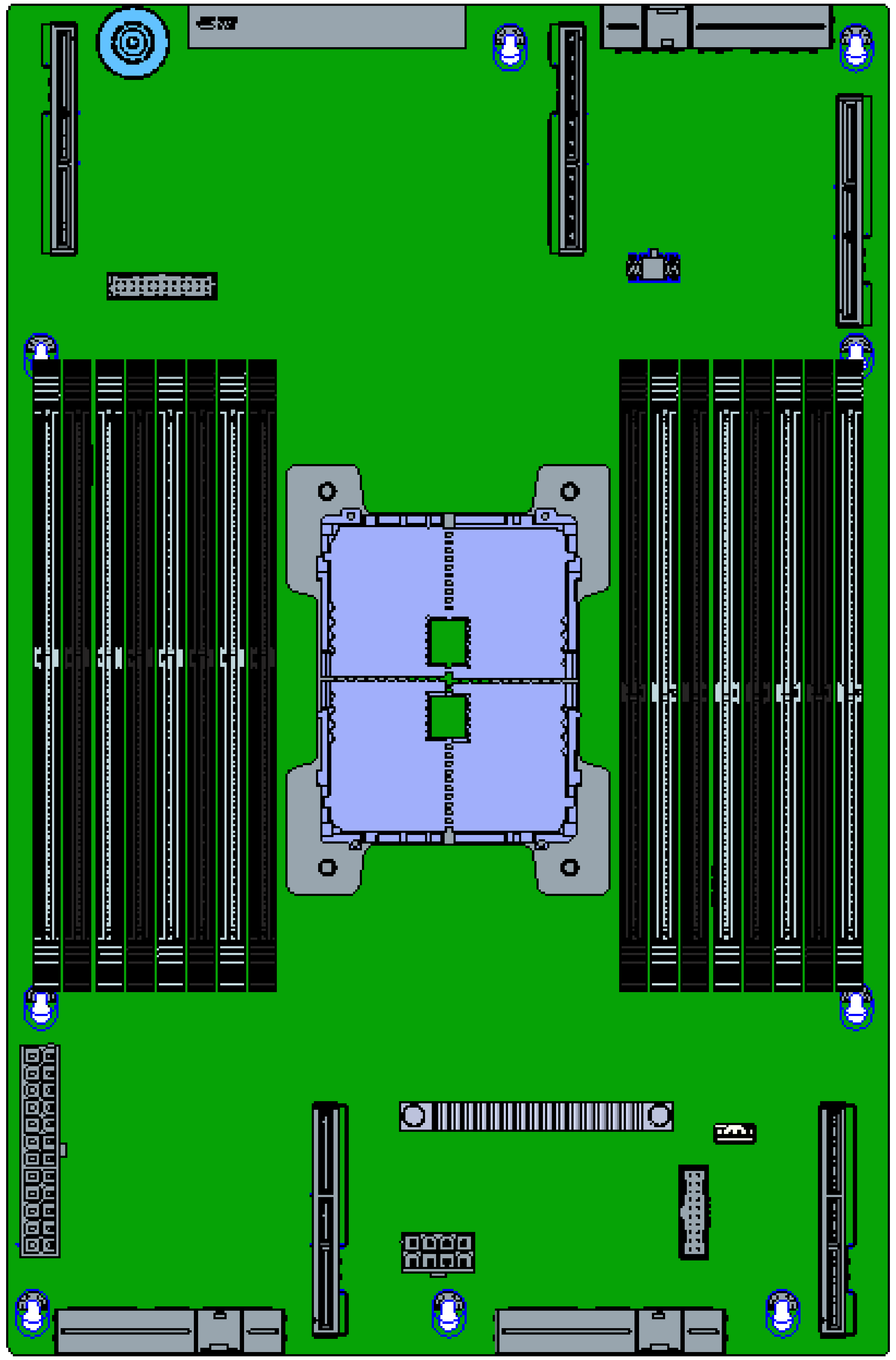
For a successful Modular Building Block Architecture, we need:

- Compute Modules (CPU/Memory/IO) (**CMIOM**)
- IO & Accelerator Add-in Card Modules (**AIC**)
- Security, Control, and Management (**SCM**)
- Data-plane Control
- An Interconnect

CPU/Mem/IO Module

Just the essential Central Compute Elements

And high-speed Memory and IO Connectors

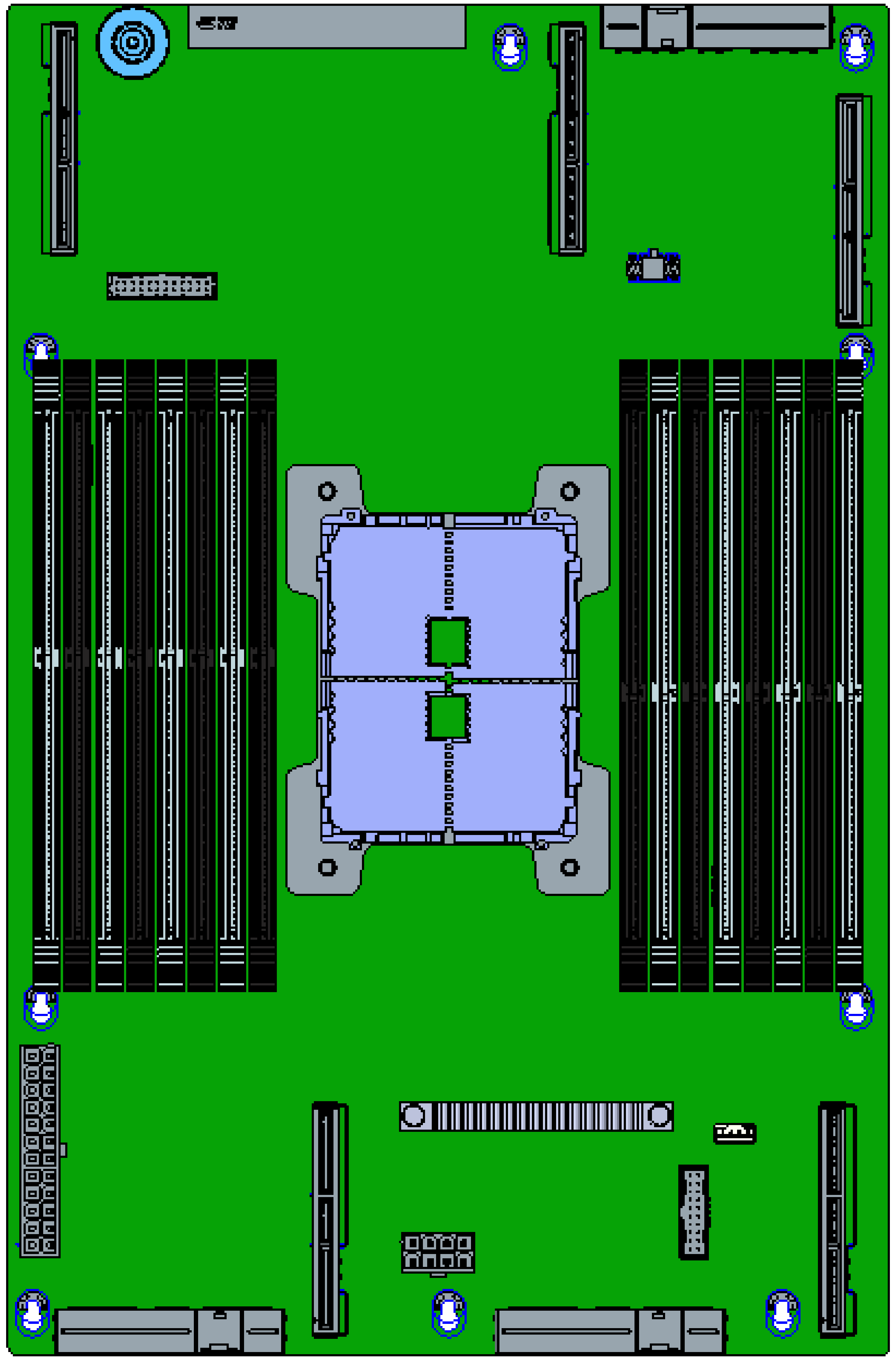


CPU/Mem/IO Module

Just the essential Central Compute Elements

And high-speed Memory and IO Connectors Close to the SoC

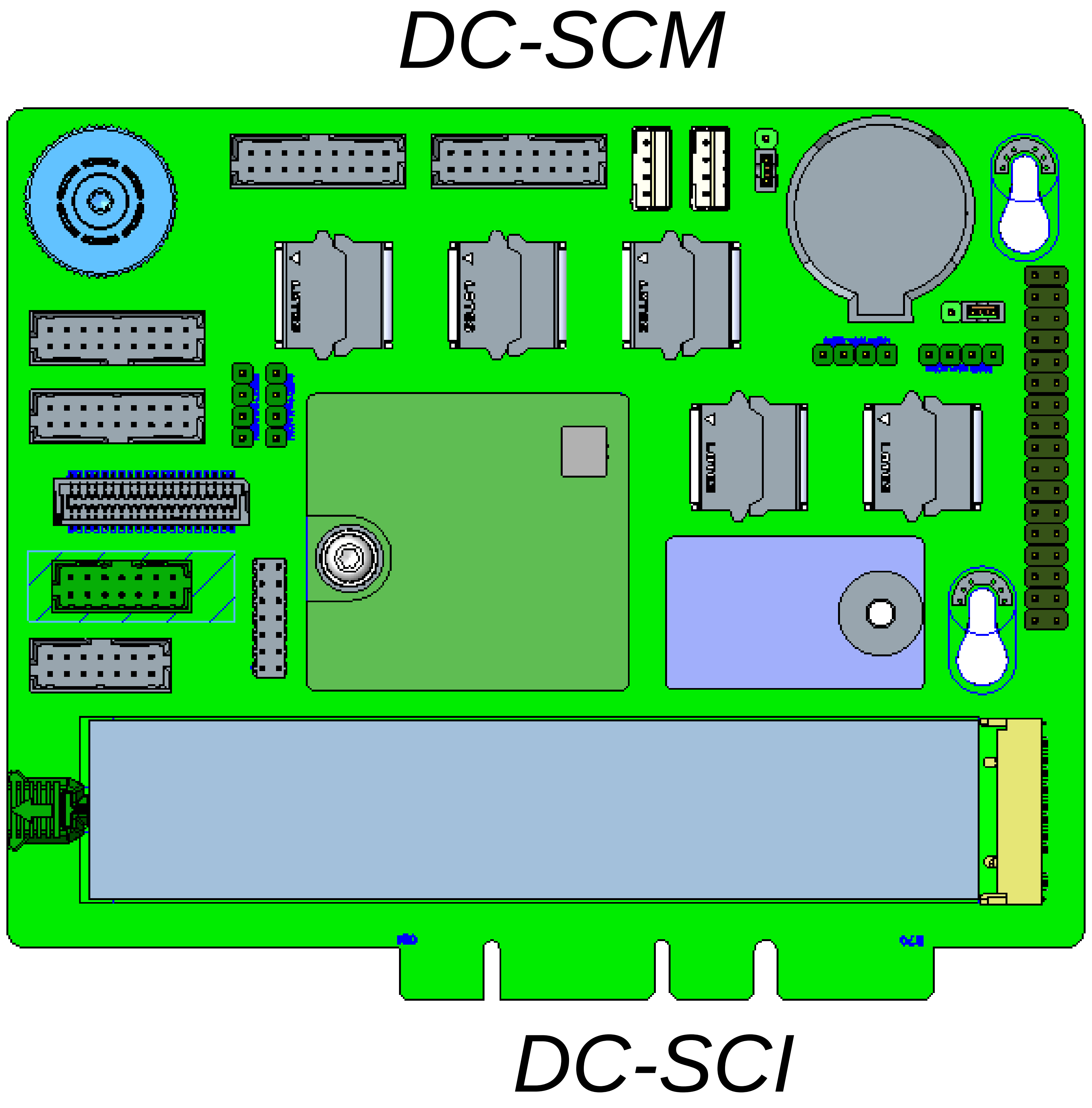
Get ready for *PCIe Gen-5!*



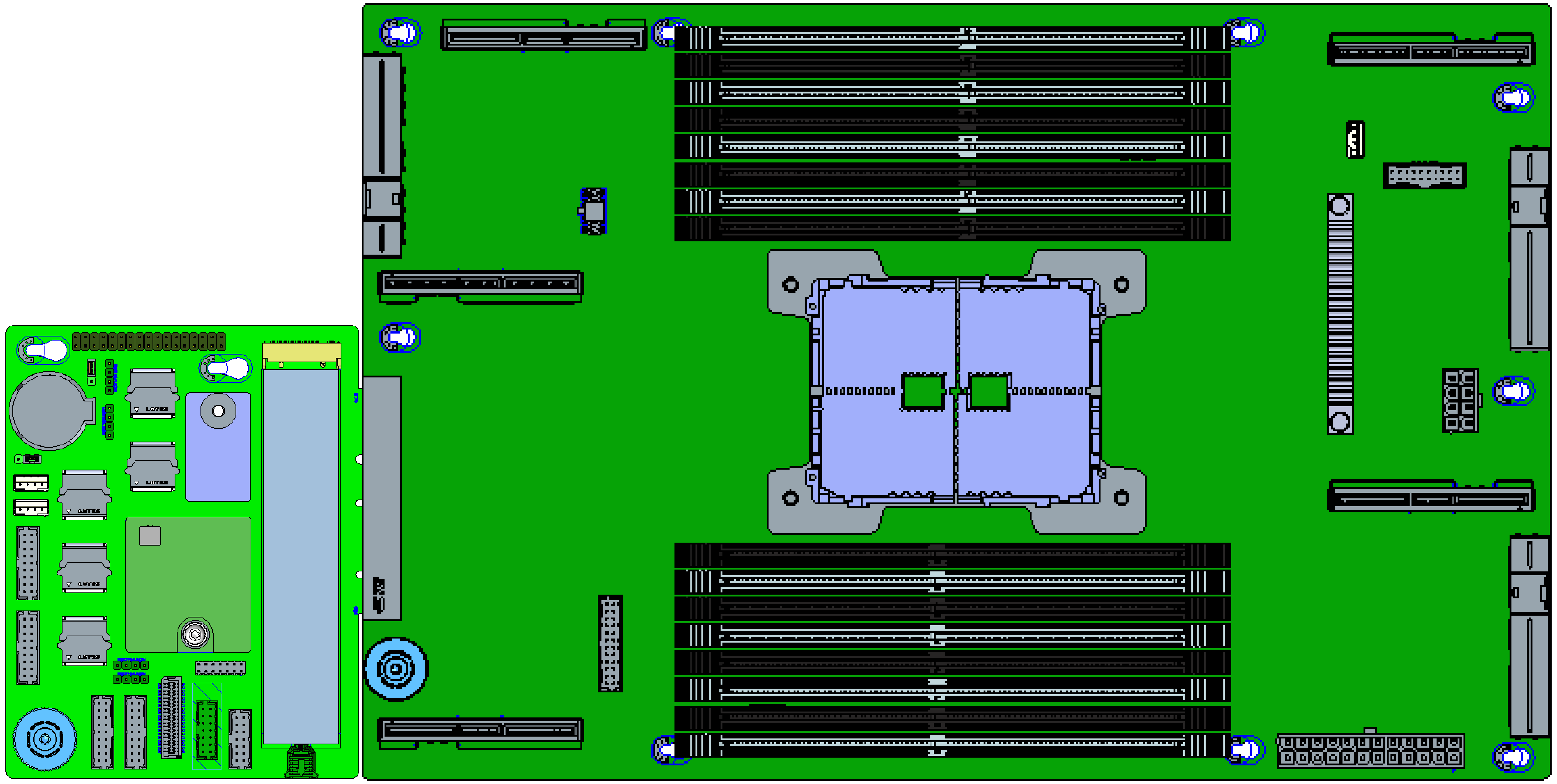
DC-SCM

Everything Else!

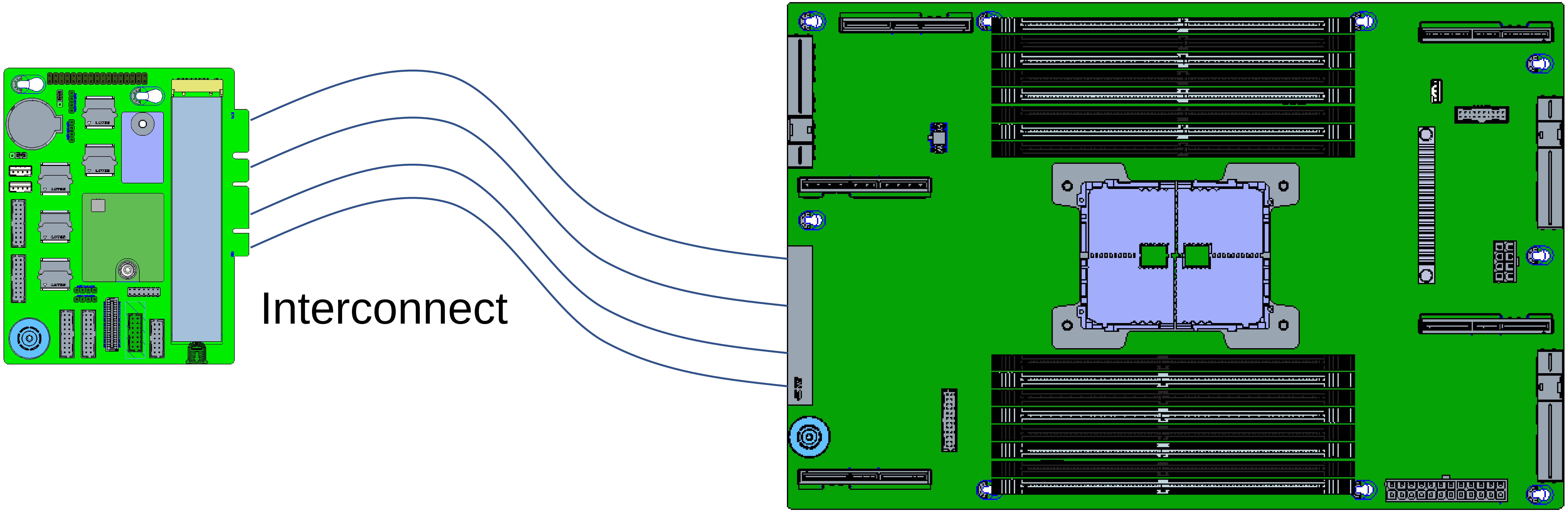
Security, Control, Management



DC-SCM + CPU/Mem/IO

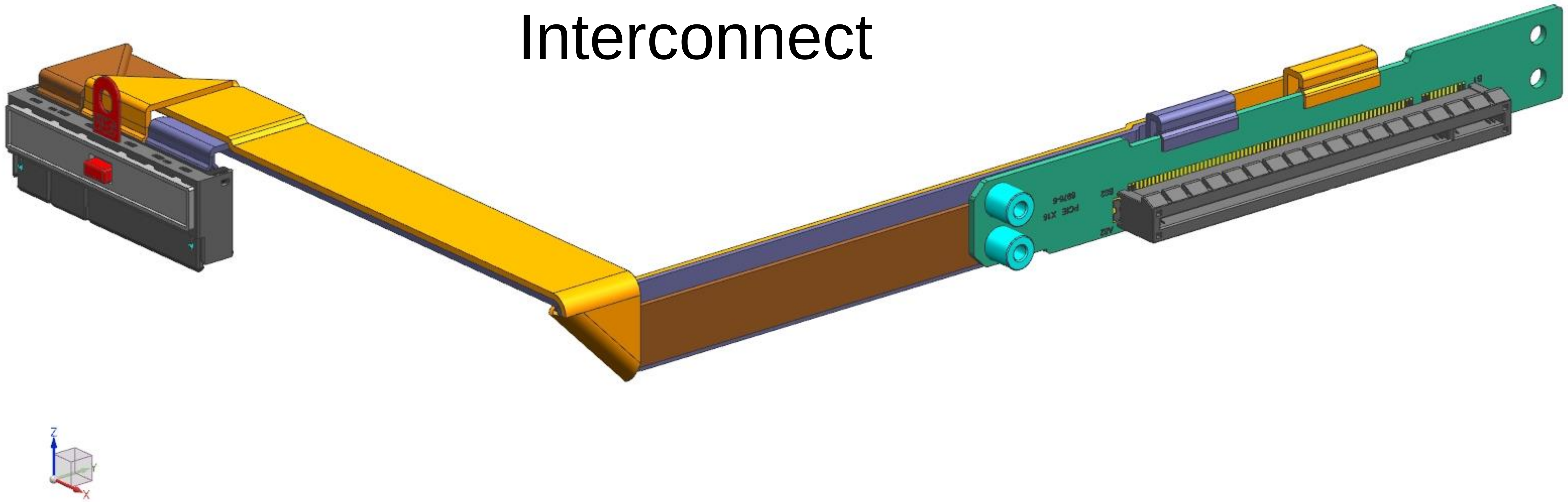


DC-SCM + CPU/Mem/IO



AIC Attachment

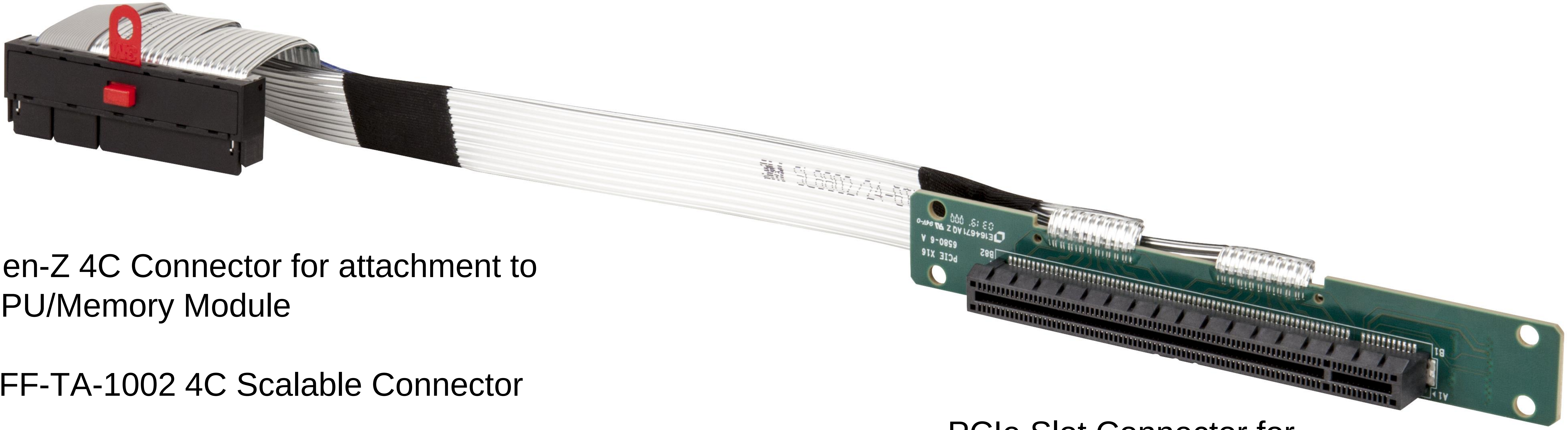
IO Slot to CPU Board Cable Harness



Interconnect

AIC Attachment

IO Slot to CPU Board Cable Harness



Gen-Z 4C Connector for attachment to CPU/Memory Module

SFF-TA-1002 4C Scalable Connector

PCIe Slot Connector for
an Add-in Card in
PCIe CEM form factor

PCIe Slot to Gen-Z Pin Map

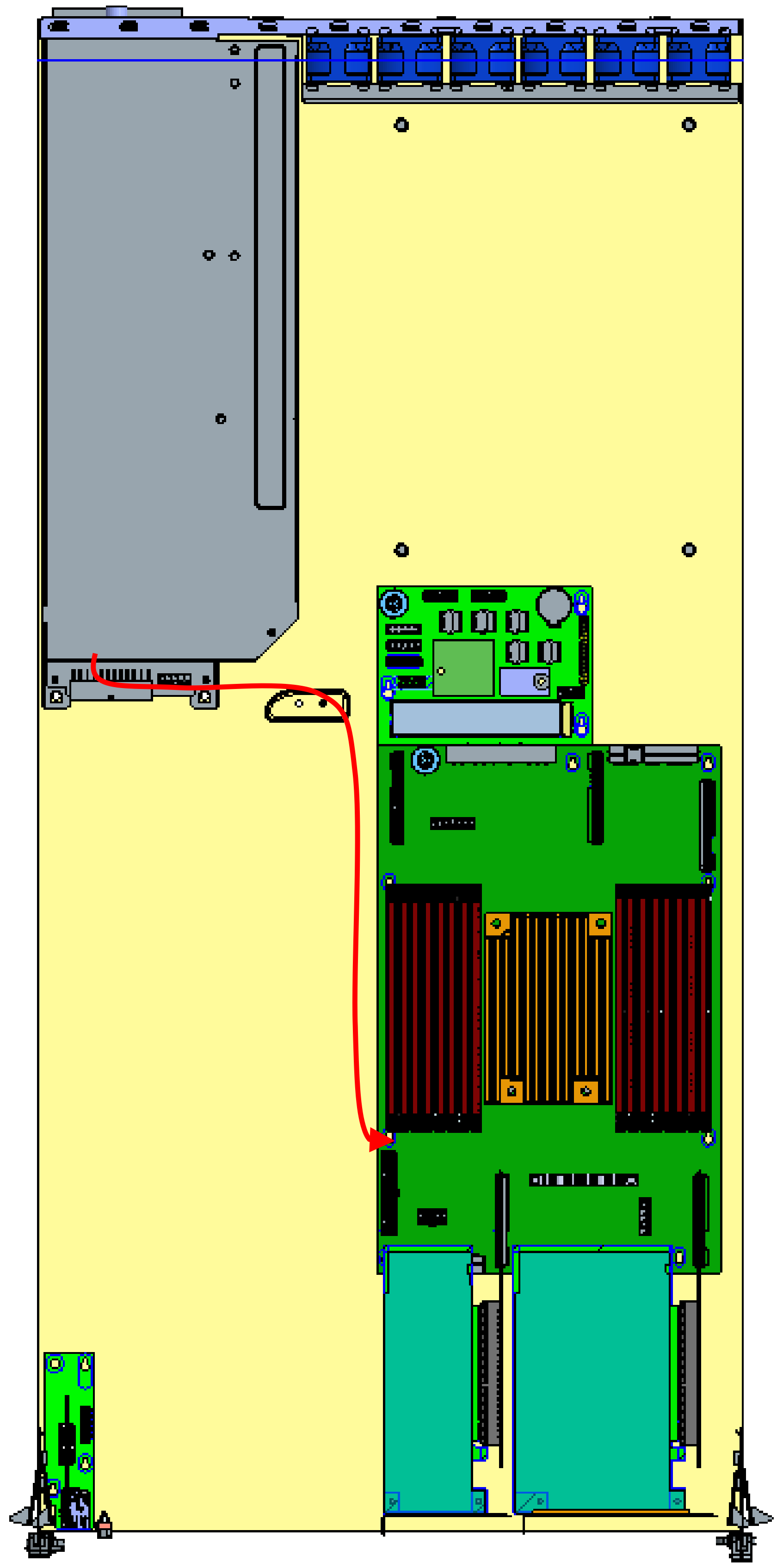
ASSEMBLY PINOUT TABLE		
PCIe Side-A		Gen-Z Side-A
P1	Description	P2
A1	PRSNT_1	A1
A2, A3	P12V	B1/B2/B3/B4/B5/B6
A4	GND	A5
A5	JTAG2	A42
A6	JTAG3	A2
A7	JTAG4	A3
A8	JTAG5	A4
A9, A10	P3.3V	A69/B68/B69
A11	PWRGD	B10
A12	GND	A6
A13	REFCLK_P	B15
A14	REFCLK_N	B14
A15	GND	A13
A16	HSIN_0 (RX)	A17
A17	HSIP_0 (RX)	A18
A18	GND	A16
A19	NC_RSVD_1	NC
A20	GND	A19
A21	HSIN_1(RX)	A20
A22	HSIP_1(RX)	A21
A23, A24	GND	A22
A25	HSIN_2(RX)	A23
A26	HSIP_2(RX)	A24
A27, A28	GND	A25
A29	HSIN_3(RX)	A26
A30	HSIP_3(RX)	A27
A31	GND	A28
A32	NC_RSVD_2	NC
A33	NC_RSVD_3	NC
A34	GND	A29
A35	HSIN_4(RX)	A30
A36	HSIP_4(RX)	A31
A37, A38	GND	A32
A39	HSIN_5(RX)	A33
A40	HSIP_5(RX)	A34
A41, A42	GND	A35
A43	HSIN_6(RX)	A36
A44	HSIP_6(RX)	A37
A45, A46	GND	A38
A47	HSIN_7(RX)	A39
A48	HSIP_7(RX)	A40
A49	GND	A41
A50	NC_RSVD_5	NC
A51	GND	A43
A52	HSIN_8(RX)	A44
A53	HSIP_8(RX)	A45
A54, A55	GND	A46
A56	HSIN_9(RX)	A47
A57	HSIP_9(RX)	A48
A58, A59	GND	A49
A60	HSIN_10(RX)	A50
A61	HSIP_10(RX)	A51
A62, A63	GND	A52
A64	HSIN_11(RX)	A53
A65	HSIP_11(RX)	A54
A66, A67	GND	A55
A68	HSIN_12(RX)	A56
A69	HSIP_12(RX)	A57
A70, A71	GND	A58
A72	HSIN_13(RX)	A59
A73	HSIP_13(RX)	A60
A74, A75	GND	A61
A76	HSIN_14(RX)	A62
A77	HSIP_14(RX)	A63
A78, A79	GND	A64
A80	HSIN_15(RX)	A65
A81	HSIP_15(RX)	A66
A82	GND	A67
NC	NC_MGMT_RST	A9
NC	NC_LED/ACTIVITY	A10
NC	NC_REFCLK1_P	A14
NC	NC_REFCLK1_N	A15

ASSEMBLY PINOUT TABLE		
PCIe Side-B		Gen-Z Side-B
P1	Description	P2
B1, B2, B3	P12V	B1/B2/B3/B4/B5/B6
B4	GND	GND
B5	SMCLK	A7
B6	SMDAT	A8
B7	GND	B13
B8	P3.3V	A69/B68/B69
B9	JTAG1	A68
B10	P3.3V_AUX	B11
B11	WAKE	A70
B12	CLKREQ	A11
B13	GND	B16
B14	HSON_0 (TX)	B17
B15	HSOP_0(TX)	B18
B16	GND	GND
B17	NC_PRSNT_2_B17	NC
B18	GND	B19
B19	HSON_1(TX)	B20
B20	HSOP_1(TX)	B21
B21, B22	GND	B22
B23	HSON_2(TX)	B23
B24	HSOP_2(TX)	B24
B25, B26	GND	B25
B27	HSON_3(TX)	B26
B28	HSOP_3(TX)	B27
B29	GND	B28
B30	PWRBRK	B8
B31	PRSNT_2_B31	A12
B32	GND	B29
B33	HSON_4(TX)	B30
B34	HSOP_4(TX)	B31
B35, B36	GND	B32
B37	HSON_5(TX)	B33
B38	HSOP_5(TX)	B34
B39, B40	GND	B35
B41	HSON_6(TX)	B36
B42	HSOP_6(TX)	B37
B43, B44	GND	B38
B45	HSON_7(TX)	B39
B46	HSOP_7(TX)	B40
B47	GND	B41
B48	PRSNT_2_B48	B42
B49	GND	B43
B50	HSON_8(TX)	B44
B51	HSOP_8(TX)	B45
B52, B53	GND	B46
B54	HSON_9(TX)	B47
B55	HSOP_9(TX)	B48
B56, B57	GND	B49
B58	HSON_10(TX)	B50
B59	HSOP_10(TX)	B51
B60, B61	GND	B52
B62	HSON_11(TX)	B53
B63	HSOP_11(TX)	B54
B64, B65	GND	B55
B66	HSON_12(TX)	B56
B67	HSOP_12(TX)	B57
B68, B69	GND	B58
B70	HSON_13(TX)	B59
B71	HSOP_13(TX)	B60
B72, B73	GND	B61
B74	HSON_14(TX)	B62
B75	HSOP_14(TX)	B63
B76, B77	GND	B64
B78	HSON_15(TX)	B65
B79	HSOP_15(TX)	B66
B80	GND	B67
B81	PRSNT_2_B81	B70
B82	GND	GND
NC	NC_MFG	B7
NC	NC_DUALPORTEN	B9
NC	NC_PWRDIS	B12

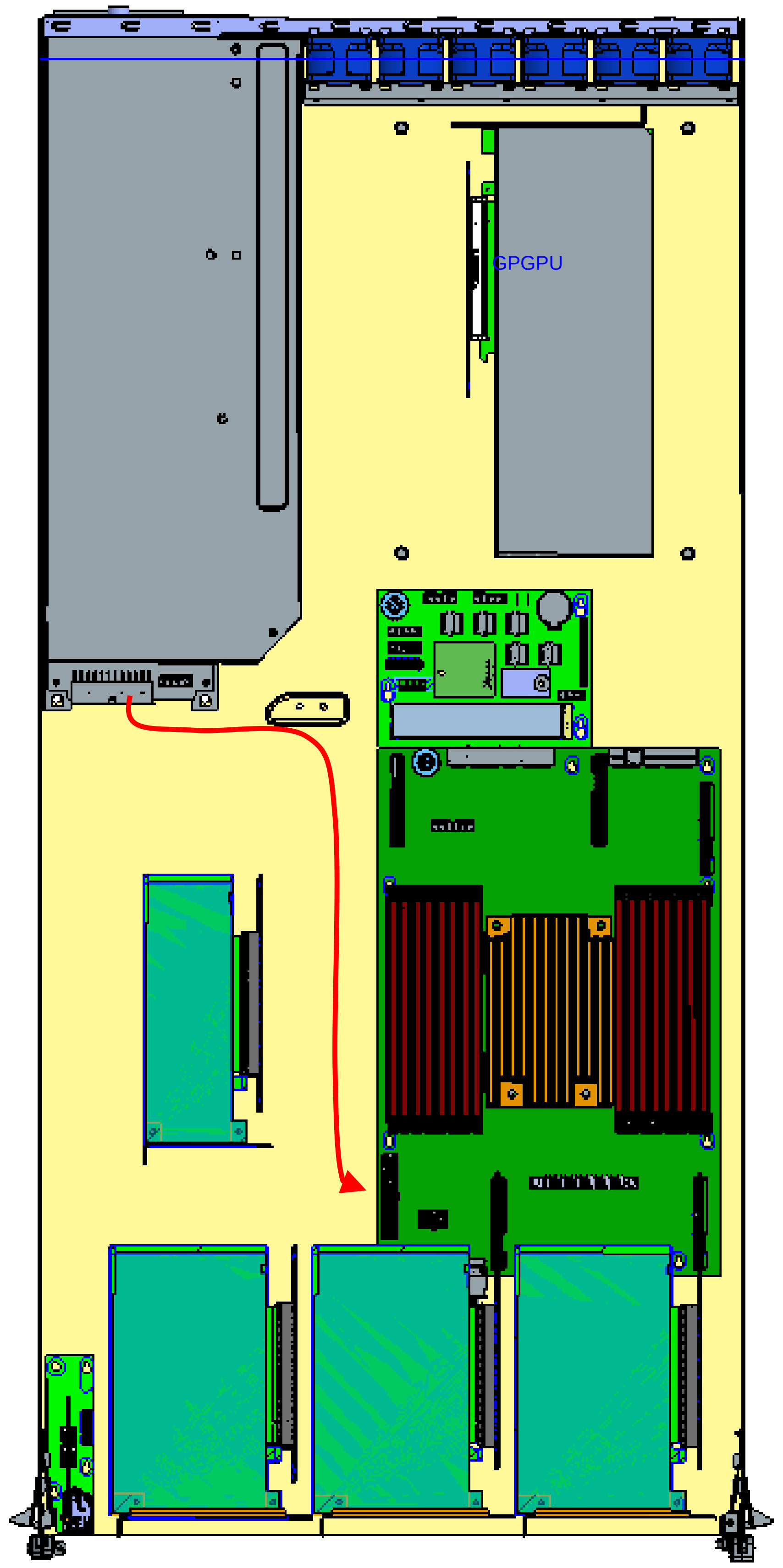
Ground pin	Zero volt reference, all tied together
Power pin	Supplies power to the card
High speed pin	High speed signals
Detect	Sense Pin
Other aux	May be pulled low or sensed by multiple cards
Reserved	Reserved for future use and no connect

Examples of Modular Designs

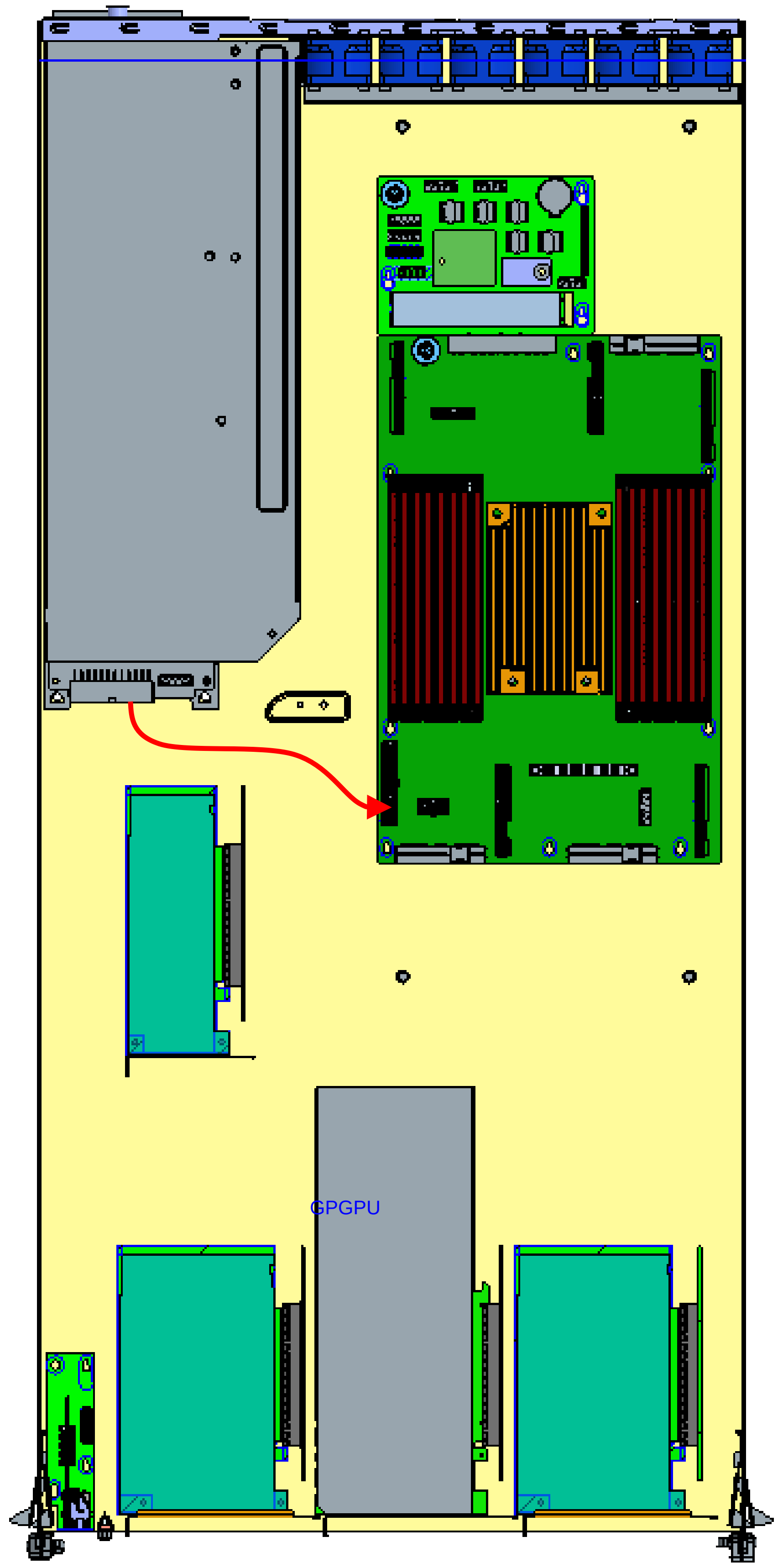
Two Riser Boards & Two AICs



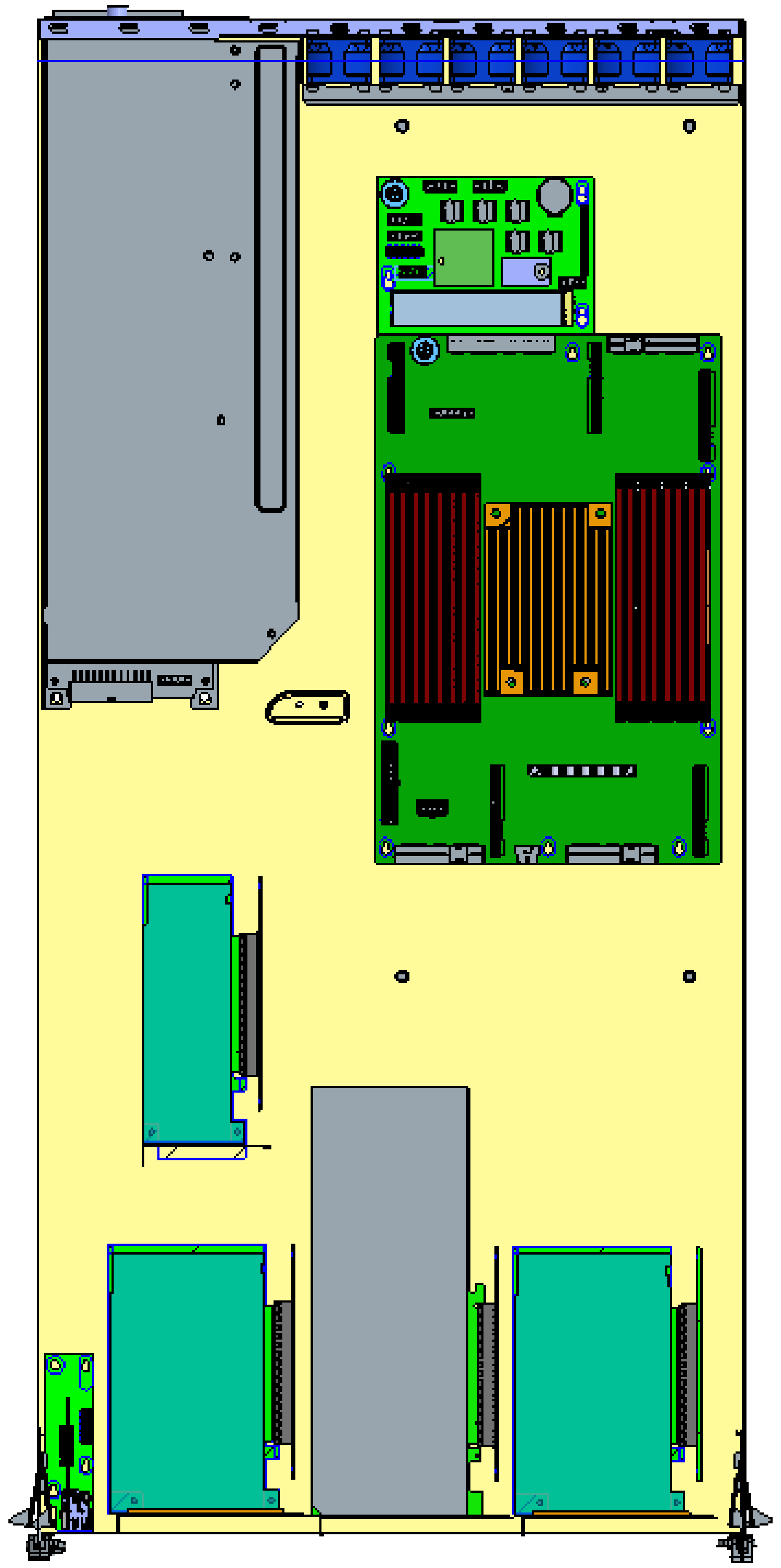
Six AICs (Two Risers, Four Cables)



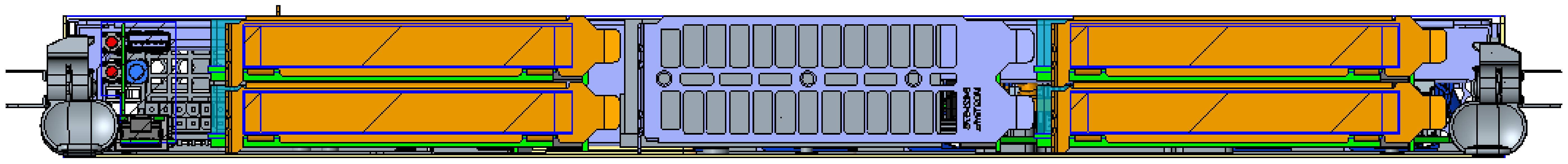
Eight AICs with Cables (or DW AICs)



Two IO Slots in 1U Height



Front View:



The Datacenter-Ready Secure Control Module ***(DC-SCM)***

&

The Datacenter-Ready Secure Control Interface ***(DC-SCI)***

DC-SCM displaces the Motherboard

Do it once, step, and repeat

Intercept Innovative Work

DC-SCM (in a nutshell)

- DC-SCM is “the heart of the motherboard” when we extract CPU(PCH)/Memory and IO Slots

- Given a traditional 1S, 2S, 4S, … Motherboard, extract CPU/PCH, DIMM Slots, IO Slots, and the associated VRs, Clock Drivers, and Reset Circuitry, and move them to a new Module

- The residual is the DC-SCM which will include everything else such as BMC, RoT, Flash, Boot SSD, connectors for Fan control, and PSU control

DC-SCM

- Reduce the problem to that which has been solved before
- ***“Same as before”*** with F/W, S/W, and Services:

  maintaining the established tools and solutions experience

  Same management, power sequencing, reset, FRU ID, VPD, …
-

DC-SCM

- Reduce the problem to that which has been solved before
- *"Same as before"* with F/W, S/W, and Services:

  maintaining the established tools and solutions experience

  Same management, power sequencing, reset, FRU ID, VPD, …

- A vehicle to drive a common Boot, Monitoring, Control, and Remote Debug procedures for

  Xeon, EPYC, ARM64, Power Servers

  firmware, diagnostic tools, manufacturing tools

Software Standardization

- Collaborating with CPU suppliers, Open Computing Project community (**OCP**), and Open Software Foundation (**OSF**) to standardize the hardware and software interfaces for the industry **OpenBMC** with **RedFish** transport and for the system BIOS/UEFI based on **EDK-II**

Collaboration and Feedback

- To ensure we have all aspects covered for upcoming servers

- We sought feedback from our industry colleagues encouraging them to provide part- and pin-list for DC-SCM

- We have compiled the feedback and proposals into a useful specification for various server types and expansion chassis so that they may be datacenter-ready

An example of DC-SCM

DC-SCM

- Receives Power
- Remote Control at Cloud Scale

CPU/Memory/IO Module (Xeon, EPYC, ARM64)

Expansion Chassis (JBOD, JBOG)

Fans, PSUs

- Includes

BMC and Rack Management Interface

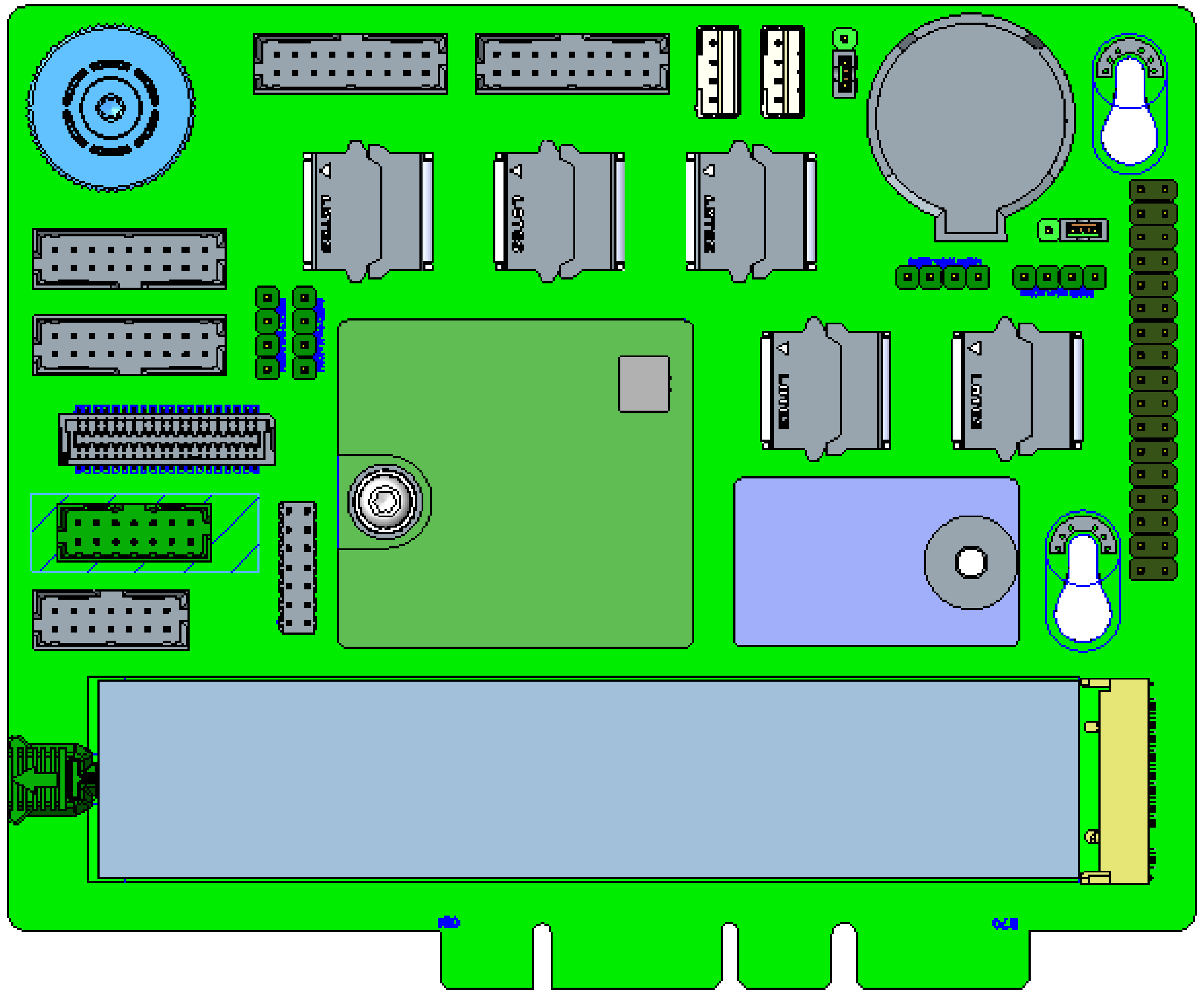
Flash Devices (all FW)

RoT and TPM for Security

Optional Boot SSD

Remote, at-scale Debug

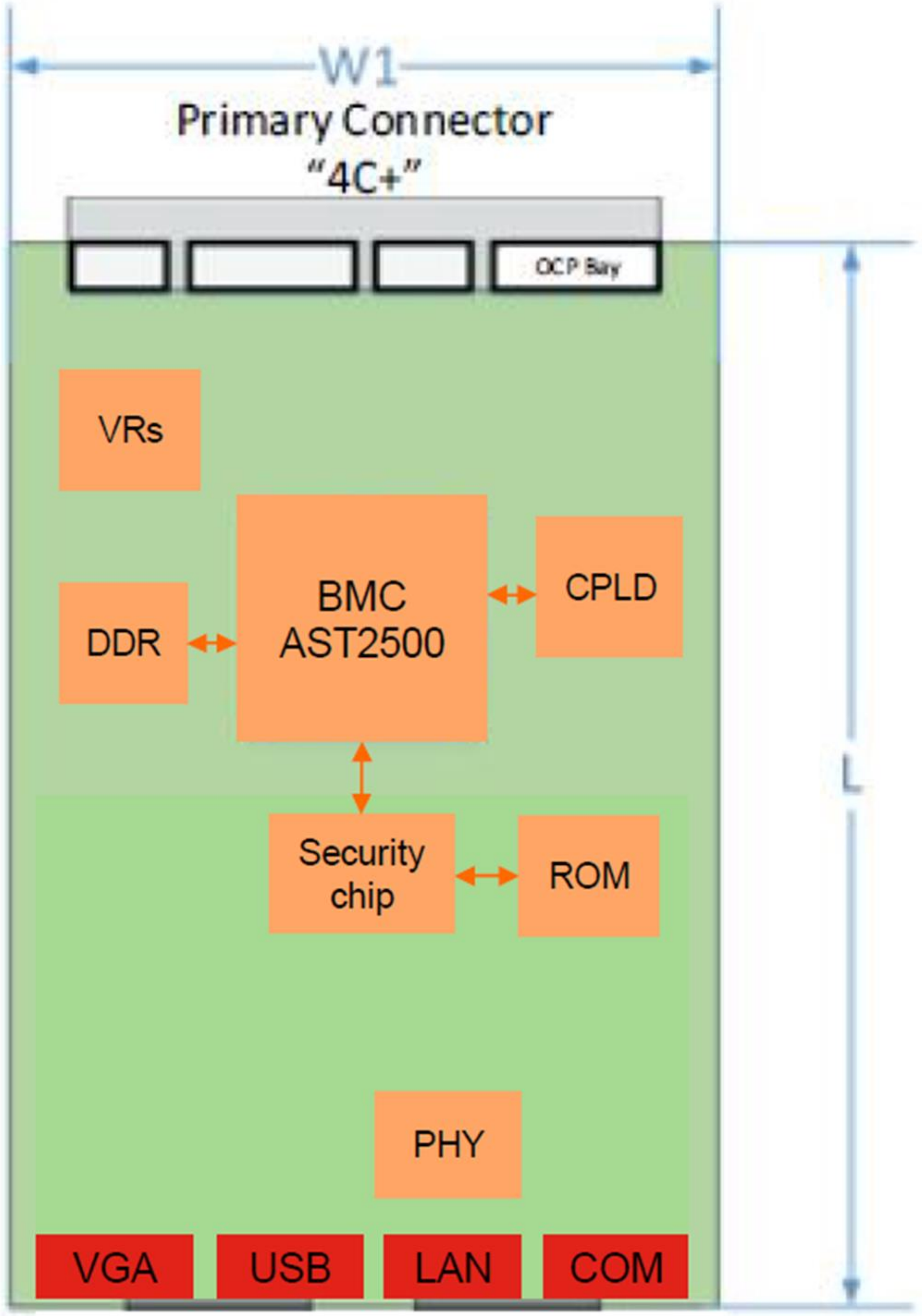
DC-SCM

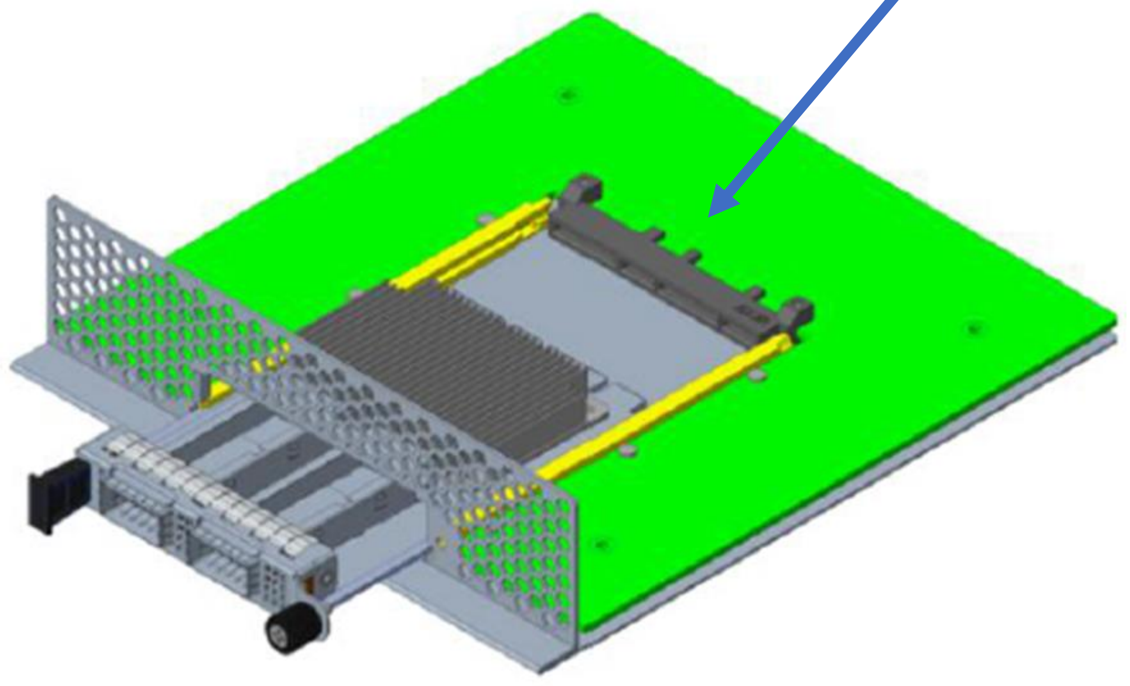


DC-SCI

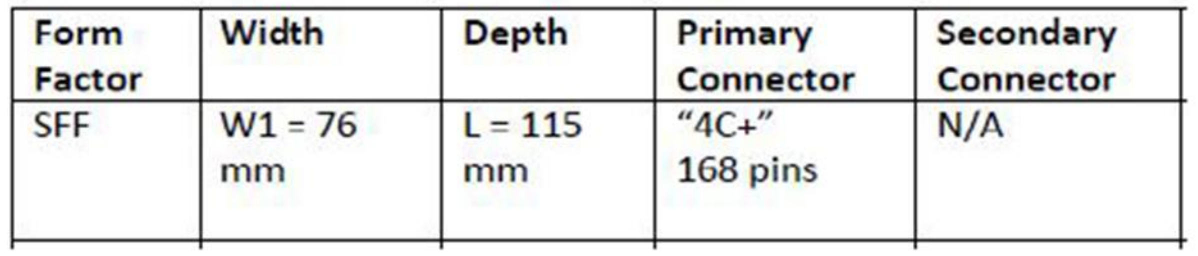
SFF-TA-1002 4C+
168-pin, Scalable Connector

Another Example of DC-SCM (OCP NIC3 Form Factor)





SFF-TA-1002 4C+
168-pin, Scalable Connector



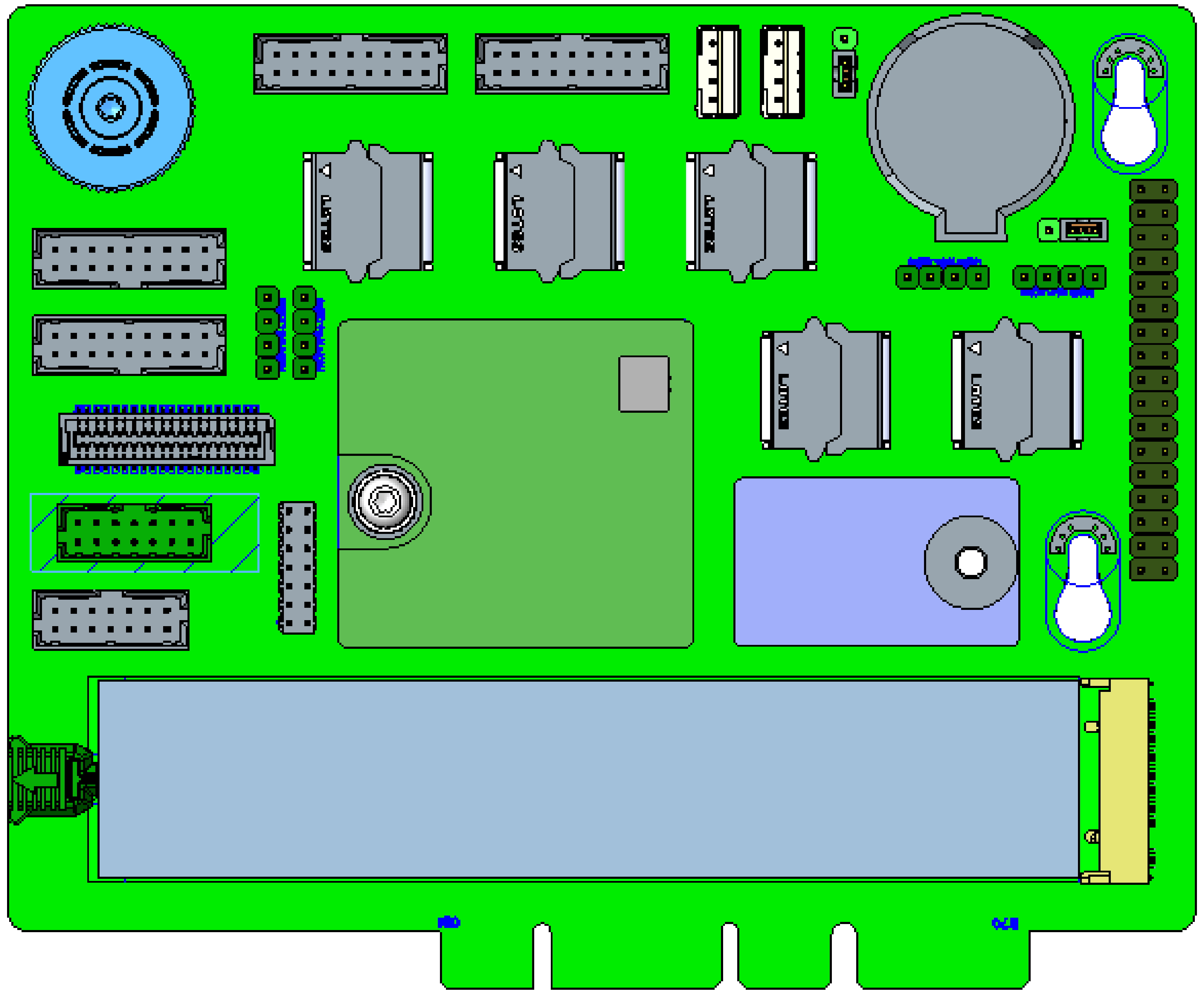
Form Factor	Width	Depth	Primary Connector	Secondary Connector
SFF	W1 = 76 mm	L = 115 mm	"4C+" 168 pins	N/A

DC-SCM

- The secure control module (SCM) includes all other system related components normally present on Motherboards
- Baseboard Management Controller (BMC), Realtime Clock (RTC), FAN/PSU Control, Root of Trust Chip (RoT: Cerberus and the associated circuitry), BIOS & BMC Flash, and the Boot Device
- The SCM is small enough to fit anywhere in the Chassis (not necessarily at the front, coplanar, vertical attached, cabled)
- Most building blocks are stateless
- SCM holds control bits secured

An Example of SCM Expander Connectors

Item	Function	Qty
1	M2 socket	1
2	TPM connector	1
3	SPI socket	5
4	RoT connector	1
5	VGA cable connector	1
6	NCSI cable connector	1
7	Front panel cable connector	1
8	FAN cable connector	2
9	BMC debug UART Pin header	1
10	HOST UART Pin header	1
11	Auxiliary UART Pin header	1
12	Reserved UART Pin header	1
13	Pin header	2
14	PSU cable connector	2
15	JATG cable connector	1
16	Reserved USB cable header	1
17	ID LED header	1
18	Battery	1
19	I2C Header	1
23	Golden Finger	1

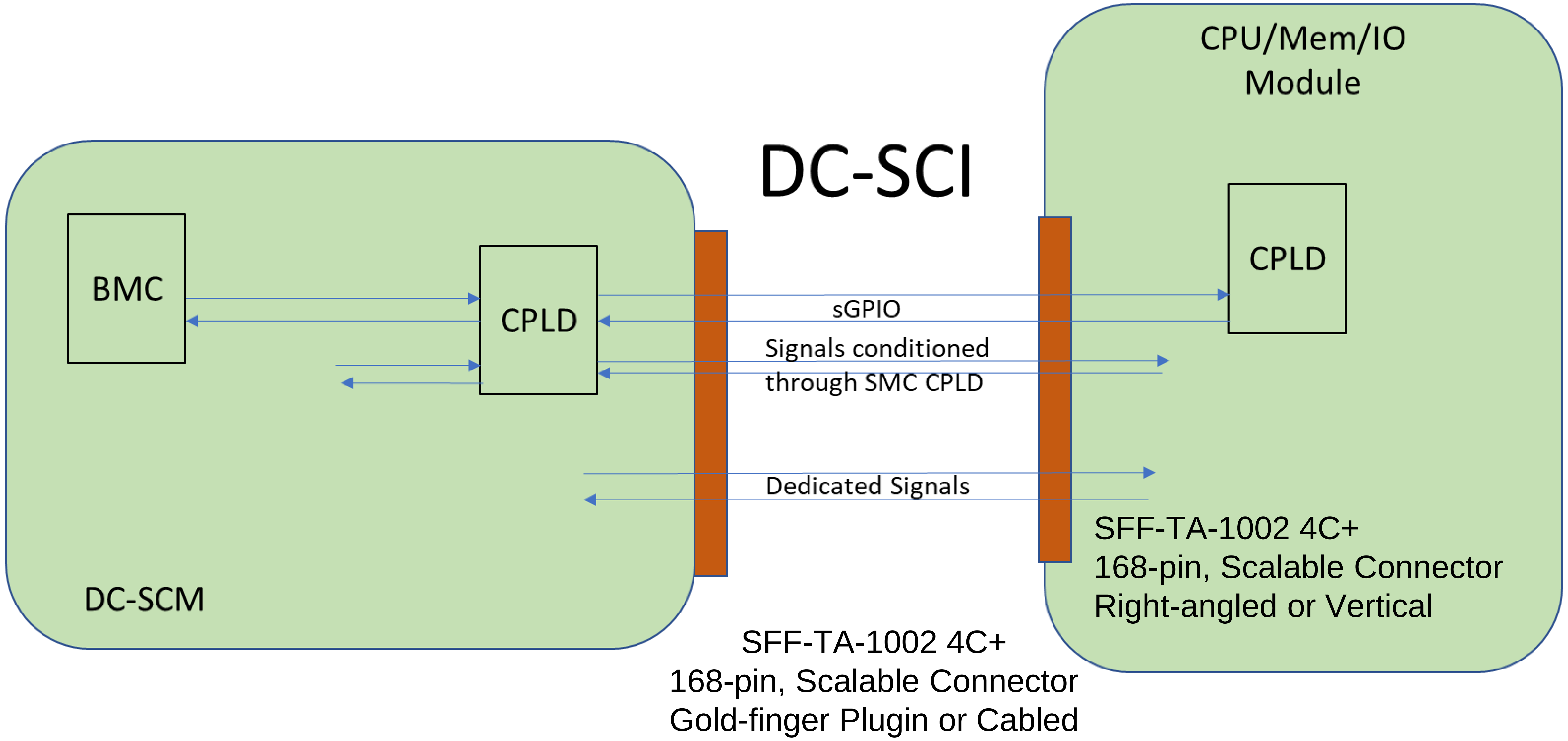


SFF-TA-1002 4C+
168-pin, Scalable Connector

DC-SCM to CPU/Mem Module Interface (DC-SCI)

Pinout and definition

Pin Reduction via SGPIO



I2C with Alert#	36	ADR_Complete	1
PECI	2	ModulePowerEn	1
JTAG	5	ModulePowerOK	1
Remote Debug	6	SysHardReset_N	1
LPC/eSPI	12	WarmReset_N	1
SPI	6	IO_Reset_N	1
SPI (Second Port)	4	Expander Reset_L	1
TPM	2	SYS_PWRBTN_N	1
USB	2	RSTBTN_OUT_N	1
UART	2	ME Recovery	1
PCIe Clock	2	SLP_A	1
BMC PCIe	5	GPIO Reset_L	1
M.2 PCIe	18	SGPIO 1	4
USB 2.0	2	SGPIO 2	4
BMC_NMI_N	1	Power	5
SMI_Active_N	1	GND/Return	26
CPU CatErr_N	1	Reserved	8
Inter-BMC_CPU	2	Total	168

DC-SCI (pinout proposal)

Connector Type:
- SFF-TA-1002 4C+
- 168-pin, Scalable Connector

DC-SCM Connector:
- Gold-finger

CPU/Memory/IO Module Connector:
- Right-angled or Vertical Recepticle

I2C with Alert#	36	ADR_Complete	1
PECI	2	ModulePowerEn	1
JTAG	5	ModulePowerOK	1
Remote Debug	6	SysHardReset_N	1
LPC/eSPI	12	WarmReset_N	1
SPI	6	IO_Reset_N	1
SPI (Second Port)	4	Expander Reset_L	1
TPM	2	SYS_PWRBTN_N	1
USB	2	RSTBTN_OUT_N	1
UART	2	ME Recovery	1
PCIe Clock	2	SLP_A	1
BMC PCIe	5	GPIO Reset_L	1
M.2 PCIe	18	SGPIO 1	4
USB 2.0	2	SGPIO 2	4
BMC_NMI_N	1	Power	5
SMI_Active_N	1	GND/Return	26
CPU CatErr_N	1	Reserved	8
Inter-BMC_CPU	2	Total	168

Win-win

DC-SCM accelerates deploying servers from various suppliers into the datacenter

Standardizing DC-SCI for ease of integration into various datacenters

- From a Datacenter point of view:

  with one DC-SCM, a datacenter may support multiple variants of servers (AMD-, Xeon-, ARM64-based 1S, 2S, 4S, …) and expansion chassis, JBODs, JBOG, JBOFs, …

- From OEM/ODMs' point of view:

  A product will fit datacenters of various CSPs or Hyperscalers

If we are smart, one DC-SCM may enable supplier products into different DC types; otherwise, each Datacenter Provider may have its own version of DC-SCM

Call to Action

Design your Servers, Expansion Chassis, JBODs, JBOGs, JBOFs, etc. with DC-SCI connector in mind.

Make your solution Datacenter-Ready!

Visit Microsoft Booth for an example of DC-SCM and MBA

Join the effort to enhance DC-SCM and DC-SCI

https://www.opencompute.org/projects/server

Q&A

Presenter

Siamak Tavallaei is a Principal Architect at Microsoft Azure and co-chair of OCP Server Project. Collaborating with industry partners, he drives several initiatives in research, design, and deployment of hardware for Microsoft's cloud-scale services at Azure. He is interested in Big Compute, Big Data, and Artificial Intelligence solutions based on distributed, heterogeneous, accelerated, and energy-efficient computing. His current focus is the optimization of large-scale, mega-datacenters for general-purpose computing and accelerated, tightly-connected, problem-solving machines built on collaborative designs of hardware, software, and management.

Open. Together.

OCP Global Summit | March 14–15, 2019

OCP SUMMIT