

IPU-MACHINE: M2000

Datasheet

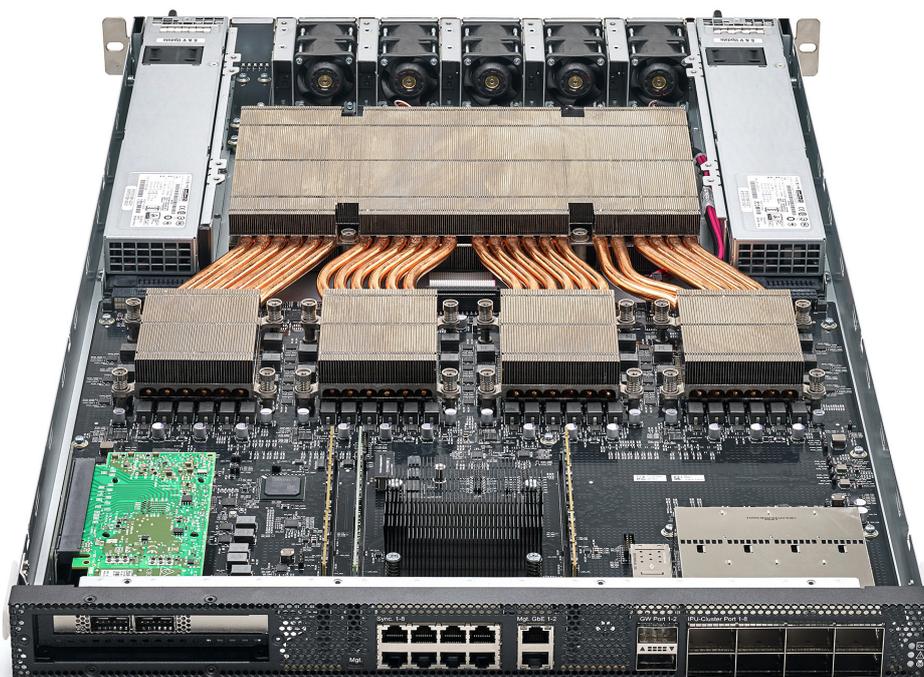




Table of contents

1	Overview	3
2	Product description	4
2.1	Product configurations	4
2.1.1	IPU-M2000 x1 direct attach	4
2.1.2	IPU-M2000 x4 direct attach	5
2.1.3	Four IPU-M2000s pre-packaged with server	6
2.1.4	IPU- POD reference design for scale-out	7
2.1.5	Software support	8
2.2	Technical specifications	10
2.3	Environmental characteristics	11
2.4	Standards compliance	11
2.5	Ordering information	11
3	Document details.....	12
3.1	Revision history	12



1 Overview

The IPU-Machine: M2000 is a 1U compute platform for AI infrastructure scalable from a direct attach development or deployment platform up to a 64K-IPU scale-out configuration. The IPU-M2000 is characterised by the following high-level features:

- 4x GC200 IPU
 - 1 petaFLOPS FP16.16 AI Compute
 - 5,888 processor cores
 - 35,000 independent parallel threads
- Up to ~450GB Exchange Memory™ comprised of:
 - Up to 448GB Streaming Memory™
 - 3.6GB In-Processor-Memory™
- IPU-Fabric™ for compiled-in networking comprised of:
 - IPU-Link™ - 512Gbps for intra IPU-POD₆₄ communication
 - GW-Link - 2x 100Gbps Gateway-Links for inter IPU-POD₆₄ communication
 - Sync-Link - dedicated hardware signalling for BSP, low jitter on IPU to IPU synchronisation
 - Host-Link - PCIe Gen4 RoCEv2 NIC/SmartNIC Interface for IPU-M2000 to server communication
- M.2 slot for SSD or media accelerator

IPU-Machine: M2000

4 x Colossus™ GC200 IPU
1 petaFLOPS AI compute
Up to 450GB Exchange Memory™
2.8Tbps IPU-Fabric™

Each Colossus™ GC200 IPU

59.4Bn transistors, TSMC 7nm @ 823mm²
250 teraFLOPS AI compute
1472 independent processor cores
8832 separate parallel threads

IPU-Gateway SoC

Arm Cortex-A quad-core SoC
Super low latency IPU-Fabric™ interconnect

Board Management Controller

RoCEv2/SmartNIC Connector

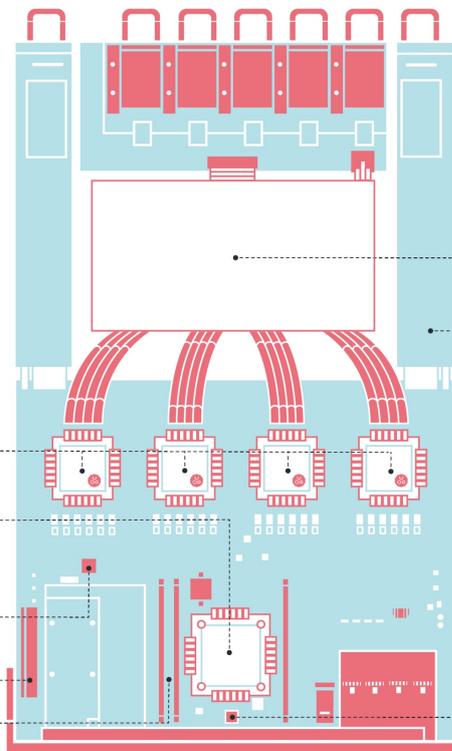
DDR4 DIMM DRAM x 2

Advanced air cooling system

Power Supply Unit (x2)

Ultra compact 1U server chassis

SSD Connector



2 Product description

2.1 Product configurations

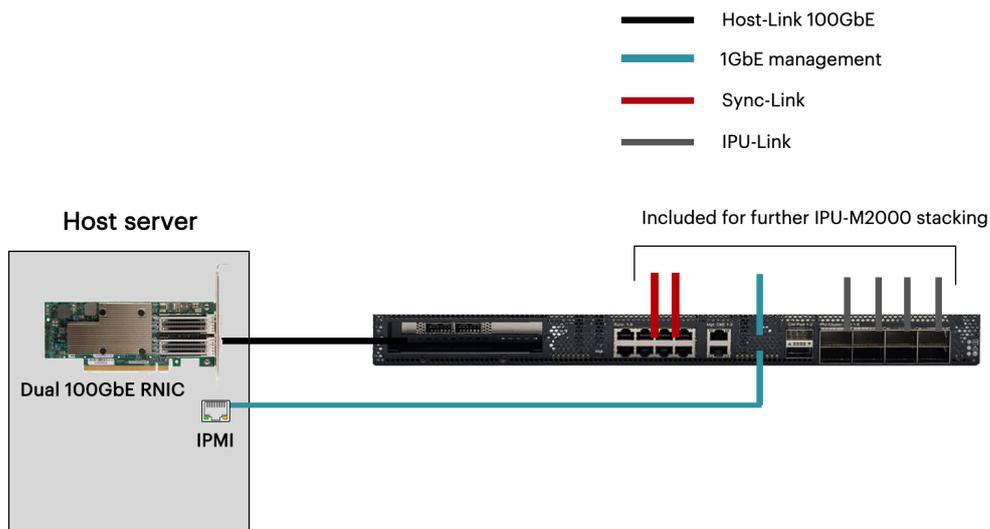
2.1.1 IPU-M2000 x1 direct attach

The IPU-M2000 blade in a direct attach system is pre-configured with Virtual-IPU (V-IPU) management software ensuring easy installation and integration with pre-existing infrastructure. The host server required to run the Poplar SDK is not included. The Dell R6525 1U server is the default server which is fully qualified by Graphcore as a host for IPU-M2000 direct attach systems. Full configuration details can be found in the [“Approved servers”](#) document on the [Graphcore documents](#) page.

The “Founder’s Edition” IPU-M2000 x1 direct attach system comes complete with all the cables required.

IPU-M2000 DIRECT ATTACH

“Founders-Edition” IPU-M2000 with included connecting cables



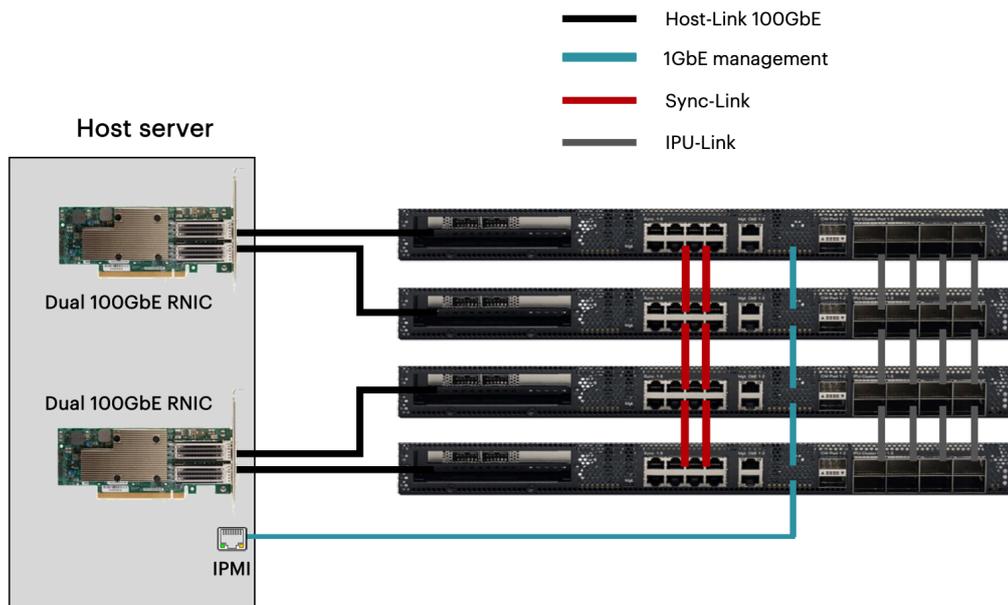
2.1.2 IPU-M2000 x4 direct attach

The IPU-M2000 is also stackable in a x4 direct attach configuration - the four IPU-M2000s are connected together as a direct attach system, then connected to a common host server (the host server is not provided by Graphcore for direct attach systems). It is easy to assemble and test the IPU-M2000 x4 direct attach configuration using Graphcore's "IPU-M2000 direct attach installation and integration guide" (which will be available on the developer page (<https://www.graphcore.ai/developer>) in early 2021) or work with Graphcore's channel partners for integration support.

A high-level cabling diagram for the IPU-M2000 x4 direct attach configuration is shown in the figure below. Note that this figure assumes a default configuration where the server is provided with 2x dual port 100Gbps Ethernet RoCEv2 NICs to support the Host-Link connectivity to the IPU-M2000s using the IPU over Fabric (IPUoF) protocol. Several RoCEv2 NIC configurations are possible, please contact either Graphcore or its channel or OEM partners for other 100Gbps RoCEv2 NIC options for the host server.

IPU-M2000 DIRECT ATTACH STACK (X4)

Two PCIe slots in the host server - 4x100GbE : 16 IPU



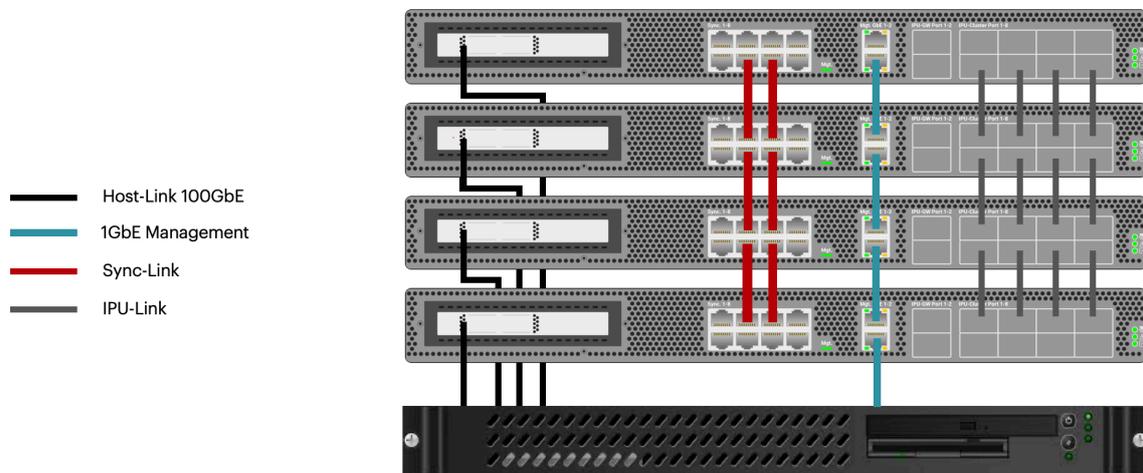


2.1.3 Four IPU-M2000s pre-packaged with server

Graphcore's channel and OEM partners also offer a pre-packaged four IPU-M2000 configuration as a 5U server offering. In this configuration, the four IPU-M2000s are connected to the host 1U server as part of an integrated 5U server product. The 1U server and the detailed specifications for this system will vary with reseller or OEM options. Please consult Graphcore sales for more information. A generic view of the four IPU-M2000 5U server configuration is shown in the figure below.

FOUR IPU-M2000 5U SERVER SYSTEM

Pre-packaged and qualified 5U server chassis with four IPU-M2000s - 16 IPU





2.1.4 IPU- POD reference design for scale-out

The IPU-POD reference design is currently available in an IPU-POD₁₆ configuration with 4 x IPU-M2000s, as well as an IPU-POD₆₄ configuration with 16 x IPU-M2000s. IPU-POD₆₄ racks can be scaled for systems ranging from 64 to 64K IPU processors in switched or direct 3D torus IPU-Fabric™ configurations.

Other configurations such as IPU-POD₃₂ and larger scale-out systems (IPU-POD₁₂₈ and IPU-POD₂₅₆) will be available in 2021 – please contact Graphcore sales for more information.

The IPU-POD₆₄ reference design combines the sixteen IPU-M2000s with switches and a host server in a pre-qualified scale-out system. Note that Graphcore has pre-qualified the IPU-POD₆₄ reference design with:

- One Dell R6525 server
- One Arista 7060X switch (32x100G + 2 10G) for the TOR switch
- One Arista 7010T switch (48p 1G+ 4x1/10G) as the management switch

Please contact Graphcore sales for alternative offerings from our reseller and OEM partners.

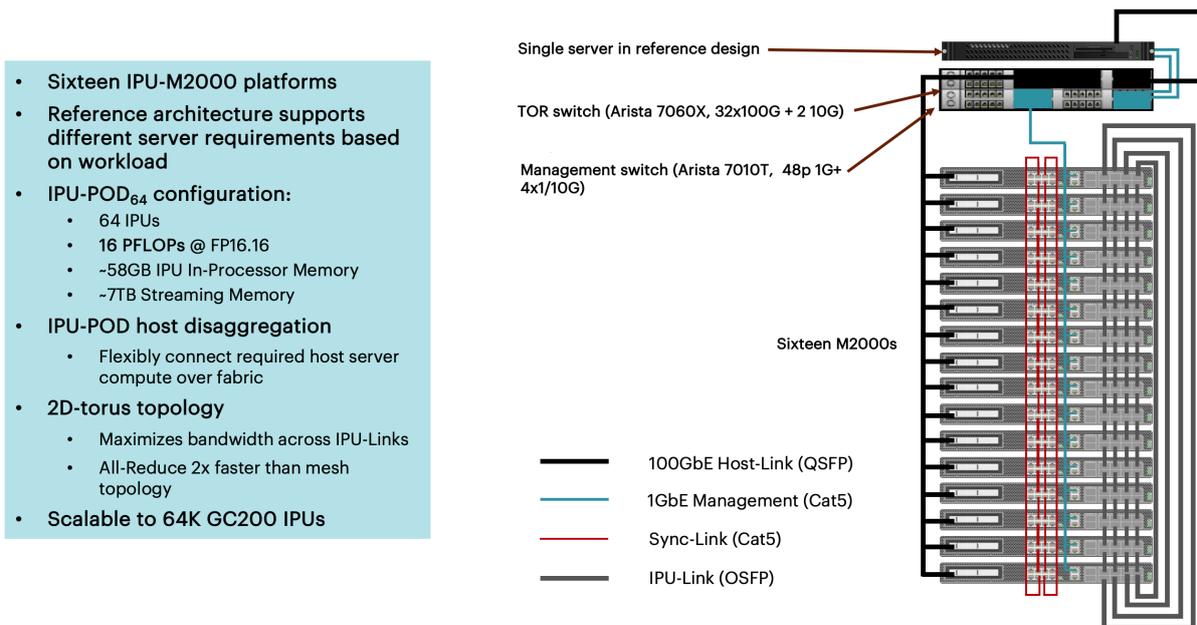
The IPU-POD₆₄ based on the IPU-M2000 is characterised by the following features:

- Disaggregated host architecture allows for different server requirements based on workload
- IPU-POD₆₄ configuration with sixteen IPU-M2000s supporting:
 - 16 petaFLOPS @ FP16.16
 - ~7.2 TBytes of Exchange Memory
- 2D-torus topology
- Scalable to 64,000 GC200 IPU

A high-level view of the IPU-POD₆₄ cabling diagram is shown in the figure below.

IPU-POD₆₄ REFERENCE DESIGN

IPU-POD₆₄ with default options for host server and switches



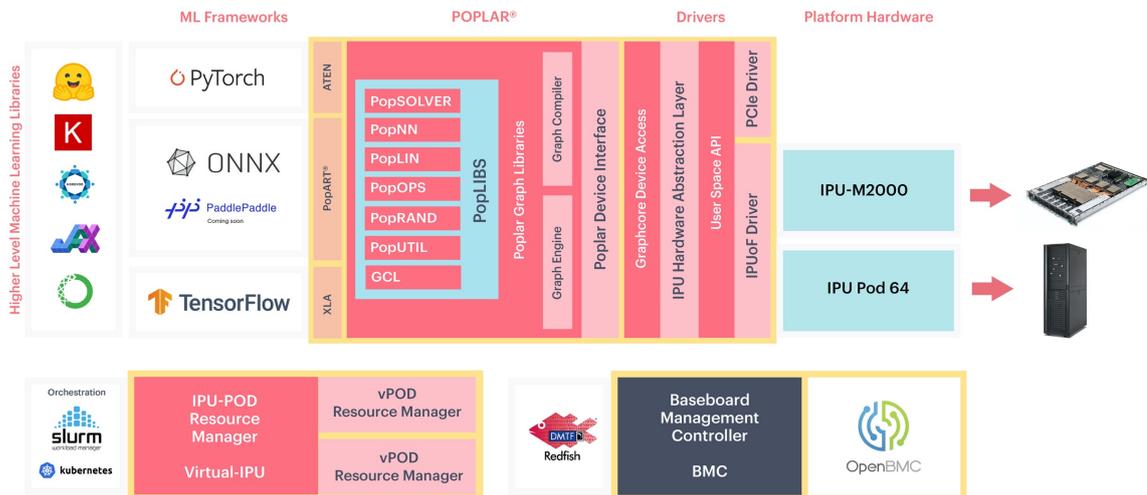
- Sixteen IPU-M2000 platforms
- Reference architecture supports different server requirements based on workload
- IPU-POD₆₄ configuration:
 - 64 IPU
 - 16 PFLOPs @ FP16.16
 - ~58GB IPU In-Processor Memory
 - ~7TB Streaming Memory
- IPU-POD host disaggregation
 - Flexibly connect required host server compute over fabric
- 2D-torus topology
 - Maximizes bandwidth across IPU-Links
 - All-Reduce 2x faster than mesh topology
- Scalable to 64K GC200 IPU



The IPU-POD₆₄ reference design is available through Graphcore’s network of reseller and OEM partners. Alternatively, customers may directly implement the IPU-POD₆₄ reference design with the help of the “IPU-POD₆₄ build and test guide” available from the [Graphcore documents](#) page. The associated “IPU-POD₆₄ installation and integration guide” provides more details about IPU-POD₆₄ power, thermal characteristics and data centre implementation requirements. Please contact Graphcore sales for further information.

2.1.5 Software support

IPU-M2000s are fully supported by Graphcore’s Poplar® software development environment, providing a complete scalable platform for accelerated development. Existing ML frameworks such as TensorFlow, ONNX, and PyTorch are fully supported as well as industry standard converged infrastructure management tools including Open BMC, Redfish, Docker containers, and orchestration with Slurm and Kubernetes. The PopVision™ visualisation and analysis tools provide monitoring of performance across one or more IPU’s - the graphical analysis enables detailed inspection of all processing activities.



See the “Getting Started with an IPU System” guide and the “Poplar and PopLibs User Guide” on the developer page (<https://www.graphcore.ai/developer>) for details of Poplar installation and use.



Poplar SDK	Complete end-to-end software stack for developing, deploying and monitoring AI model training jobs as well as inference applications on the Graphcore IPU
ML frameworks	ONNX, TensorFlow, and PyTorch
Deployment options	Bare metal (Linux), VM (HyperV), containers (Docker) To see a full list of supported OS, VM and container options go to the Graphcore support portal https://www.graphcore.ai/support
Host-Links	RDMA based disaggregation between a host and IPU over 100Gbps RoCEv2 NIC, using the IPU over Fabric (IPUoF) protocol. Host-to-IPU ratios supported: 1:1 up to 1:64
Graphcore Communication Library (GCL)	IPU-optimized communication and collective library integrated with the Poplar SDK stack. Support all-reduce (sum,max), all-gather, reduce, broadcast Scale at near linear performance to 64.000 IPUs
PopVision	Visualization and analysis tools
Graphcore virtual IPU SW	IPU-M2000 and IPU-POD ₆₄ resource manager IPU-Fabric topology discovery and validation
Provisioning	REST API and SSH/CLI for IPU allocation/ de-allocation into secure domains (vPODs) Plug-ins for SLURM and Kubernetes (K8)
Resource monitoring	REST API and SSH/CLI for accessing the IPU-M2000 monitoring service Prometheus node exporter and Grafana (visualization) support
Lights out management	Baseboard Management Controller (OpenBMC) Dual-image firmware with local rollback support Console support, CLI/SSH based Serial-over-Lan and Redfish REST API User management (LDAP, RADIUS)



2.2 Technical specifications

IPU processors	4 Colossus GC200 IPU processors (IPU frequency 1.325GHz) 5,888 IPU-Cores™ with independent code execution on 35,328 worker threads
AI compute	1 petaFLOPS AI compute 0.25 petaFLOPS FP32 compute
Exchange Memory	Up to 450GB Exchange Memory (3.6GB In-Processor Memory™ plus up to 448GB Streaming Memory™) 180TB/s Exchange Memory bandwidth
Streaming Memory	2 DDR4-2400 DIMM DRAM Options: 2x 64GB (default SKU in IPU-M2000 Founder's Edition) or 2x 128GB or 2x 256GB (contact sales)
IPU-Fabric	8x IPU-Links supporting 2Tbps bi-directional bandwidth 8x OSFP ports Switch-less scalability Up to 8 M2000s in directly connected stacked systems Up to 16 M2000s in IPU-POD systems 2x IPU-GW-Links (IPU-Link extension over 100GbE) 2 QSFP28 ports Switch or Switch-less scalability supporting 400Gbp bi-directional bandwidth Up to 1024 IPU-M2000s connected
IPU-Gateway	1 IPU-Gateway with integrated Arm Cortex quad-core A-series SoC
Internal SSD	32GB eMMC 1TB M.2 SSD
NIC	RoCEv2 NIC (1 PCIe G4 x16 FH ³ / ₄ L slot) Standard QSFP ports
Lights-out management	OpenBMC AST2520 2x1GbE RJ45 management ports
Thermal	Air cooled with built-in cooling system N+1 hot-plug fans
Airflow	Front-to-rear (IO to PSU) (standard) Rear-to-front (PSU to IO) (optional)
Airflow rate	60 CFM at nominal fan speed
Power	2x 1500W hot-plug PSUs (standard SSI slim type 54mm)
Input power (V _{ac})	100 - 240 V _{ac} (115 - 230 V _{ac} nominal)
Power consumption	1100W (typical)
PowerCap	User configurable system input PowerCap
Mechanical	1U 19inch chassis (Open Compute compliant) 440mm (width) x 728mm (depth) x 1U (height) Weight: 16.395kg (36.14lbs)



2.3 Environmental characteristics

Operating temperature and humidity (inlet air)	10-35°C (50 to 95°F) at 5%-85% RH (*)
--	---------------------------------------

Operating altitude	0 to 3,048m (0-10,000ft) (**)
--------------------	-------------------------------

(*) Altitude less than 900m/3000ft and non-condensing environment

(**) Max. ambient temperature is de-rated by 1°C per 300m above 900m

2.4 Standards compliance

EMC standards	Emissions: FCC CFR 47, ICES-003, EN55032, EN61000-3-2, EN61000-3-3, VCCI 32-1 Immunity: EN55024, EN61000-4-2, EN61000-4-3, EN61000-4-4, EN61000-4-5, EN61000-4-6, EN61000-4-8, EN61000-4-11
---------------	--

Safety standards	IEC62368, IEC60950
------------------	--------------------

Certifications	North America (FCC), Europe (CE), UK (UKCA), Australia (RCM), Taiwan (BSMI), Japan (VCCI), South Korea (KC), China (CQC) CB-62368, CB-60950
----------------	--

Environmental standards	EU 2011/65/EU RoHS Directive, XVII REACH 1907/2006, 2012/19/EU WEEE Directive
-------------------------	---

2.5 Ordering information

Part number	Description
GC-ADA2-00	IPU-Machine: M2000
GC-ADA2-FE	IPU-Machine: M2000 Founder's Edition



3 Document details

3.1 Revision history

This document's revision history is as follows:

Version	Date	Notes
1.0	30 th September 2020	First release
1.1	22 nd October 2020	Input power numbers and server position updated
1.2	2 nd December 2020	Updated information in software table and added IPU-POD reference design information

The European Directive 2012/19/EU on Waste Electrical and Electronic Equipment (WEEE) states that these appliances should not be disposed of as part of the routine solid urban waste cycle, but collected separately in order to optimise the recovery and recycling flow of the materials they contain, while also preventing potential damage to human health and the environment arising from the presence of potentially hazardous substances.



The crossed-out bin symbol is printed on all products as a reminder.

Waste may be taken to special collection site or can be delivered free of charge to the dealer when purchasing a new equivalent or without obligation to make a new purchase for equipment smaller than 25cm.

For more information on proper disposal of these devices, kindly refer to the public utility service.

Trademarks & copyright

Graphcore® and Poplar® are Registered Trademarks of Graphcore Ltd.

Colossus™, IPU-Core™, In-Processor-Memory™, Exchange Memory™, Streaming Memory™, IPU-Tile™, IPU-Exchange™, IPU-Machine™, IPU-M2000™, IPU-POD™, IPU-Link™, Virtual-IPU™, AI-Float™, IPU-Fabric™, PopART™, PopLibs™, PopTorch™ and PopVision™ are Trademarks of Graphcore Ltd.

All other trademarks are the property of their respective owners.

Design and specifications subject to change without prior notice.

© Copyright 2020, Graphcore Ltd.