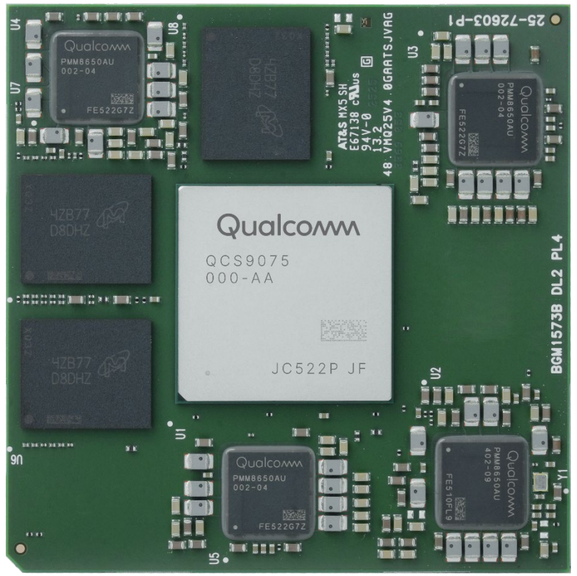
CORE-9075G

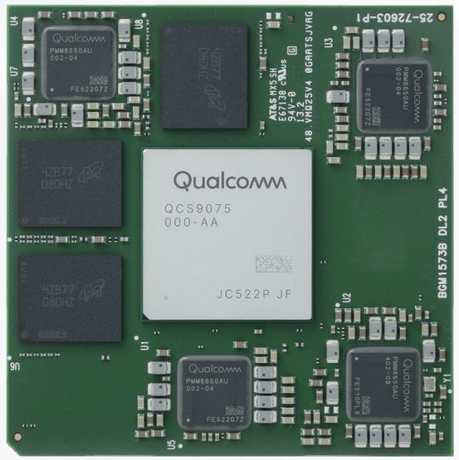
IQ-9075 SoM



2026-11-6

FIREFLY TECHNOLOGY

Product features



Qualcomm AI Processor IQ-9075

Integrated with an octa-core full-performance Kryo Gen6 CPU, with a maximum main frequency of 2.36 GHz and a general computing power of 230 KDMIPS. It has a built-in MCU-like security subsystem equipped with four real-time cores.

200 TOPS Flagship-grade NPU

Dual Hexagon NPU, delivering up to 200 TOPS (sparse INT8) / 100 TOPS (dense INT8). It supports hardware collaborative acceleration of CPU/GPU/NPU, enabling private deployment of edge-side large models and generative AI, with large model inference performance up to 12 Tokens/s@13B.

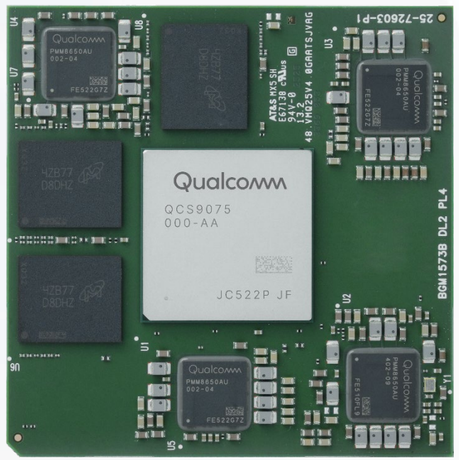
36GB LPDDR5 High-Speed Memory

On-board 36GB LPDDR5 high-speed memory, 96-bit width, 3200MHz clock frequency, with Inline ECC support. Capable of handling memory-intensive workloads including LLM weights, KV cache, and multi-channel video frame buffers.

Complete Software Ecosystem

Relying on the Qualcomm Linux software stack, it natively supports Ubuntu and Yocto systems, is compatible with deep learning frameworks such as TensorFlow, PyTorch and ONNX, and can connect to model marketplaces and massive model repositories to quickly complete model conversion, tuning and validation.

Product features



Powerful Graphics and Secure Compute

Integrated Adreno 663 GPU with secure general-purpose computing capability. Supports Vulkan 1.2, OpenGL ES 3.2 graphics APIs, as well as OpenCL 2.0 FP and Adreno NN Direct compute APIs.

8K Decoding & 4K Decoding

Supports decoding of 1×8K@60 / 4×4K@60 / 32×1080P@30 for AV1, H.265 and VP9; supports encoding of 2×4K@60 / 32×1080P@30 for H.265, HEIF/HEIC. Mixed concurrent operation of 2×4K@60 encoding plus 2×4K@60 decoding is available.

Industrial Wide-Temperature Design

Supports wide-temperature operation from -40℃ to 85℃. Equipped with USB3.1, PCIe4.0, MIPI (CSI/DSI), DP, CAN-FD, UART, I2C, SPI and other interfaces to meet expansion requirements for edge computing, industrial vision and other scenarios.

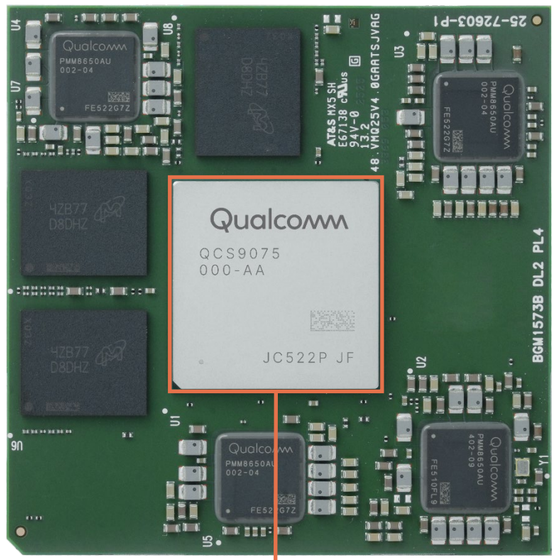
Wide Range of Application Scenarios

Widely applicable to edge computing, private LLM deployment, robotics, computer vision, intelligent security, industrial control, retail automation, medical self-service terminals and other products and fields.

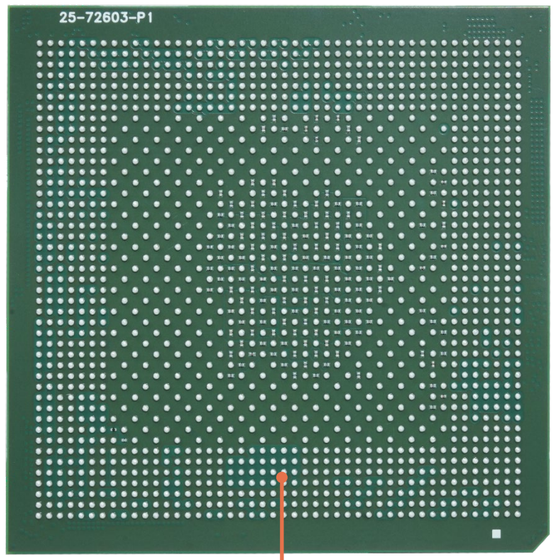
Specifications

Specifications		
Basic Specifications	SOC	Qualcomm IQ-9075
	CPU	Qualcomm Kryo® Gen6 CPU: Octa-core 64-bit (8×Kryo Gold Prime), up to 2.36 GHz
	MCU-Like subsystem	Integrated quad-core Cortex-R52 real-time subsystem, single-core frequency up to 1.85 GHz; supports independent RTOS boot via OSPI interface
	GPU	Adreno 663 GPU with secure general-purpose computing capability. Sustained maximum frequency 800 MHz under industrial temperature; non-industrial hardware peak frequency up to 840 MHz · Graphics APIs: Vulkan 1.2, OpenGL ES 3.2 · Compute APIs: Vulkan 1.2, OpenCL 2.0 FP, Adreno NN Direct
	ISP	Qualcomm Spectra 690 ISP Image Signal Processor, featuring 2× Image Front End (IFE) + 5× Lightweight Image Front End (IFE_L). Supports 24-bit HDR Bayer processing, lens distortion correction, advanced tone mapping, offset correction, lens vignetting correction, dead pixel correction, directional scaling, color lookup table, color space conversion and noise reduction.
	AI Performance	Dual Hexagon Tensor Processors (HTP) with built-in Hexagon DSP, frequency up to 1.42 GHz. Equipped with four-lane Hexagon Vector eXtensions (HVX) and dual-lane Hexagon Matrix eXtensions (HMX) coprocessors. Peak computing power: up to 100 TOPS (dense INT8) / 200 TOPS (sparse INT8). Smoothly runs large models within 13B parameters and vision algorithms; Llama2-7B achieves 22 Token/s. LPASS audio DSP delivers computing performance up to 1980 MPPS.
	DPU	Dual Adreno DPU1199 supports image processing features including target scaler, exclusion rectangle extraction, inline rotation, 17×17×17 3D LUT (ViG/DSPP), HDR10 enhancement, wide gamut WCG, corner rounding, and CCCS fixed-point conversion. It also supports UBWC 4.0 and DSC v1.2 image compression technologies.
	Video codec	Video decoding: 1×8K@60fps, 2×8K@30fps, 4×4K@60fps, 8×4K@30fps, 16×1080p@60fps, 32×1080p@30fps AV1, H.264, H.265, VP9, MPEG2 Video encoding: 2×4K@60fps, 4×4K@30fps, 8×1080p@60fps, 16×1080p@30fps H.264, H.265, HEIF/HEIC
	RAM	36GB LPDDR5 (supports Inline ECC memory error correction)
	Power supply	DC 3.3V (supports 3.1V~3.5V wide voltage input)
	Power consumption	Normal: 1.98W(3.3V/600mA), Max: 11.88W(3.3V/3600mA)
	OS/Software	Ubuntu, Yocto Linux Software Stack
	Software support	Supports private deployment of ultra-large parameter models under the Transformer architecture, including the Deepseek-R1 series, Gemma series, Llama series, Qwen series, Phi series and other large language models. Supports private deployment of image generation models such as Stable Diffusion. Supports the QNN AI inference framework, as well as multiple deep learning frameworks including TensorFlow, TensorFlow Lite, PyTorch, ONNX, Keras and Caffe.
	Package type	BGA1573 (1573-ball, ball pitch: 1.27mm)
	Size	65.0mm × 65.0mm × 3.75mm
	Weight	≈33g
	Environment	Operating Temperature: -40℃ ~ 85℃, Storage Temperature: -40℃ ~ 90℃, Storage Humidity: 10% ~ 90%RH (non-condensing)
Interface Specifications	Network	Wired: Main domain: 2 × 2.5 Gbps SGMII MAC (TSN/AVB capable, auto-negotiation 2.5G/1G/100M); RTSS real-time domain: 1 × 1 Gbps RGMII MAC (TSN/AVB capable) Wireless: No RF integrated on SoC; Wi-Fi 6 and Bluetooth 5.3 available with external companion chips
	Video input	4 × MIPI-CSI (4-Lane, C-PHY / D-PHY supported, lane splittable). Up to 16 cameras can work concurrently; single-sensor max. resolution 12 MP
	Video output	2 × MIPI-DSI (4-Lane), 4 × DP/eDP (DP1.4, MST supported); maximum 5-channel 4K output
	Audio	10 groups of audio interfaces (7 × LS-I2S/PCM/TDM + 3 × HS-I2S)
	PCIe	1 × PCIe Gen4 (4-Lane), 1 × PCIe Gen4 (2-Lane); supports both RC (Root Complex) and EP (Endpoint) modes
	USB	2 × USB3.1 Gen2, 1 × USB2.0; supports OTG and Debug mode
	Other interfaces	26 groups of QUP v3 serial engines (21 in AP main domain + 5 in RTSS real-time domain), configurable as UART / I2C / SPI; UART supports RS485 extension. 8 × CAN FD 149 GPIO in main domain + 79 GPIO in real-time domain

Interface description

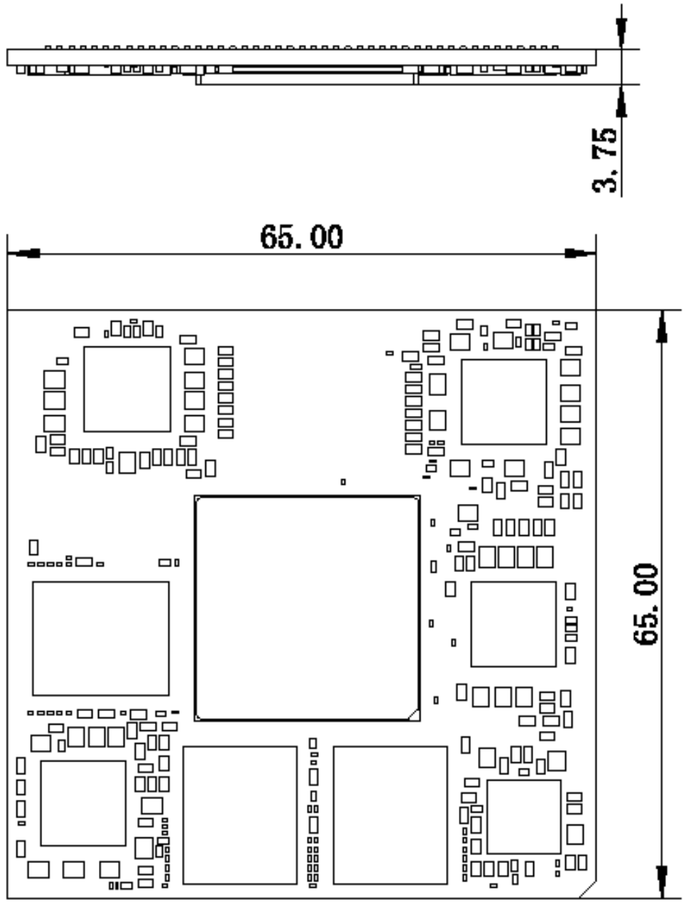


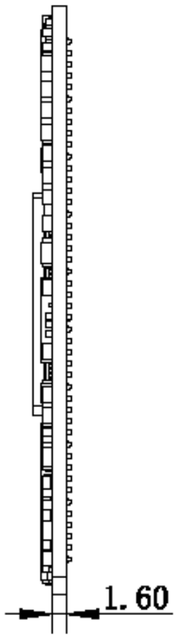
Qualcomm IQ-9075
Up to 2.36GHz

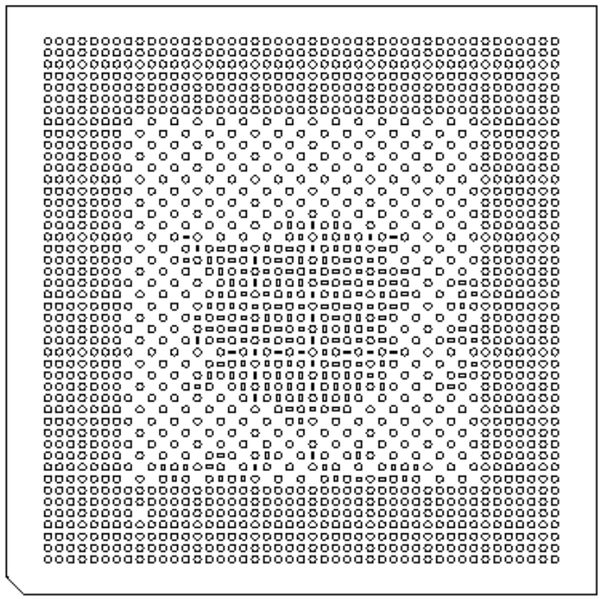


BGA1573 Package
1573-ball, ball pitch: 1.27mm

Dimension









FIREFLY TECHNOLOGY

Contact Us
(+86)18688117175

E-mail
global@t-firefly.com

Website
https://www.t-firefly.com

Address
Room 2101, Hongyu Building, #57 Zhongshan 4Rd, East District, Zhongshan, Guangdong, China.