BIOS IT

M2000 IPU SERVER
Second generation IPU systems for
AI infrastructure at scale

A core, new building block for AI infrastructure, the IPU-M2000 is powered by 4 x Colossus Mk2 GC200, Graphcore's second generation 7nm IPU. It packs 1 PetaFlop of AI compute, up to 450GB Exchange Memory and 2.8Tbps IPU-Fabric for super low latency communication, in a slim 1U blade to handle the most demanding of machine intelligence workloads.

FLEXIBLE, MODULAR DESIGN
The IPU-M2000 has a flexible, modular design, so you can start with one and scale to thousands. It works as a standalone system, eight can be stacked together or racks of 16 tightly interconnected IPU-M2000's in IPU-POD64 systems can grow to supercomputing scale thanks to 2.8Tbps high-bandwidth, near-zero latency IPU-Fabric™ interconnect architecture, built into the box.

Designed from the ground up for high performance training and inference workloads, the IPU-M2000 unifies your AI infrastructure for maximum datacentre utilization. Get started with development and experimentation then ramp to full scale production. Available to pre-order today.

POPLAR SOFTWARE
With Poplar, managing IPUs at scale is as simple as programming a single device, allowing you to focus on the data and the results.

A state of the art compiler simplifies IPU programming by handling all the scheduling and work partitioning of large models, including memory control, while the Graph Engine builds the runtime to execute your workload efficiently across as many IPUs, IPU-Machines or IPU-PODs as you have available.

KEY FEATURES
4 x Colossus™ Mk2 GC200 IPU

5888 independent processor cores

1 PetaFlop AI compute

180TB/s Exchange Memory™ bandwidth

2.8Tbps ultra-low latency IPU-Fabric™

GRAPHCORE ELITE PARTNER

Unparalleled performance and flexibility

Ultra-fast and jitter-free communications technology

Unprecedented 450GB Exchange-Memory™

ADVANCING FINANCIAL SERVICES
With the IPU, quantitative researchers can harness state of the art compute to explore and create new algorithms that have the potential to redefine financial services. With IPU technology, FSI institutes can fully utilise recent AI approaches to proactively track, anticipate and manage risk. For investment management firms, IPUs mean they can accelerate derivative pricing models and leverage alternative data faster to improve sentiment analysis and increase revenue.

ACCELERATING MEDICAL DISCOVERY
AI is enabling vital new directions for medical research. Intelligent algorithms are driving innovation in drug discovery, precision medicine and medical imaging. By harnessing the IPU, healthcare innovators can take full advantage of these new approaches. Next generation models can be run with unprecedented speed and accuracy, helping research scientists to deliver breakthroughs with potentially life-saving consequences.

POWERING TOMORROWS NETWORKS
As demand for digital services intensifies, providing world-class connectivity is becoming increasingly challenging. Telecoms innovators are looking to AI to drive the future of networking forward and enable cutting-edge applications. State of the art IPU technology enables unprecedented intelligence for networks, boosting efficiency, performance and revenue.

	Graphcore M2000 IPU Server
IPU	4 Colossus GC200 IPU processors (IPU frequency 1.325GHz) 5,888 IPU-Cores™
MEMORY	Up to 450GB Exchange Memory™ (3.6GB In-Processor Memory™ plus up to 448GB Streaming Memory™) 180TB/s Exchange Memory bandwidth
STORAGE	32GB eMMC 1TB M.2 SSD
FORM FACTOR	1U
NETWORK	2.8Tbps ultra-low latency IPU-Fabric™
POWER SUPPLY	2x 1500W hot-plug PSUs (standard SSI slim type 54mm)

CUSTOM CONFIGURATIONS ACROSS CORE COUNT, FORM FACTOR AND SUPPORT ARE ALSO AVAILABLE

For more information visit www.bios-it.com..