longsys

AILPBGA™

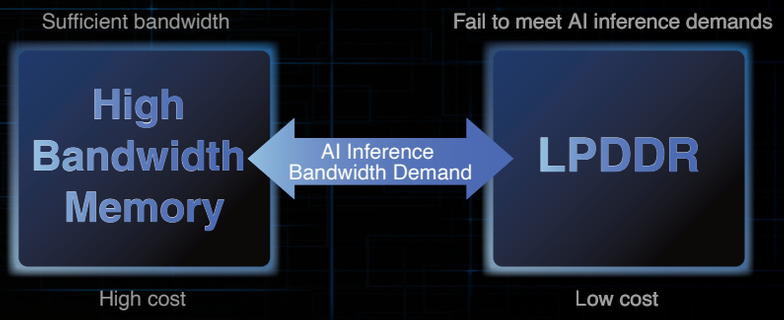
High-bandwidth memory chip compatible with LPDDR interface

256bit	307GB/s	64GB
Ultra-wide bit width	Single-chip ultra-high bandwidth	Larger capacity

Three Core Memory Pain Points In AI Inference Scenarios

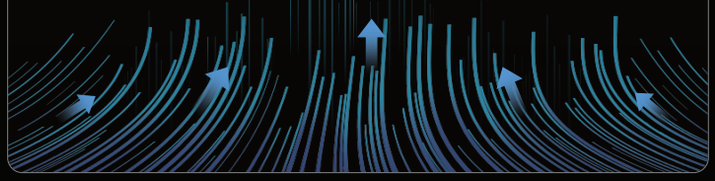
Bandwidth-cost imbalance

HBM series features sufficient bandwidth yet extremely high cost, while ordinary LPDDR is low-cost but fails to meet high bandwidth demands for AI inference.



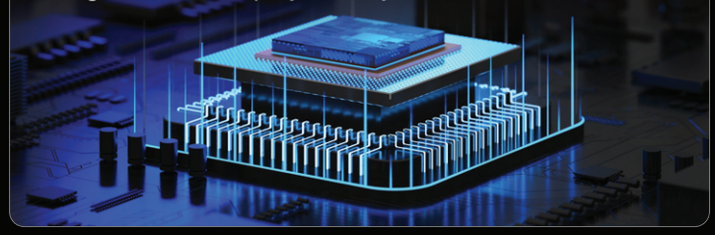
High power consumption & heavy heat pressure

Long-term operation of AI devices leads to high power consumption of traditional memory, raising operating costs and increasing heat dissipation burden on edge AI devices.



Complicated adaptation & design

Some high-end memories require dedicated controllers and advanced packaging with poor compatibility, which requiring system redesign and leads to long integration and deployment cycles.



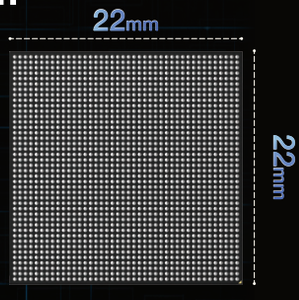
AILPBGA New Architecture & Advanced Packaging

Architecture advantages

Achieve 3D optimization of integration, bandwidth and power consumption, resolving signal & heat issues for compact AI devices.

Package specification

22×22mm, 1764 balls rationally arranged, balancing compactness and stability.



Core design

Built-in native 256-bit width,, no multi-die parallel connection needed, simplifying layout and boosting signal transmission efficiency.

Innovative Breakthroughs Enhance AI Inference Memory Performance

High cost efficiency

Deliver about 40% cloud AI memory bandwidth at merely 1/5 to 1/10 the cost of cloud high-bandwidth memory, greatly cutting AI inference deployment and operating expenses.

High Bandwidth Memory

Bandwidth ≈ 40%
Cost ≈ 20% - 10%

Ultra low power consumption

Inherit LPDDR low-power trait, operating voltage down to 0.3V, slash energy consumption and electricity cost.

LPDDR Low Power

0.3V ↓

High compatibility

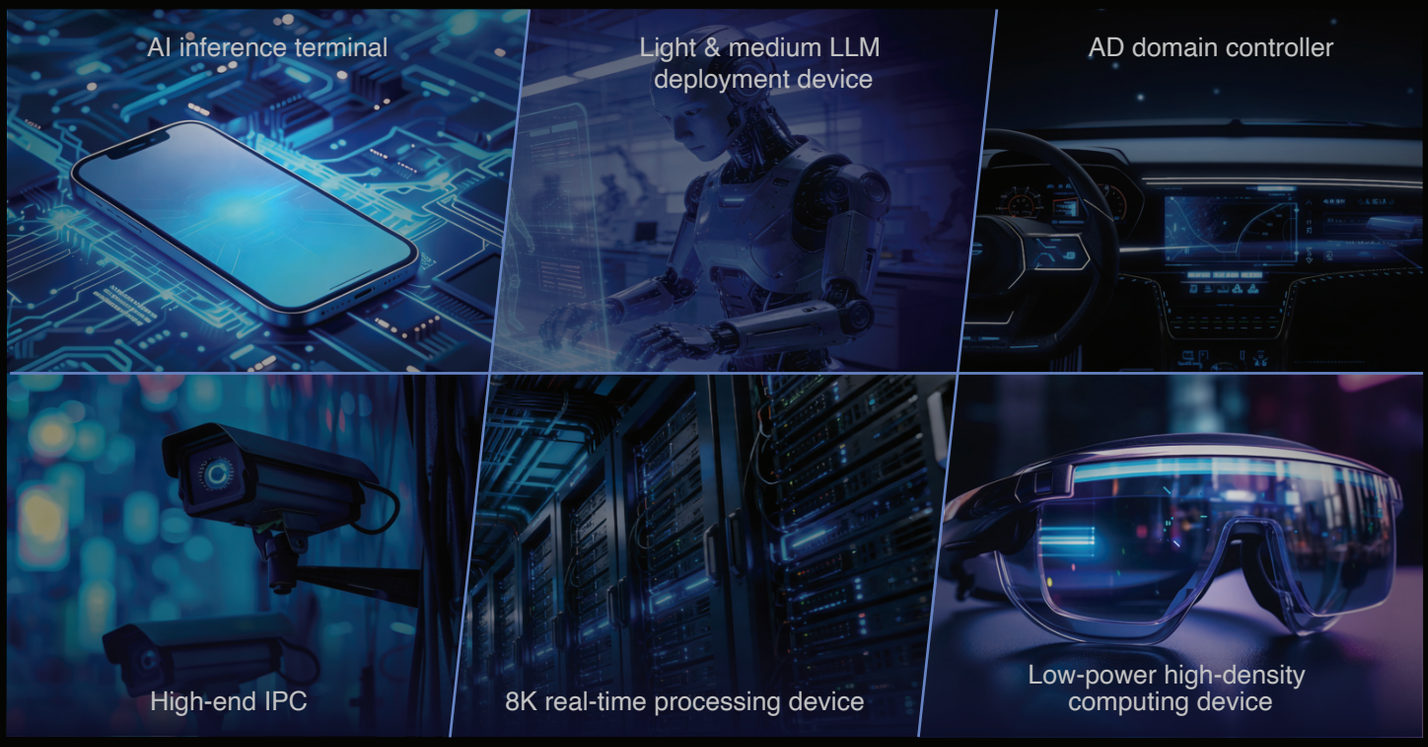
Adopt standard LPDDR5x interface, compatible with existing LPDDR and main control platforms. No system redesign required, lower adaptation threshold, accelerate customer integration verification and shorten time-to-market.

High integration

Single BGA package with innovative WBM architecture simplifies PCB layout and reduces failure points. Space-saving, strong anti-interference, ideal for mini AI terminals with more stable signal transmission.

WBM

Explore Multi-scenario Edge AI Inference Demands



*The data comes from internal tests and algorithm estimates. Performance varies by device and environment.

Cost-effective Memory Selection for AI Inference Scenarios

	Edge AI Memory		Cloud AI Memory		
Comparison dimensions	Standard LPDDR5X	**LONGSYS AILPBGA**	HBM2	HBM2E	HBM3
Interface bit width	32-bit (per standard single die)	**256-bit (native single-die)**	1024bit (8×128bit)	1024bit (8×128bit)	1024bit (16×64bit)
Single-die bandwidth	≈38.4GB/s	**≈307GB/s**	204.8~307.2 GB/s	307.2~460.8 GB/s	819GB/s
Packaging process	BGA 315	**BGA 1764**	Silicon interposer packaging (2.5D)	Silicon interposer packaging (2.5D) / Advanced 3D packaging	Silicon interposer packaging
Single-die capacity	8GB/12GB/ 16GB	**24GB/32GB/ 48GB/64GB**	4GB/8GB (Up to 16GB)	8GB/ 16GB/24GB (Up to 32GB)	36~48GB per stack
Power consumption performance	Low	**Low**	Medium-high		High
Cost	Low (Mass Prod)	**Medium-Low (Mature BGA)**	High		Ultra-high
PCB Area Routing Complexity	Multi-chips, larger PCB footprint	**Space-saving, Simplified routing, Compatiable with LPDDR platform**	Larger footprint, High complexity		Extremely large footprint, extreme complexity
Applications	Mobile phones Tablets	**AI Inference Edge LLM etc.**	AI Inference、high-end GPU、HPC、data center、professional workstation		AI training supercomputer、HPC

STORAGE FOR AI

longsys

Longsys Electronics Co., Ltd.

E-mail: FORESEE@longsys.com
Website: www.longsys.com
Facebook: Longsys
LinkedIn: Longsys