

# CXL互连技术应用探索

作者简介：

梁永贵，架构设计与技术规划专家，创新技术研发总工程师

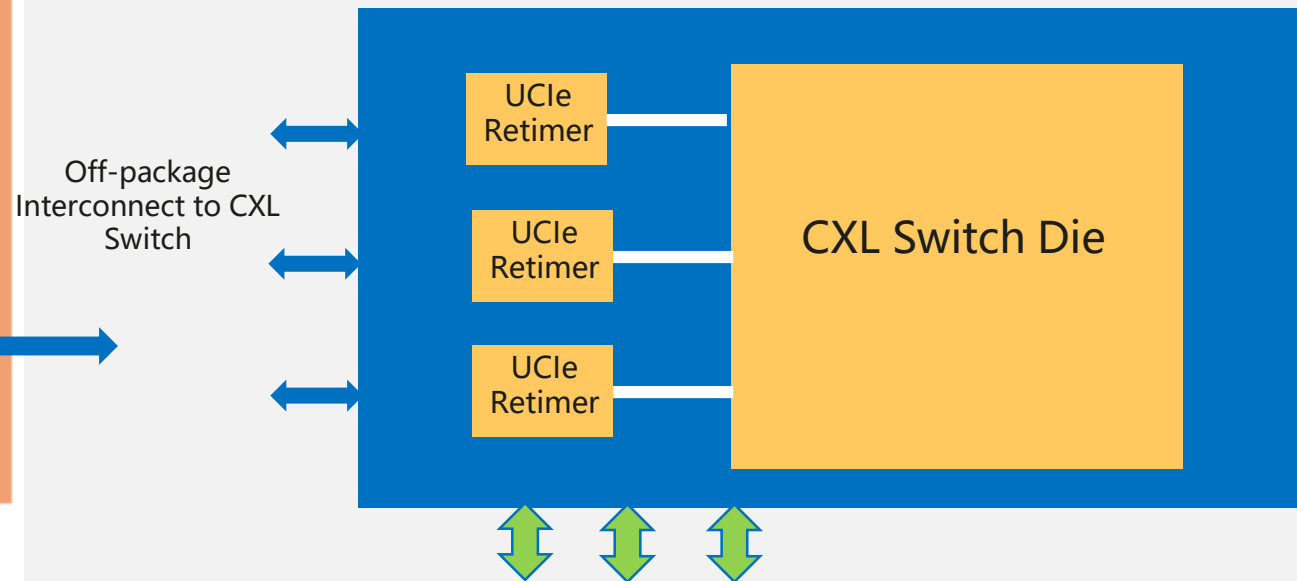
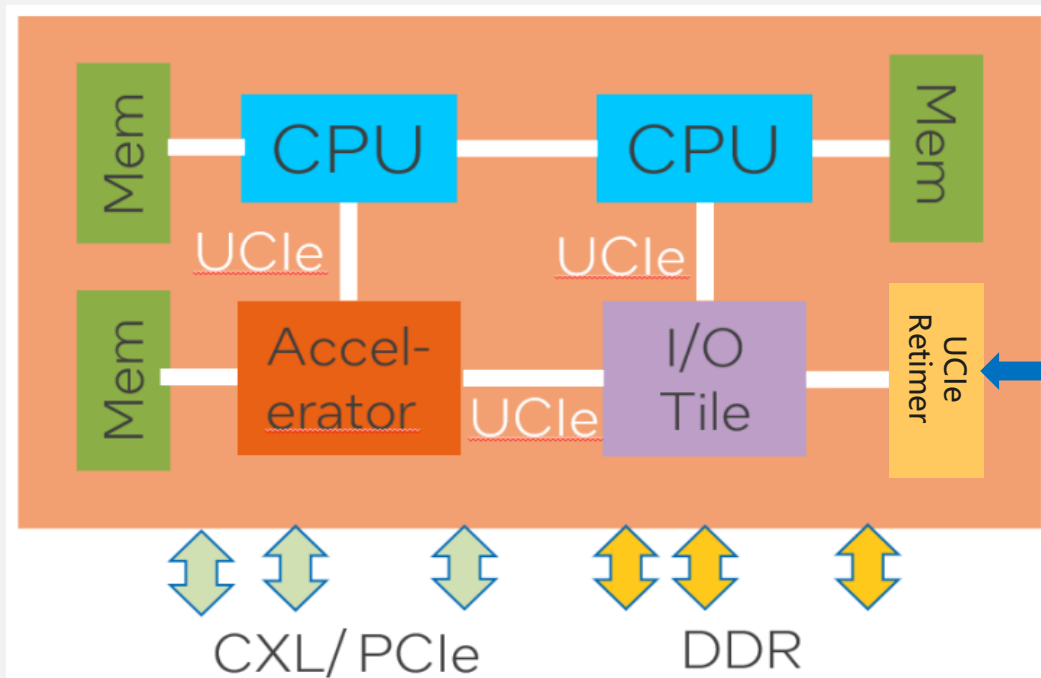


# 目录

---

- CXL互连技术介绍
- CXL互连技术应用探索
- CXL互连技术展望

# I/O近年来的两大新星：Chiplet和CXL



## Chiplet技术解决单芯片面临的挑战

- 提高芯片设计和制造的灵活性和效率
- 提高芯片系统的可扩展性和可升级性
- 降低芯片开发成本

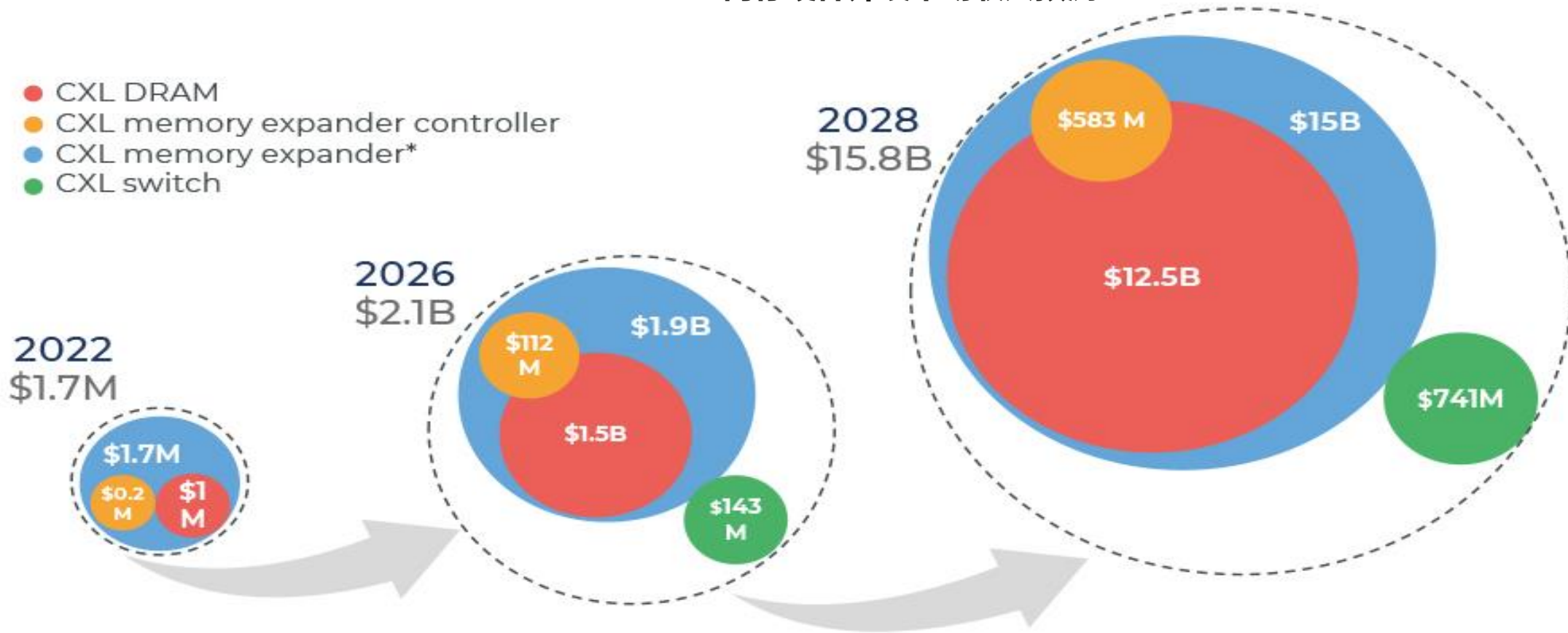
## CXL互连为不同芯片间提供通信和数据传输

- 提供更高的通信带宽和更低的延迟，提高系统性能
- 支持高速互连和共享内存，实现工作协同和数据共享
- 提供灵活的拓扑结构和可扩展性，实现资源解耦



# CXL互连技术：2028年CXL技术将开拓新兴千亿市场

## 2022-2028 CXL内存硬件外设市场收入预测



\*inclusive of DRAM and CXL memory expander controller  
CXL : Compute Express Link

# CXL互连技术：支持处理器扩展加速器和内存设备

## TYPE 1: 加速器 (带cache)



## 应用

- 高性能智能网卡
- NIC atomic

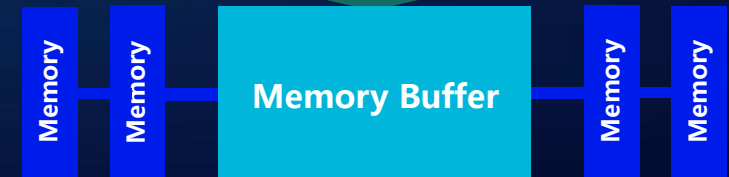
## TYPE 2: 加速器 (带内存)



## 应用

- GPU
- 高密计算

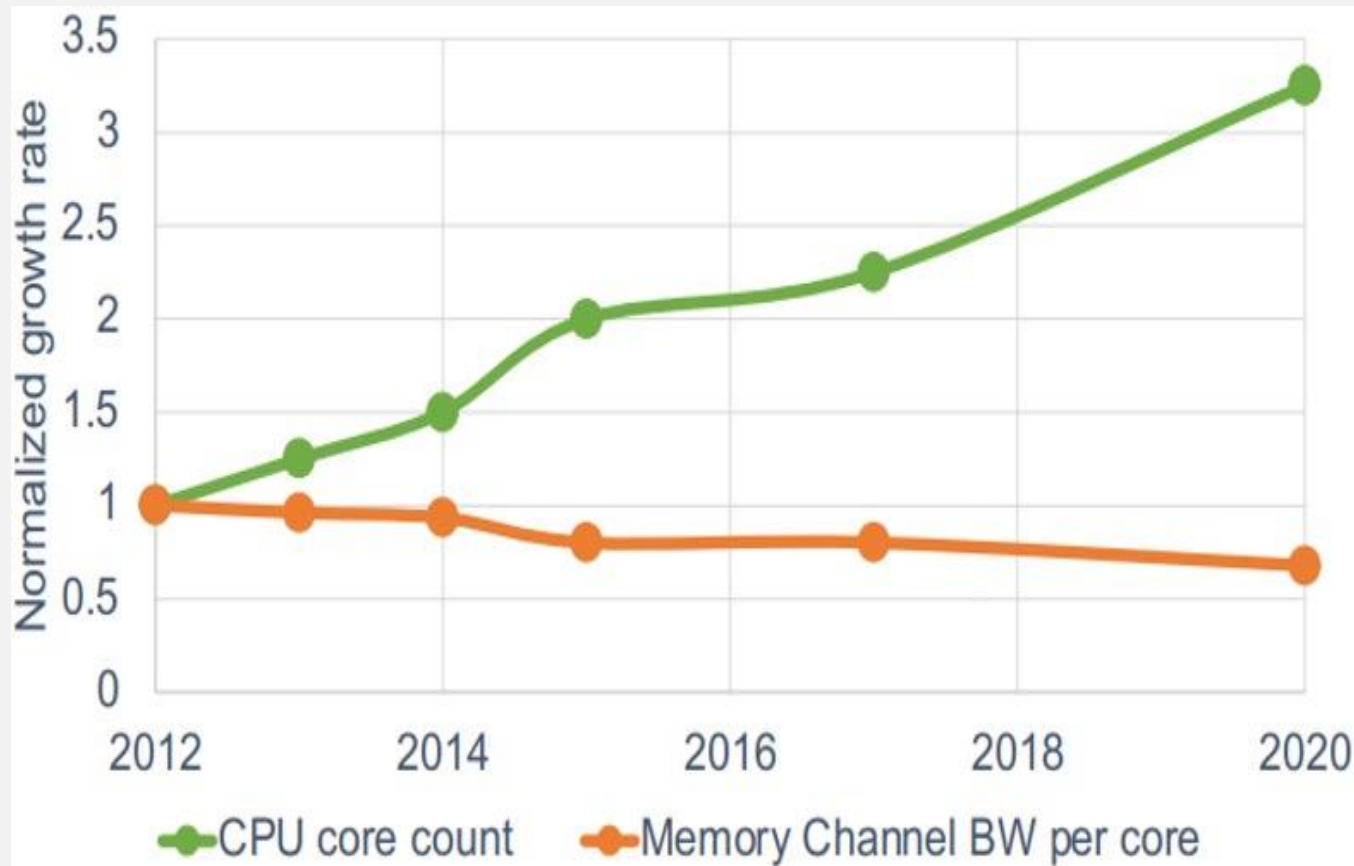
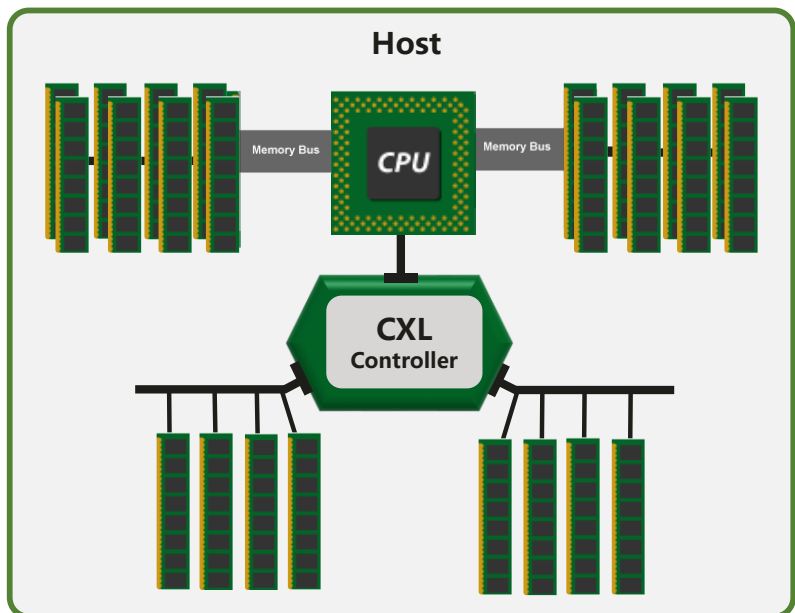
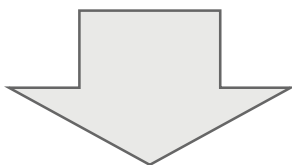
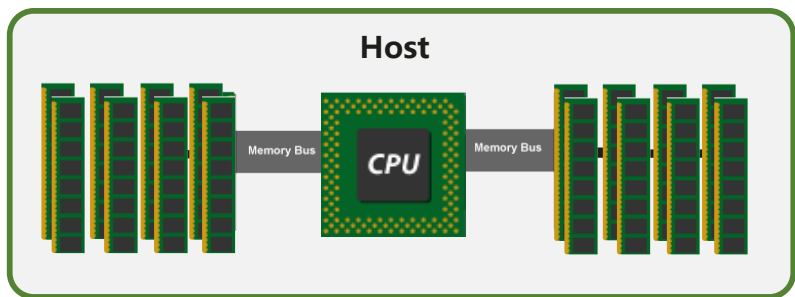
## TYPE 3: 内存扩展设备



## 应用

- 内存带宽扩展
- 内存容量扩展
- SCM扩展

# CXL互连技术：解决内存与CPU耦合、容量和带宽受CPU管脚限制的问题

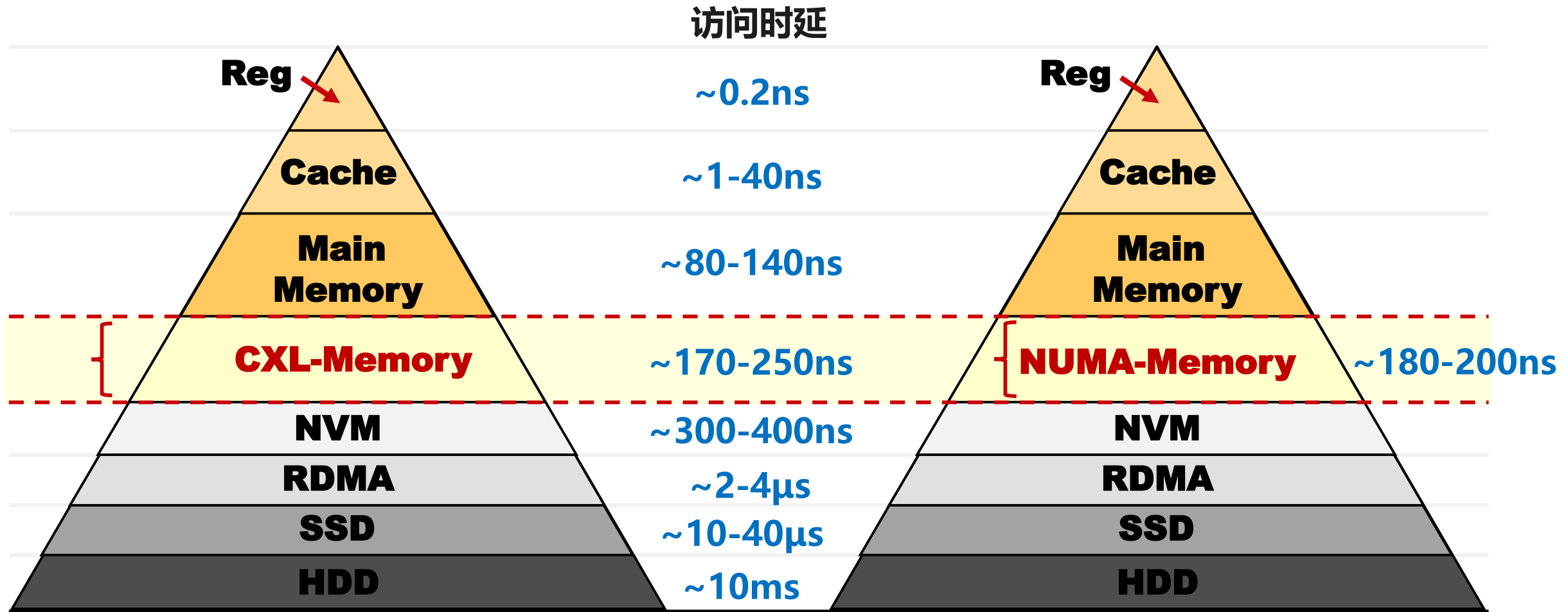


CPU 核数变化趋势 vs 单核内存变化趋势

(来源: Software Defined Memory: A Meta perspective, OCP Global Summit, 2021)

突破CPU管脚对内存限制，解决随CPU核数增长，每核可用内存下降问题

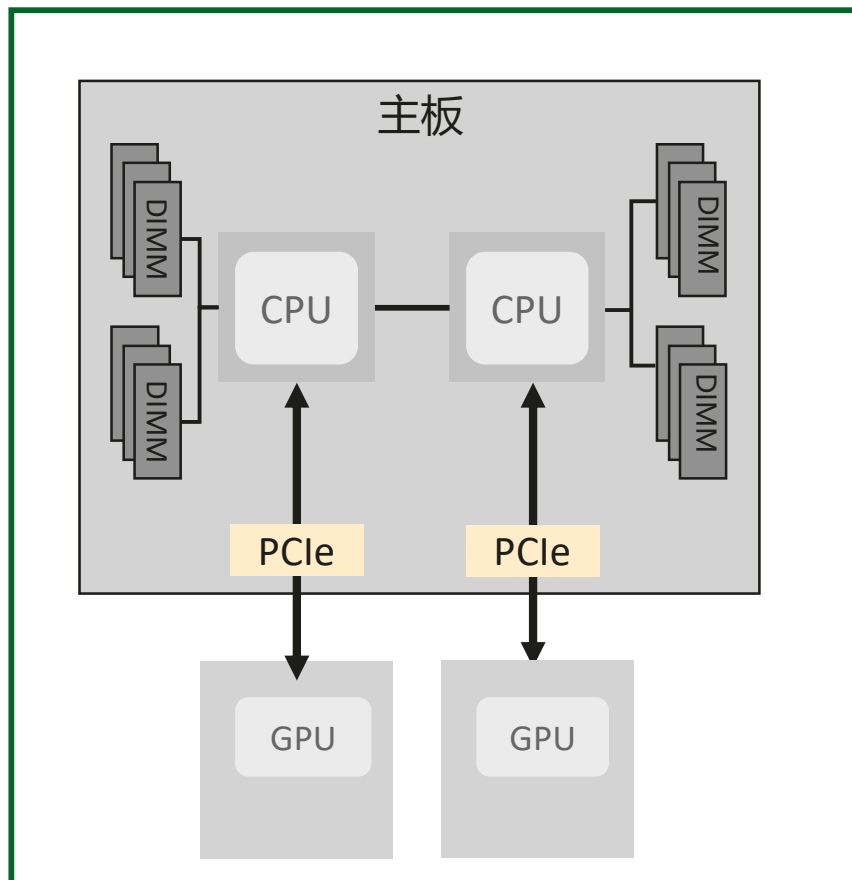
# CXL互连技术：CXL内存时延可接受，与NUMA内存时延相当



# CXL互连技术：推进全新服务器架构，突破内存墙，I/O墙

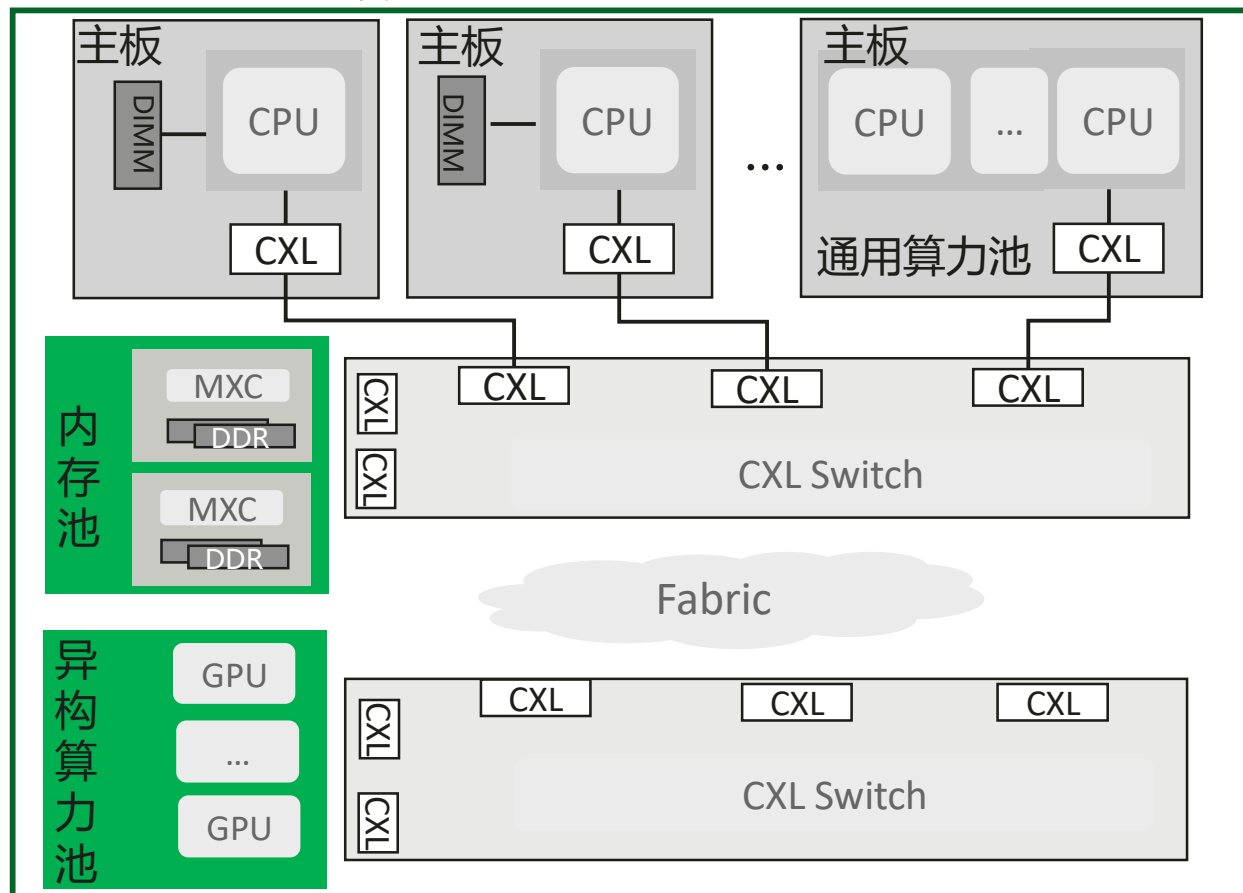
## 传统服务器架构

- 内存/存储/加速器与CPU紧耦合
- 通过网络实现资源间互通互访
- 配置固定，调整空间小



## 基于CXL的服务器架构

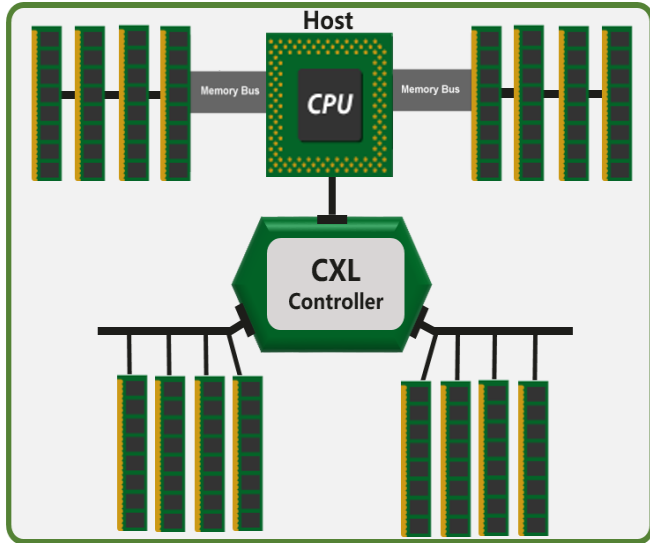
- 内存/存储/加速器与CPU解耦
- 有利于资源模块化设计
- 资源配置可共享





# CXL互连技术：内存应用技术发展趋势

## 单节点内存扩展

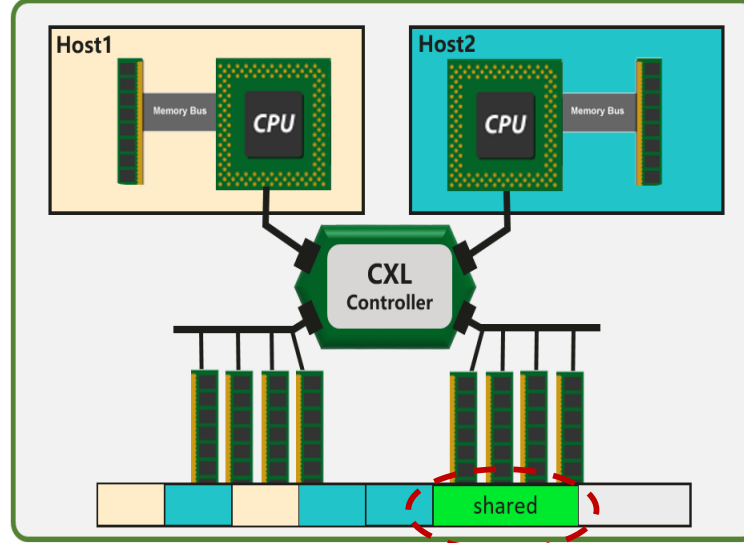


### 内存与CPU解耦

- 单节点内存容量扩展
- 单节点内存带宽扩展
- CXL 1.1

2H23 to 1H24

## 内存池化

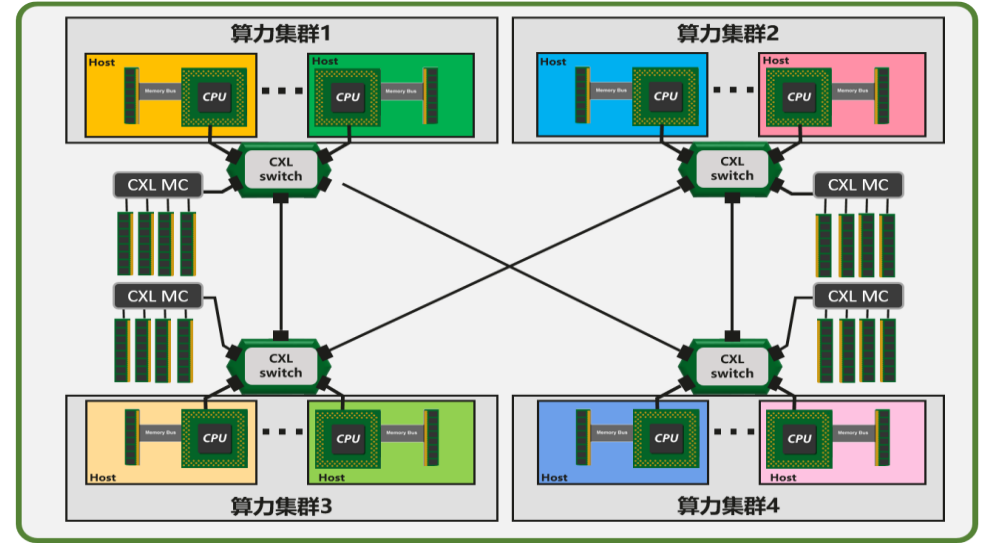


### 内存与服务器解耦

- 多节点内存池化 (机柜内, DC内)
- CXL 2.0 (标准协议仅支持到内存池化)
- **CXL 2.0+ (多节点内存共享)**

1H24 to 1H25

## 内存&数据共享



### 内存与网络解耦 (池化&共享)

- 多节点内存网络 (DC内, 跨DC)
- 灵活、极强的扩展性
- CXL 3.0

2H25 to 26+

# CXL互连技术：内存扩展解决方案

## CXL内存扩展卡原型

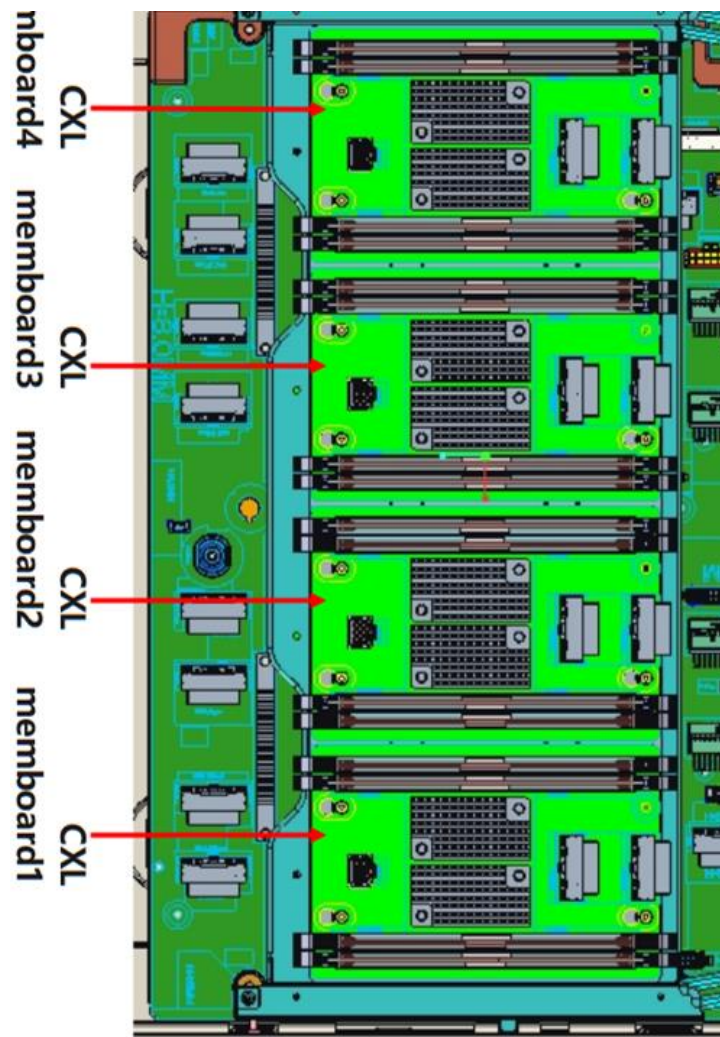


- 可支持DDR4、DDR5
- 可支持4 DIMM Slots
- 单卡最大内存容量可达1TB
- 支持NUMA Node模式
- 支持FW升级

## CXL内存扩展基础性能参考

CXL Memory Device	读时延(ns)	带宽(GB/s)	
		GCC	ICC
Local CXL DDR5	245~260	~17.5	~24.1
Remote CXL DDR5	460~475	~17.4	~16.8
Local CXL DDR4	225~236	~12.8	~13.8
Remote CXL DDR4	422~432	~12.5	~13.2

扩展卡置于CPU直连内存之上，不改造服务器、不占面板空间、内存规格提升明显





# CXL互连技术：内存扩展解决方案

<https://www.xfusion.com/cn/product/rack-server/fusionserver-2288h-v7>

CXL  
Enabled

48 x DIMM Slots

32 x DDR5 DIMMs, up to 4800 MT/s

Expandable to 48 x DIMM slots through CXL

32 X DIMM Slots to 48 x DIMM Slots

业界最早发布的内存大胖子服务器



# CXL内存模组物理形态



AIC

模组厂商主推



E3.S

内存颗粒厂商主推



异形卡

服务器厂商定制



JBOM

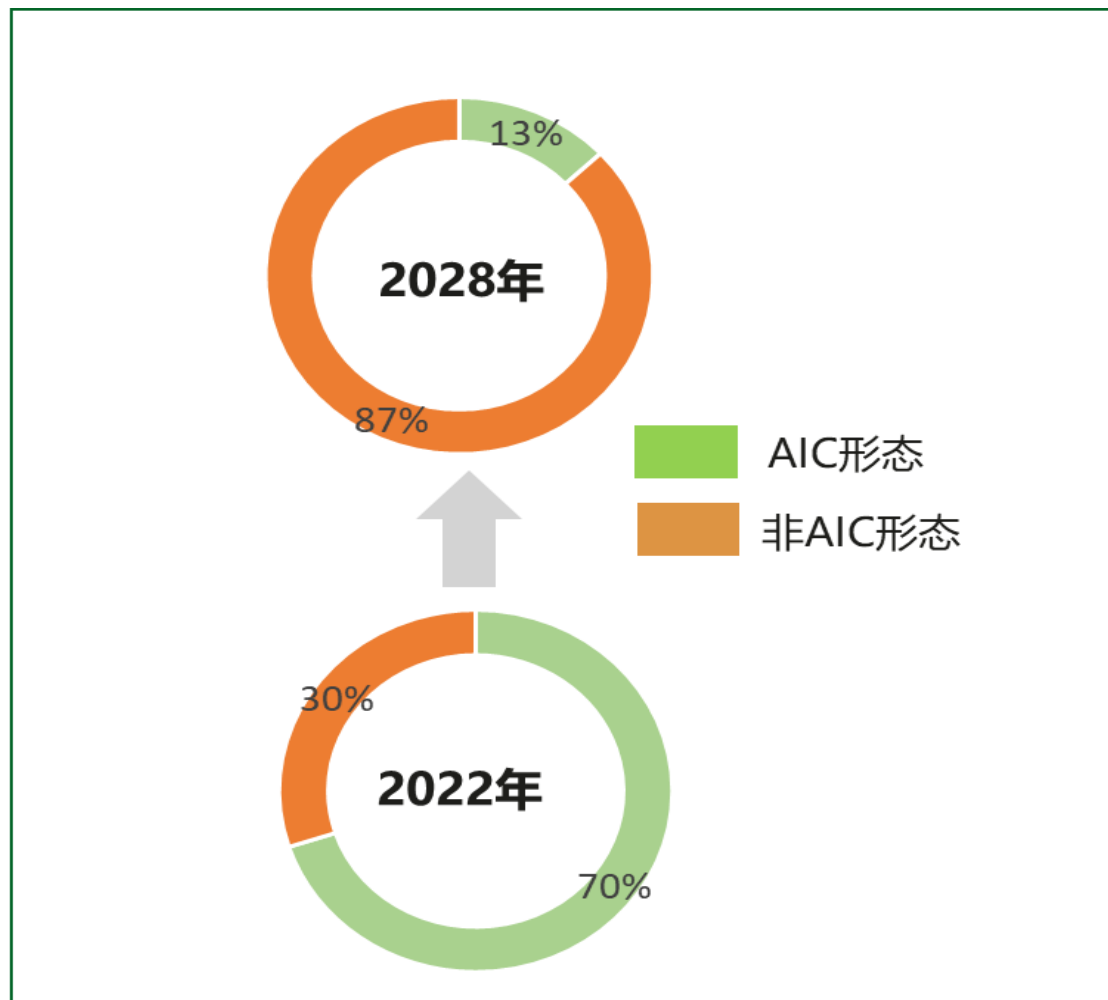
内存池设备

物理形态	大容量	易维护	易扩容	多厂商兼容	特殊需求满足 (器件利旧、空间利用)	配置灵活性 (内存和硬盘灵活转换)
PCIe-AIC	√ (512GB/1TB)	×	×	√	×	×
E3.S/E1.S	×	√	√	×	×	√
非标形态	√ (512GB/1TB)	×	×	×	√	×
JBOM	>6TB	√	√	√	√	√

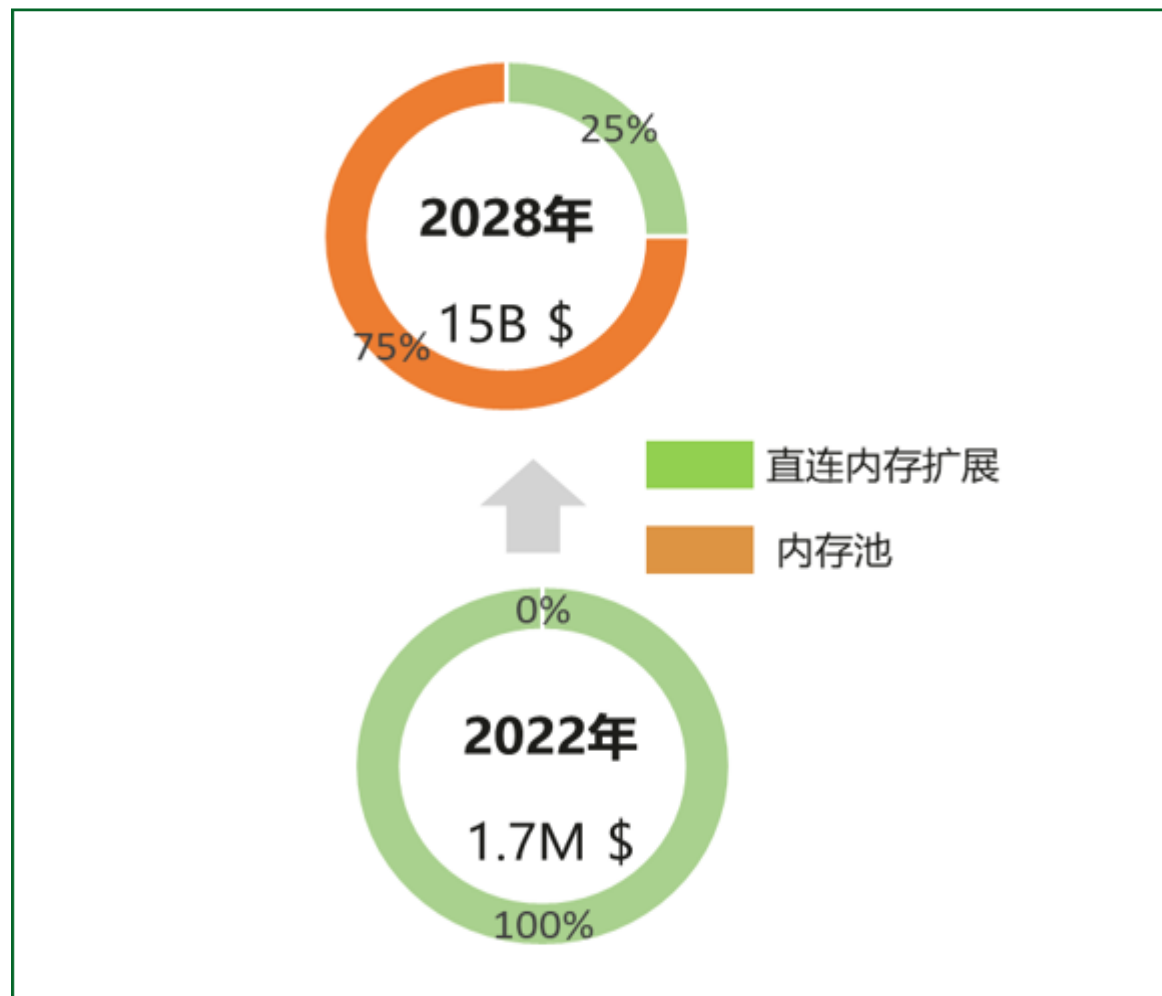
直连内存扩展多形态模组共存，JBOM内存池将成为未来CXL内存应用的主流形态

# CXL内存物理形态及占比预测

非AIC形态将快速增长成为主流



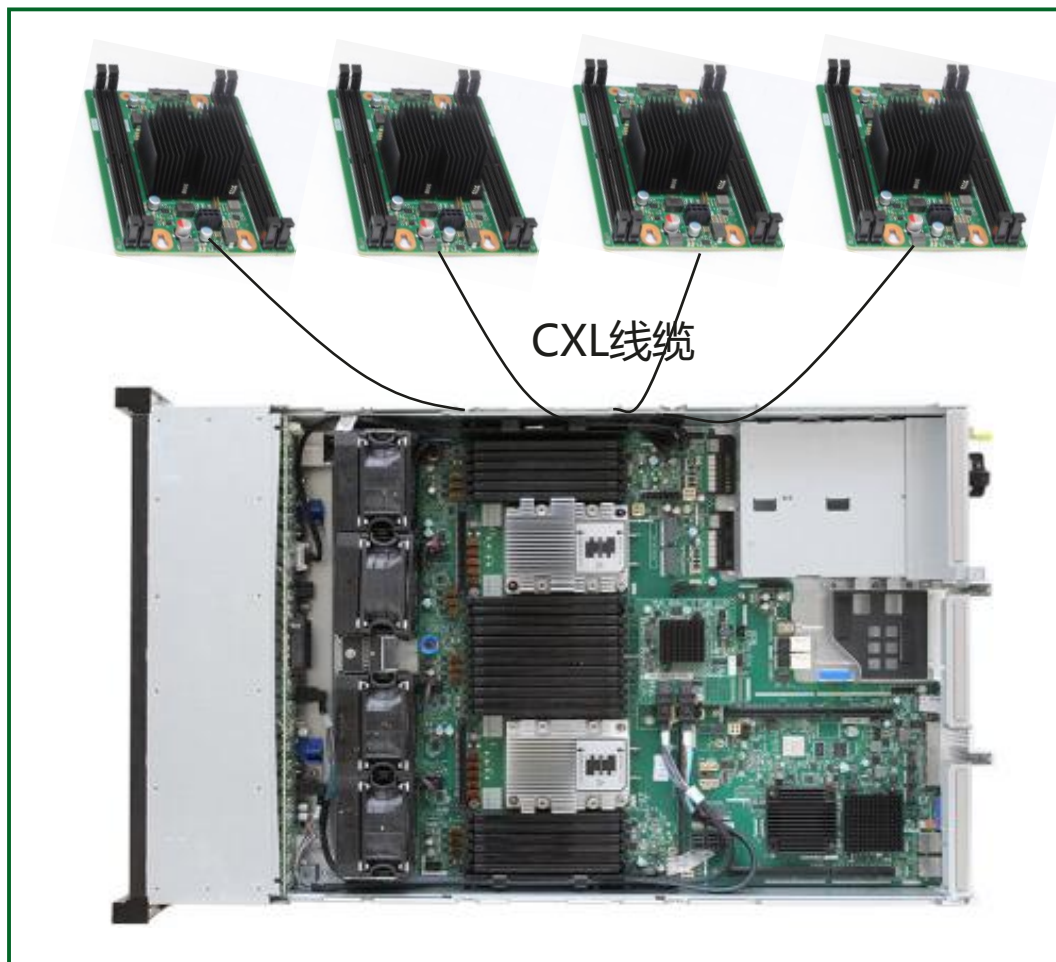
内存池将占主要市场份额





# CXL互连技术应用探索：单服务器扩内存容量，提升内存数据库解决方案性价比

基于CXL通道扩展25%的内存容量



内存数据库典型场景性能影响<5%，业务几乎无影响

+ redis + CXL			测试结果		
物理机内存配置 (本地+CXL)	增加VM 数量比例	32G虚拟机内存分配 (本地: CXL内存)	操作 方式	CXL DDR5 VM性能	CXL DDR4 VM性能
32*32GB + 8*32GB	25%	4:1	只读	100%	100%
			只写	80.6%	90.7%
			读写10:1	96.8%	96%

直连内存扩展在数据库典型应用下提升虚拟机数量，降低单虚拟机成本，提升整体解决方案性价比

# CXL互连技术应用探索：机柜内共享CXL内存，提升云化虚拟机场景内存利用率

通过CXL互连技术进行内存池化，提升机柜内内存利用率，降低内存成本

计算节点集群

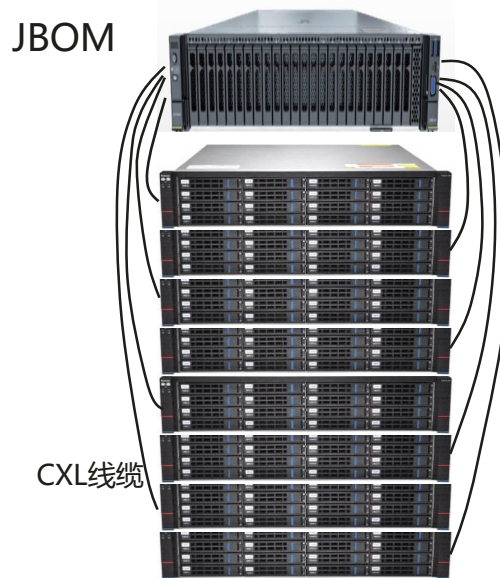


- 单服务器内存：1 TB
- 8台服务器总内存：8 TB

- 单机高配置内存，适应业务扩展需求
- 内存受物理机箱限制，无法共享
- 内存空闲率40%以上，存在资源浪费



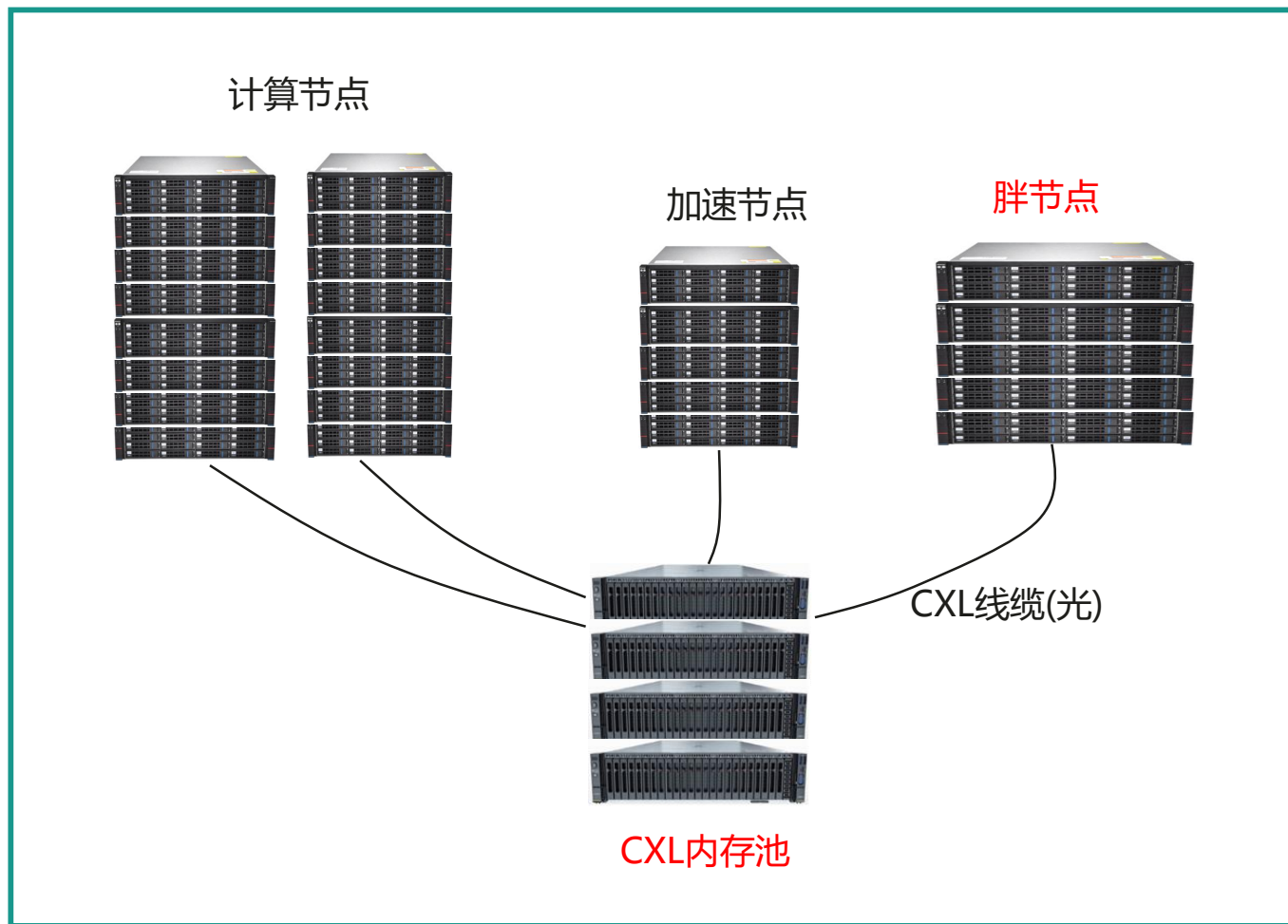
计算节点集群



- 单服务器内存：512GB
- 8台服务器总内存：4TB(8\*512GB)
- JBOM内存：3 TB
- 内存合计：7 TB (节省 12.5%)

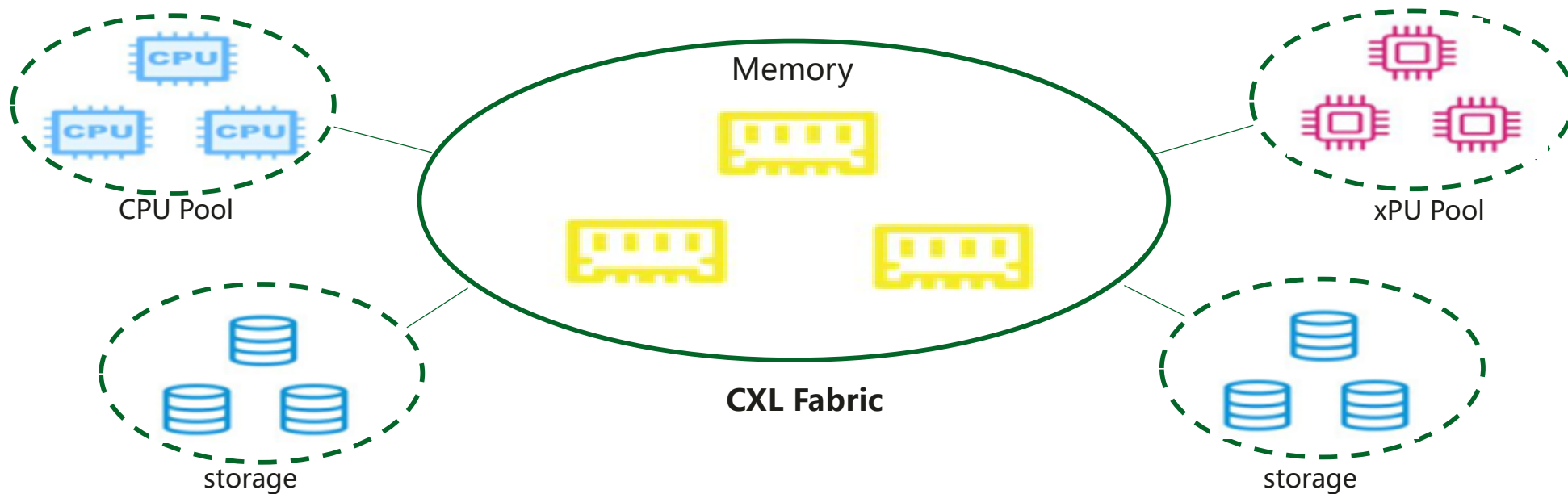
- 单机无需高配内存，通过CXL内存池弹性分配来满足业务扩展需求
- 配置内存总量减少，内存利用率提升10%以上
- 单机通过弹性分配可获取更大的内存

# CXL互连技术应用探索：机柜间共享CXL内存，提升HPC集群业务性能



- **胖节点更“胖”：**  
扩展节点内存容量及带宽，使胖节点算力更强。
- **消除Checkpoint IO瓶颈：**  
将计算过程Checkpoints保存在共享CXL内存池中，消除IO瓶颈，当发生错误时可快速恢复
- **内存共享实现高性能数据传输：**  
节点间大数据传输时，直接将本节点内存切换挂给对端节点，毫秒级完成TB级数据传递

# CXL互连技术展望：以内存为中心的DC创新架构



## 传统数据中心构架:

- “IO墙”，“内存墙”，“功耗墙”问题
- 内存/存储/加速器与CPU紧耦合
- 配置固定，调整空间小，扩展、维护成本高
- 资源利用难均衡，闲置、浪费情况严重



## 下一代数据中心构架:

- 资源模块化设计，异构算力，存储资源，内存资源全部池化
- 资源配置灵活，可调整可共享，按需分配高利用率，扩展灵活易维护



道阻且长，行则将至！



Question?  
questions



# Thank you.

让数字世界无限可能

Fusion X, Digital Infinity

Copyright©2021 xFusion Digital Technologies Co., Ltd.  
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. xFusion may change the information at any time without notice.

