Hot Chips 21

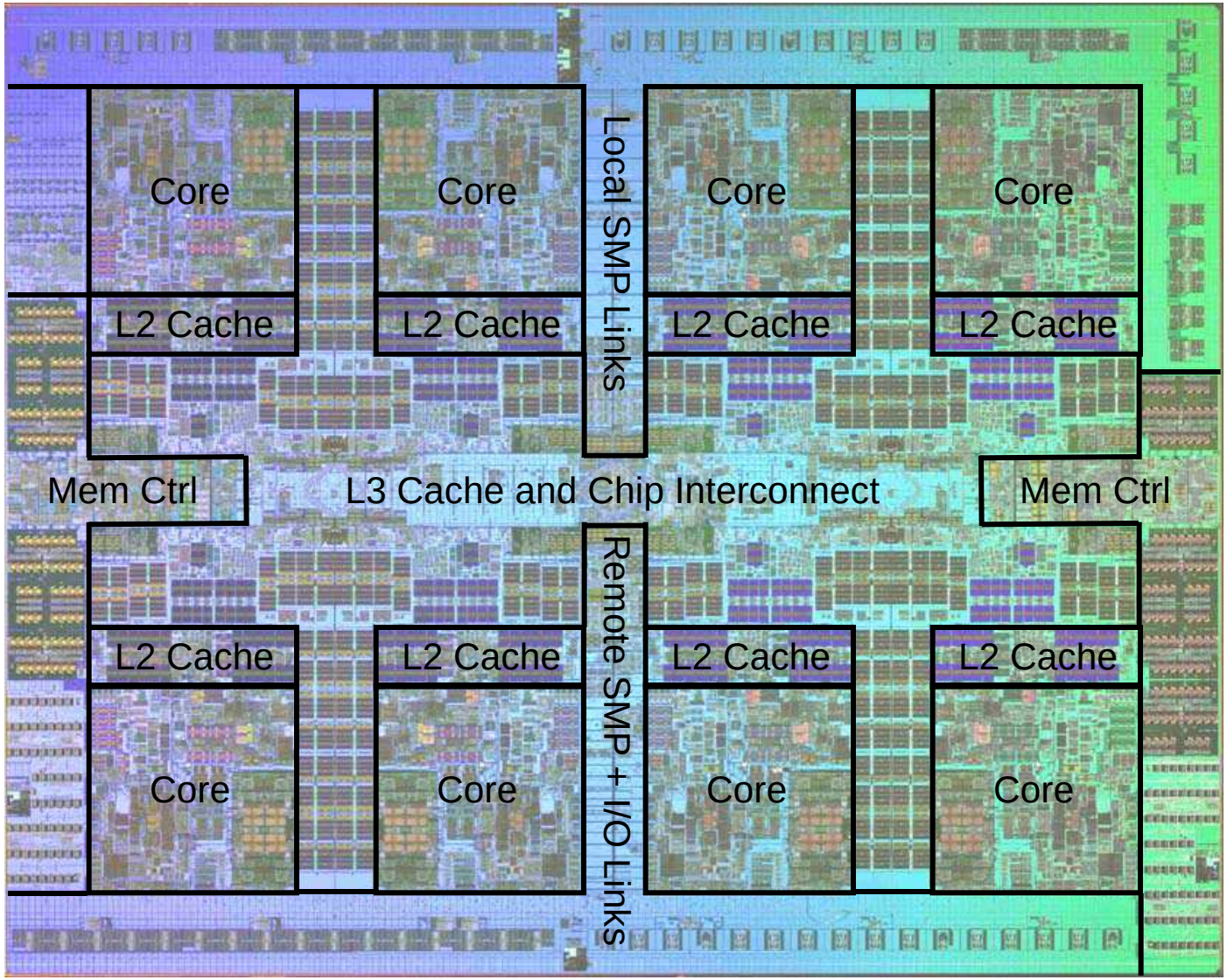
POWER7: IBM's Next Generation, Balanced POWER Server Chip

William J. Starke POWER7 Chief Storage Hierarchy and SMP Architect

Acknowledgment: This material is based upon work supported by the Defense Advanced Research Projects Agency under its Agreement No. HR0011-07-9-0002

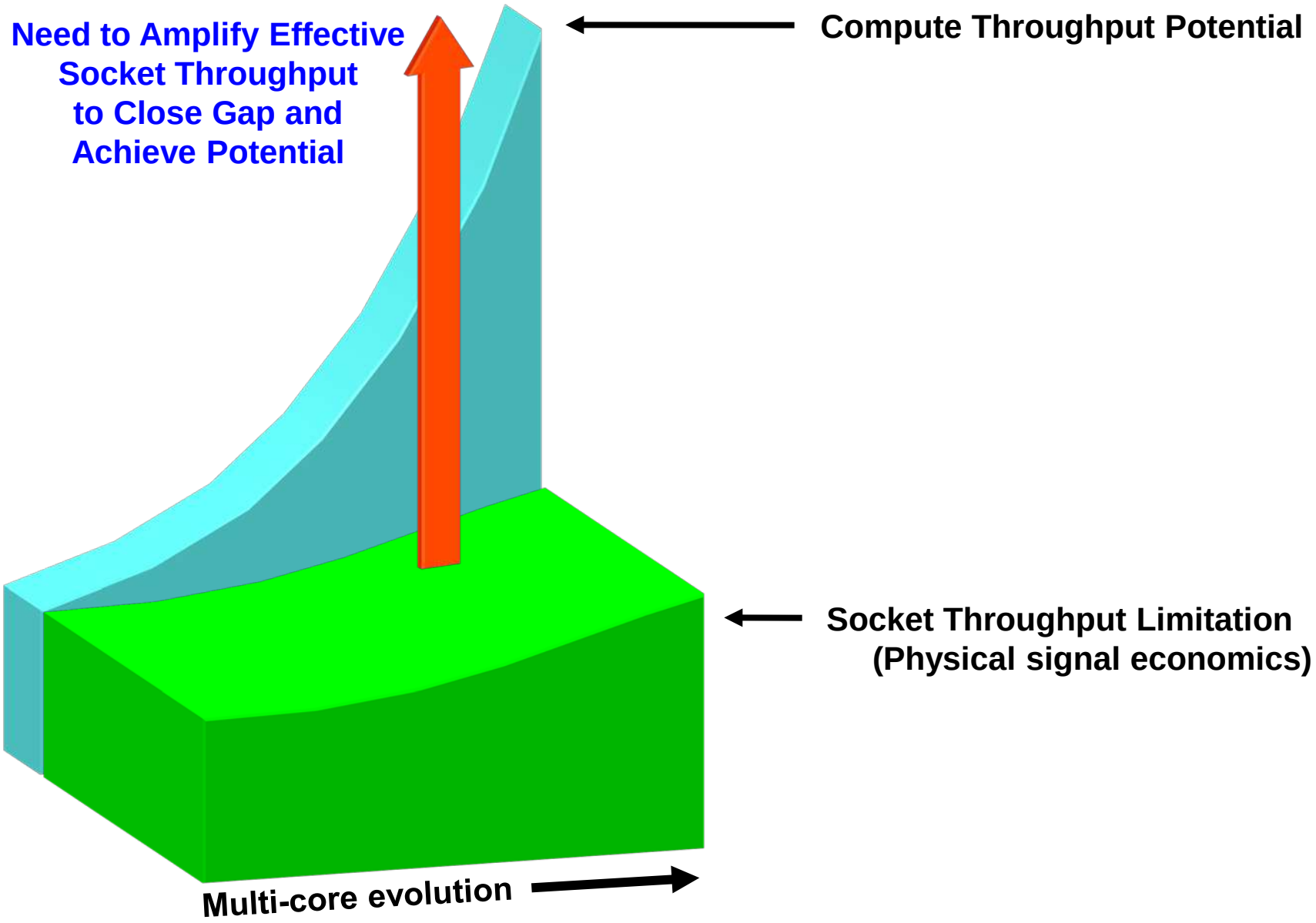
IBM

Challenge: Beating Physics to Realize Multi-core Potential

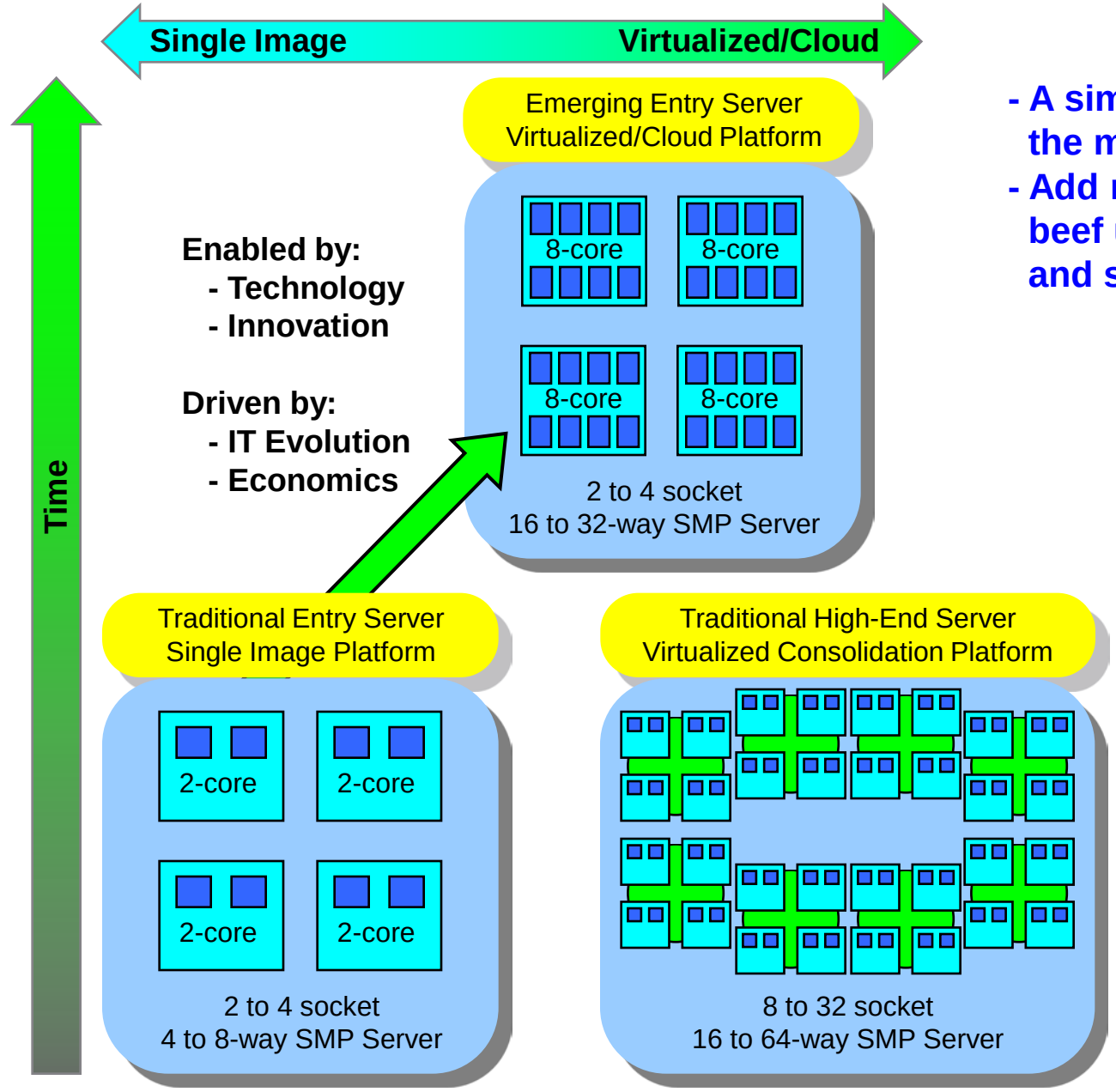


POWER7™ is an 8-core, high performance Server chip. A solid chip is a good start. But to win the race, you need a balanced system. POWER7 enables that balance.

Challenge: Beating Physics to Realize Multi-core Potential



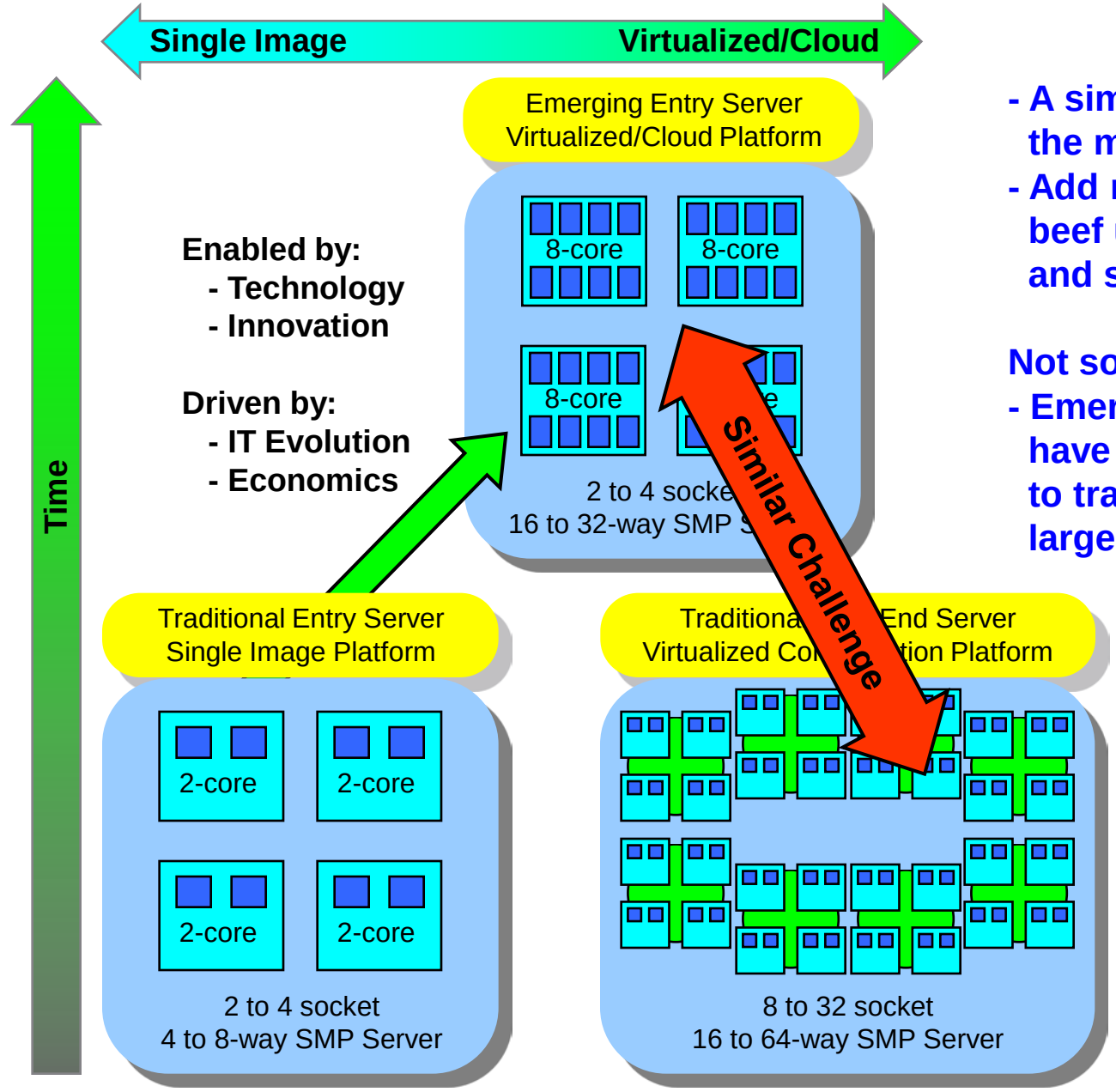
Trends in Server Evolution



- **A simple matter of riding the multi-core trend?**
- **Add more cores to the die, beef up some interfaces, and scale to a large SMP?**

* Statements regarding SMP servers do not imply that IBM will introduce a system with this capability.

Trends in Server Evolution



- A simple matter of riding the multi-core trend?
- Add more cores to the die, beef up some interfaces, and scale to a large SMP?

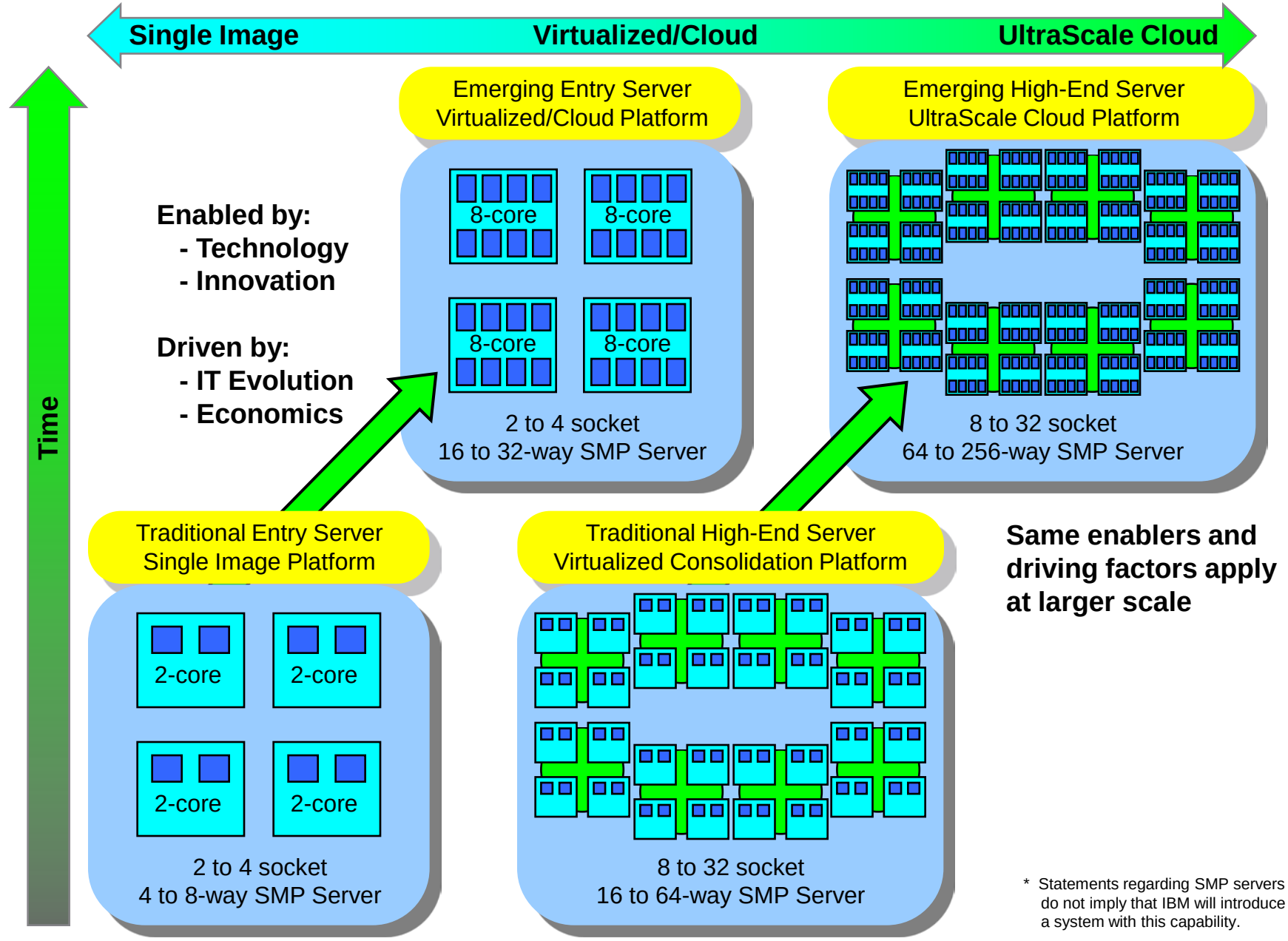
Not so simple:

- Emerging entry servers have characteristics similar to traditional high-end large SMP servers

Achieving solid virtual machine performance requires a Balanced System Structure.

* Statements regarding SMP servers do not imply that IBM will introduce a system with this capability.

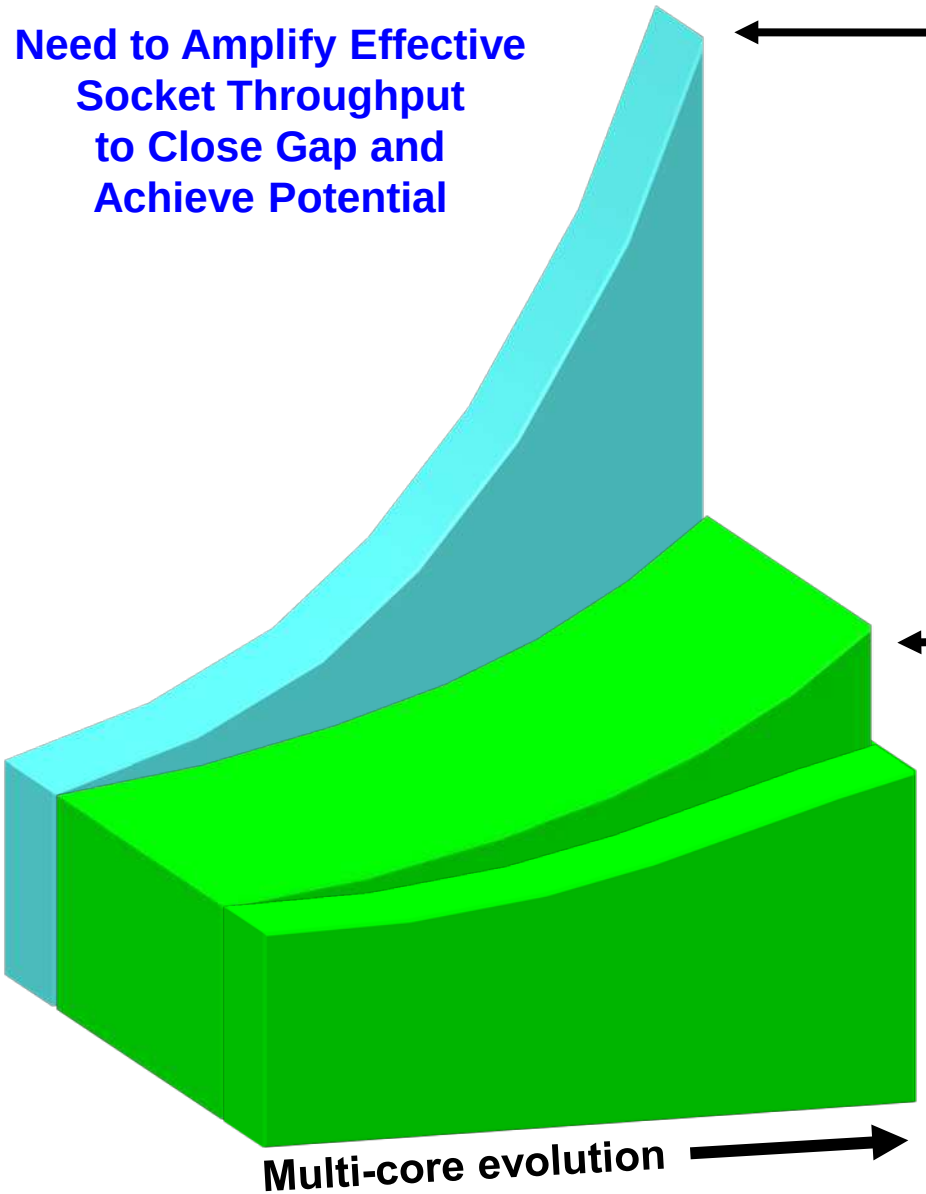
Trends in Server Evolution



Same enablers and driving factors apply at larger scale

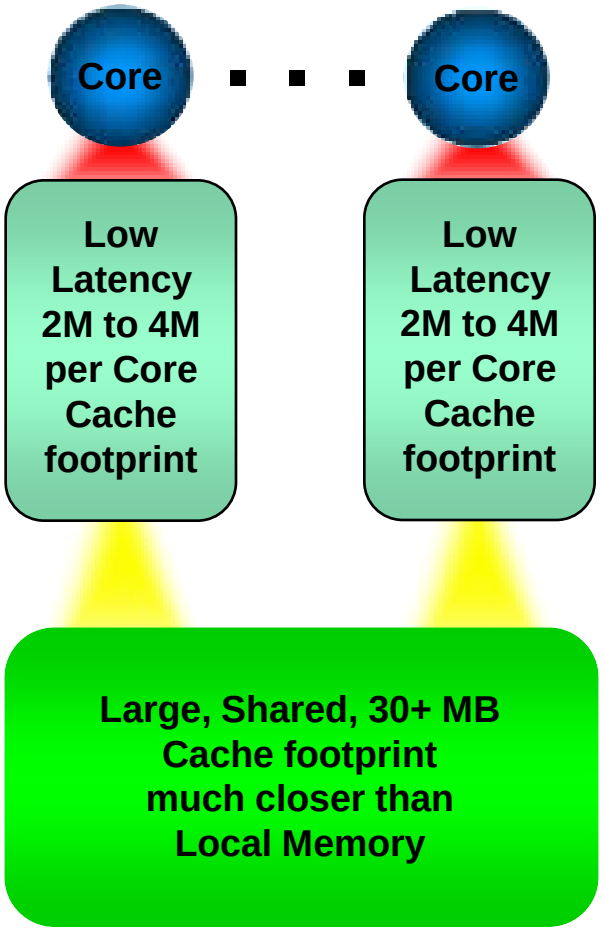
* Statements regarding SMP servers do not imply that IBM will introduce a system with this capability.

Challenge: How does POWER7 maintain the Balance?



Cache Hierarchy Technology and Innovation

Cache Hierarchy Rqmt for POWER® Servers



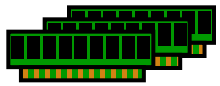
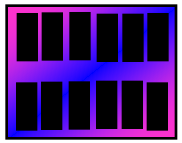
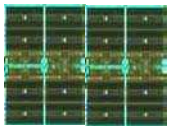
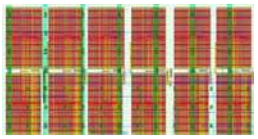
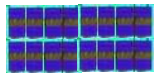
Challenge for Multi-core POWER7

POWER4™, POWER5™, and POWER6™ systems derive huge benefit from high bandwidth access to large, off-chip cache.

But socket pin count constraints prevent scaling the off-chip cache interface to support 8 cores.

Cache Hierarchy Technology and Innovation

Solution: High speed eDRAM on the processor die

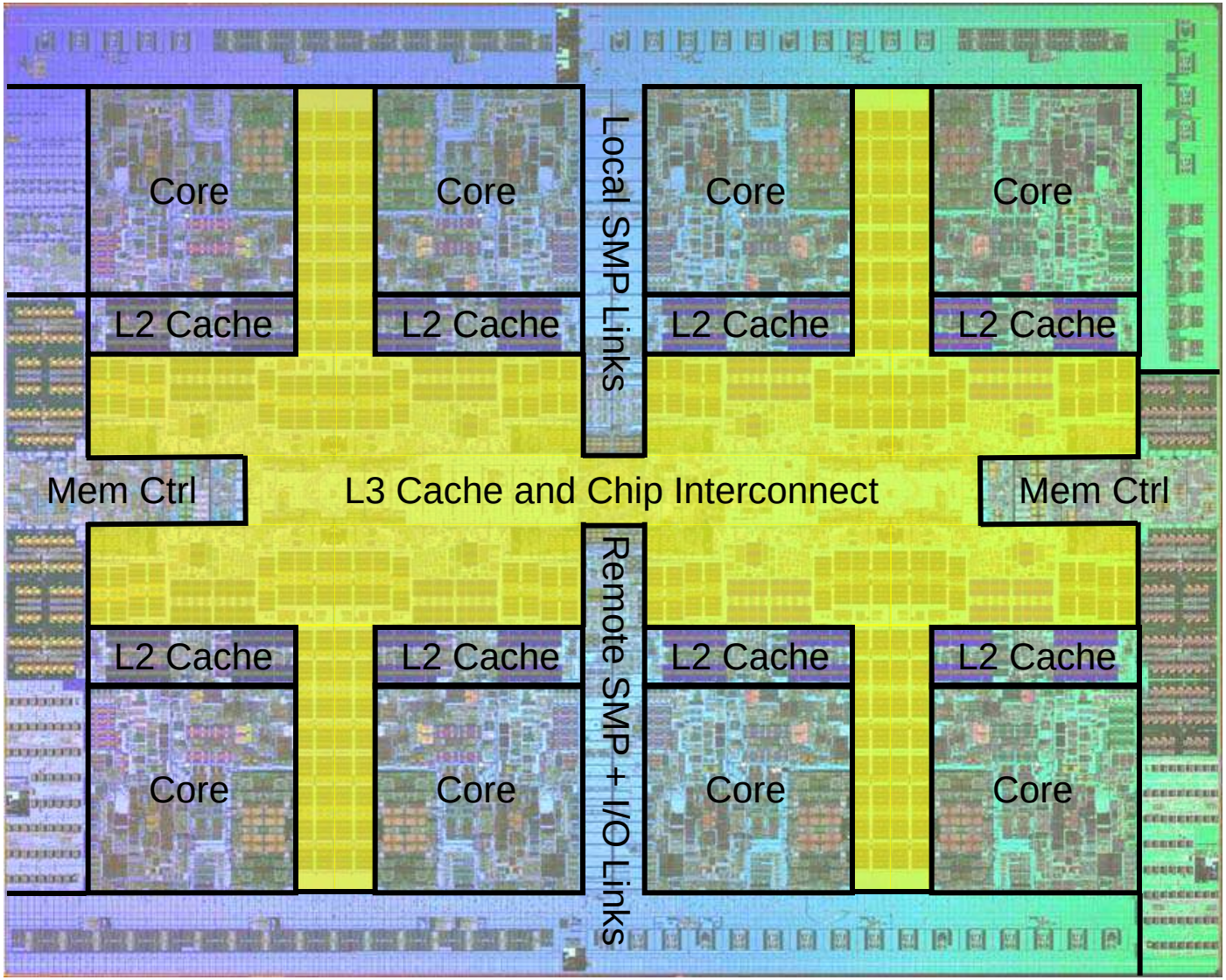
Conventional Memory DRAM	IBM ASIC eDRAM	IBM Custom eDRAM	Custom Dense SRAM	Custom Fast SRAM
Dense, low power Low speed/bandwidth		Off uP Chip | On uP Chip		High Area/power High speed/bandwidth
				
Conventional Memory DIMMs	Large, Off-chip 30+ MB Cache	On-processor 30+ MB Cache	On-processor Multi-MB Cache	Private core Sub-MB Cache

Industry Standard Caching and Memory Technologies: Conventional DIMMs, Dense and Fast SRAM's.

IBM's POWER Servers have leveraged large off-chip eDRAM caches in POWER4, 5, and 6.

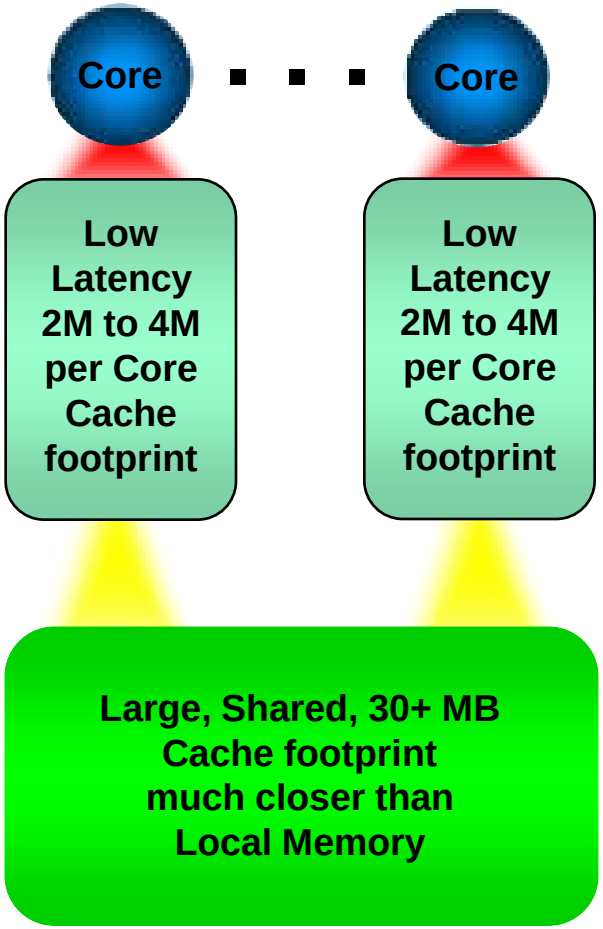
With POWER7, IBM introduces on-processor, high-speed, custom eDRAM, combining the dense, low power attributes of eDRAM with the speed and bandwidth of SRAM.

Cache Hierarchy Technology and Innovation



Cache Hierarchy Technology and Innovation

Cache Hierarchy Rqmt for POWER Servers

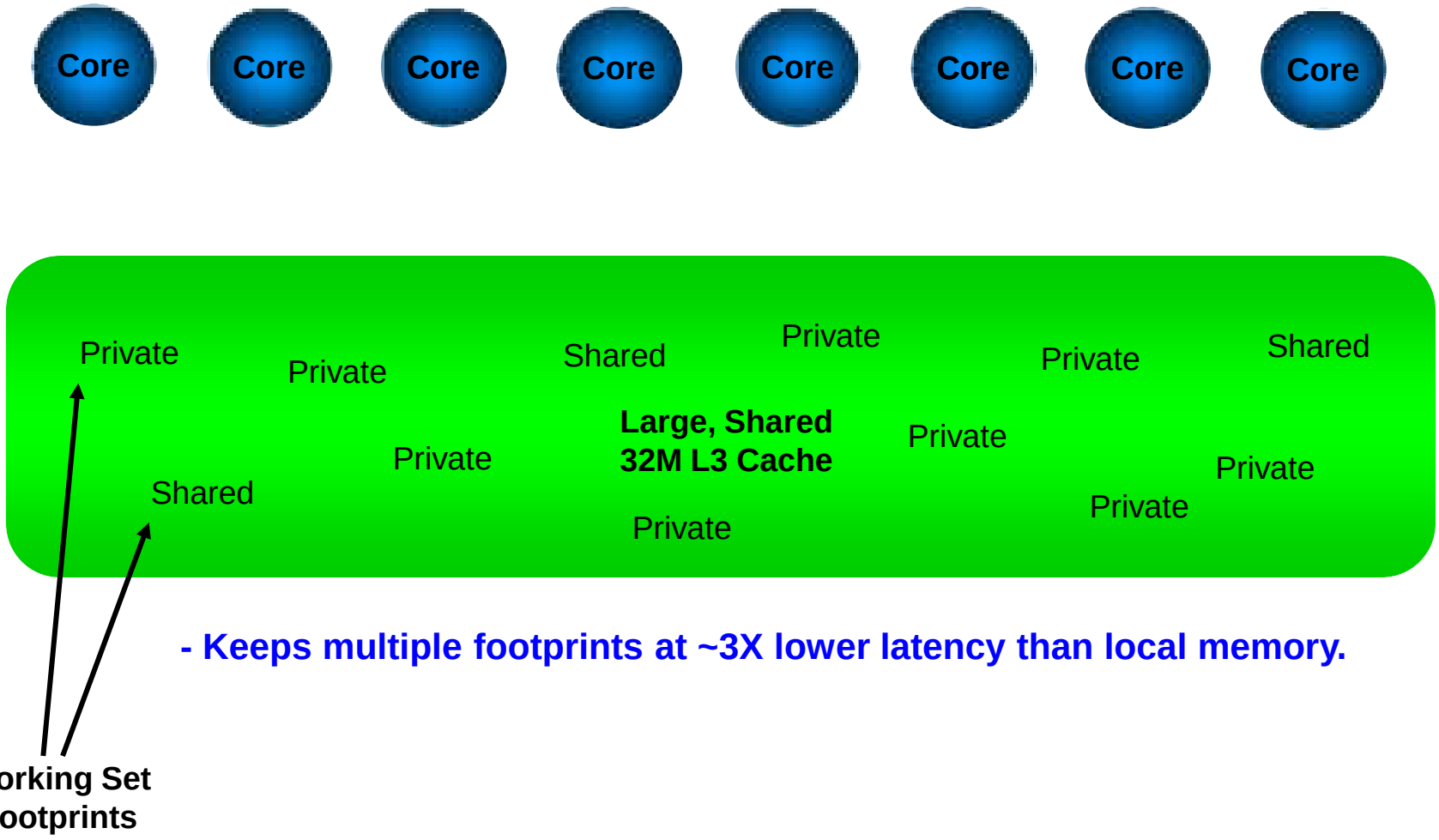


Challenge for Multi-core POWER7

Need to satisfy both caching requirements with one cache.

Cache Hierarchy Technology and Innovation

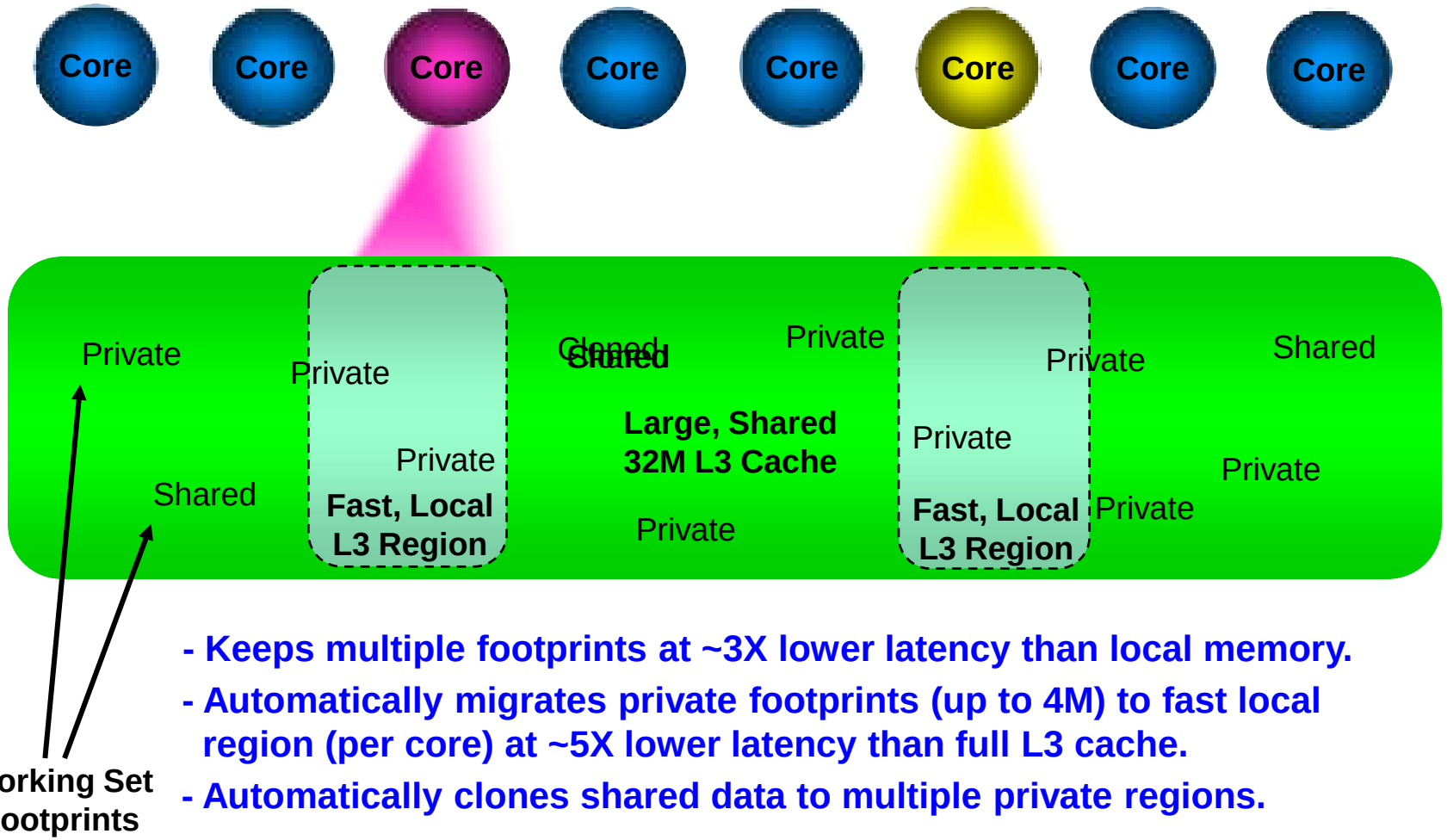
Solution: Hybrid L3 "Fluid" Cache Structure



- Keeps multiple footprints at ~3X lower latency than local memory.

Cache Hierarchy Technology and Innovation

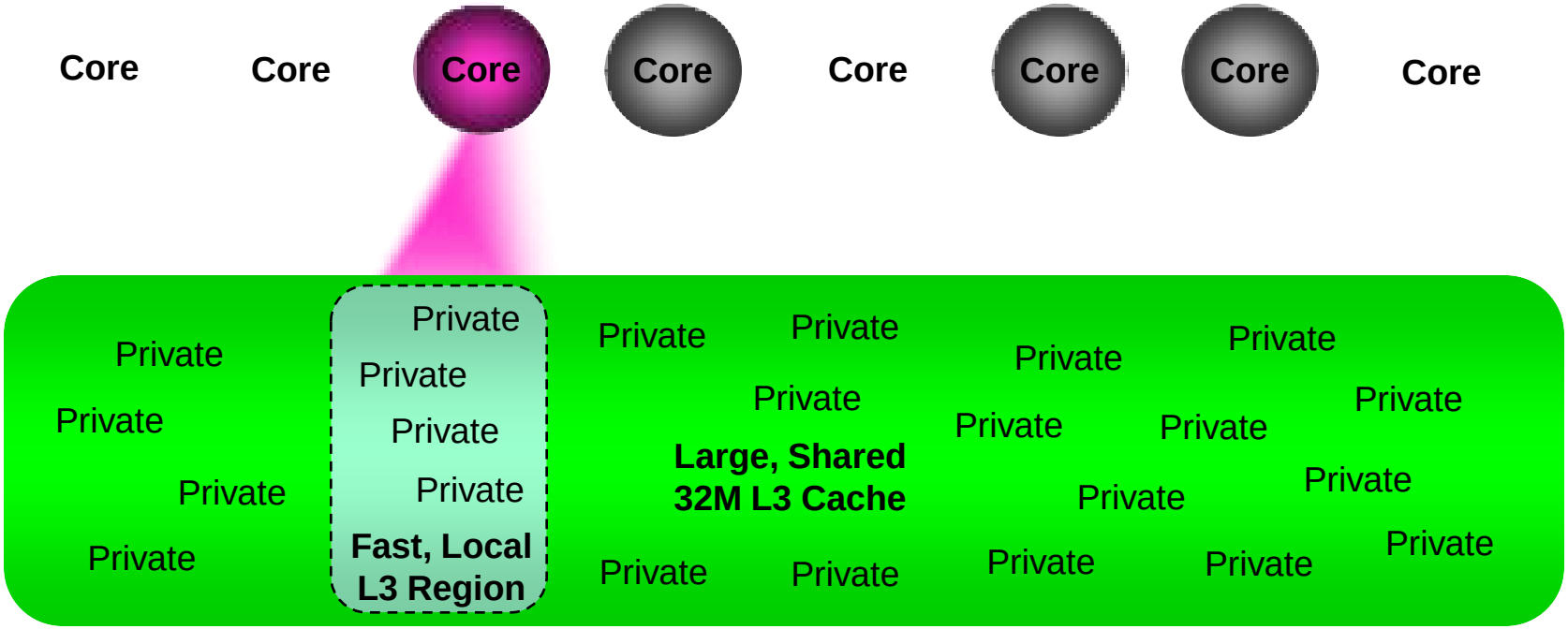
Solution: Hybrid L3 "Fluid" Cache Structure



- Keeps multiple footprints at ~3X lower latency than local memory.
- Automatically migrates private footprints (up to 4M) to fast local region (per core) at ~5X lower latency than full L3 cache.
- Automatically clones shared data to multiple private regions.

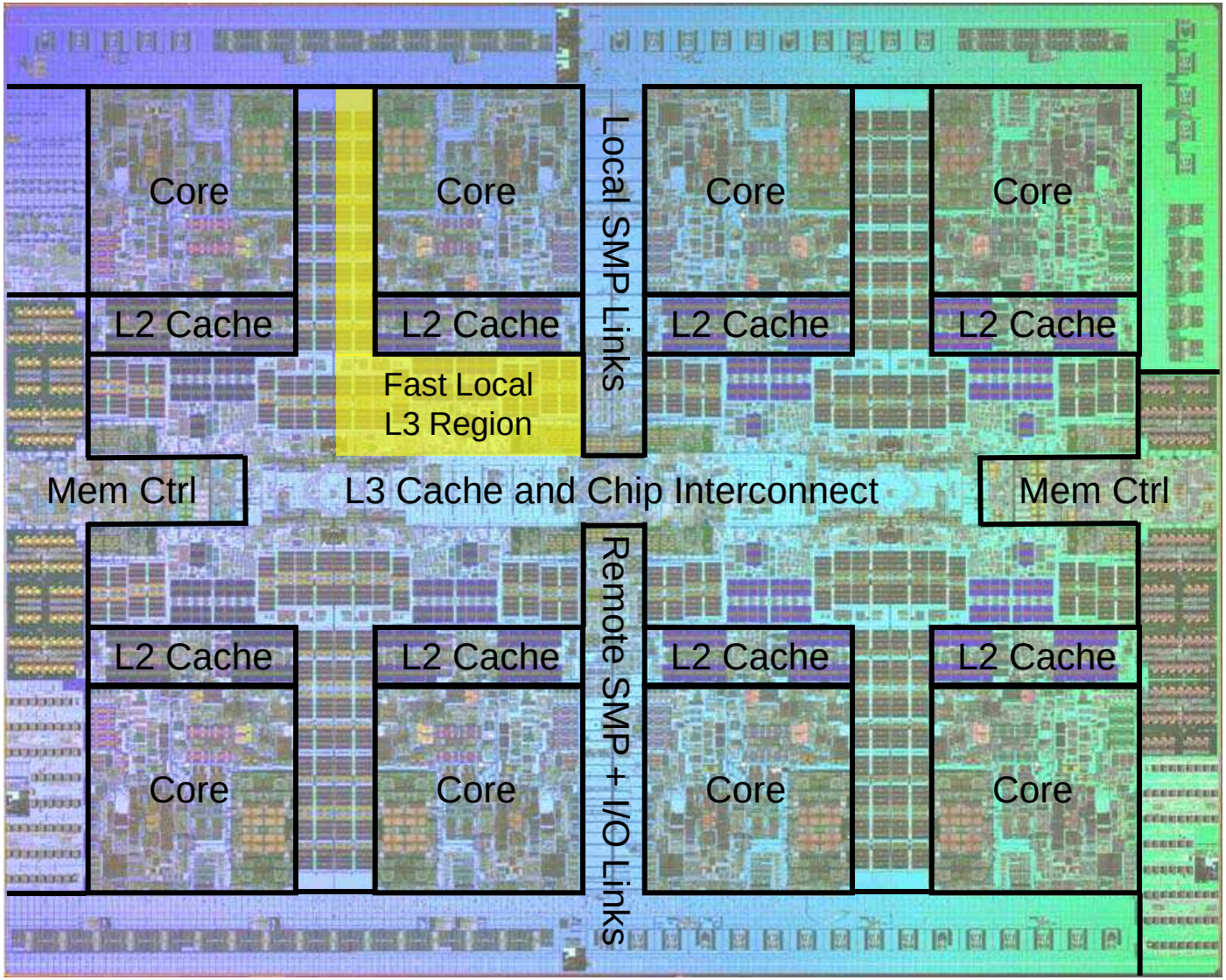
Cache Hierarchy Technology and Innovation

Solution: Hybrid L3 "Fluid" Cache Structure



- Enables a subset of the cores to utilize the entire large shared L3 cache when the remaining cores are not using it.

Cache Hierarchy Technology and Innovation



Cache Hierarchy Technology and Innovation

Cache Hierarchy Rqmt for POWER Servers

Core ▪ ▪ ▪ Core

Low Latency 2M to 4M per Core Cache footprint

Low Latency 2M to 4M per Core Cache footprint

Large, Shared, 30+ MB Cache footprint much closer than Local Memory

Challenge for Multi-core POWER7

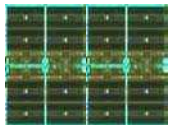
Low power, dense eDRAM is best when complemented by low latency, high bandwidth, fast SRAM structures

IBM Custom eDRAM

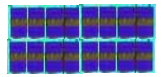
Custom Fast SRAM

Dense, low power
Lower speed/bandwidth

High Area/power
High speed/bandwidth



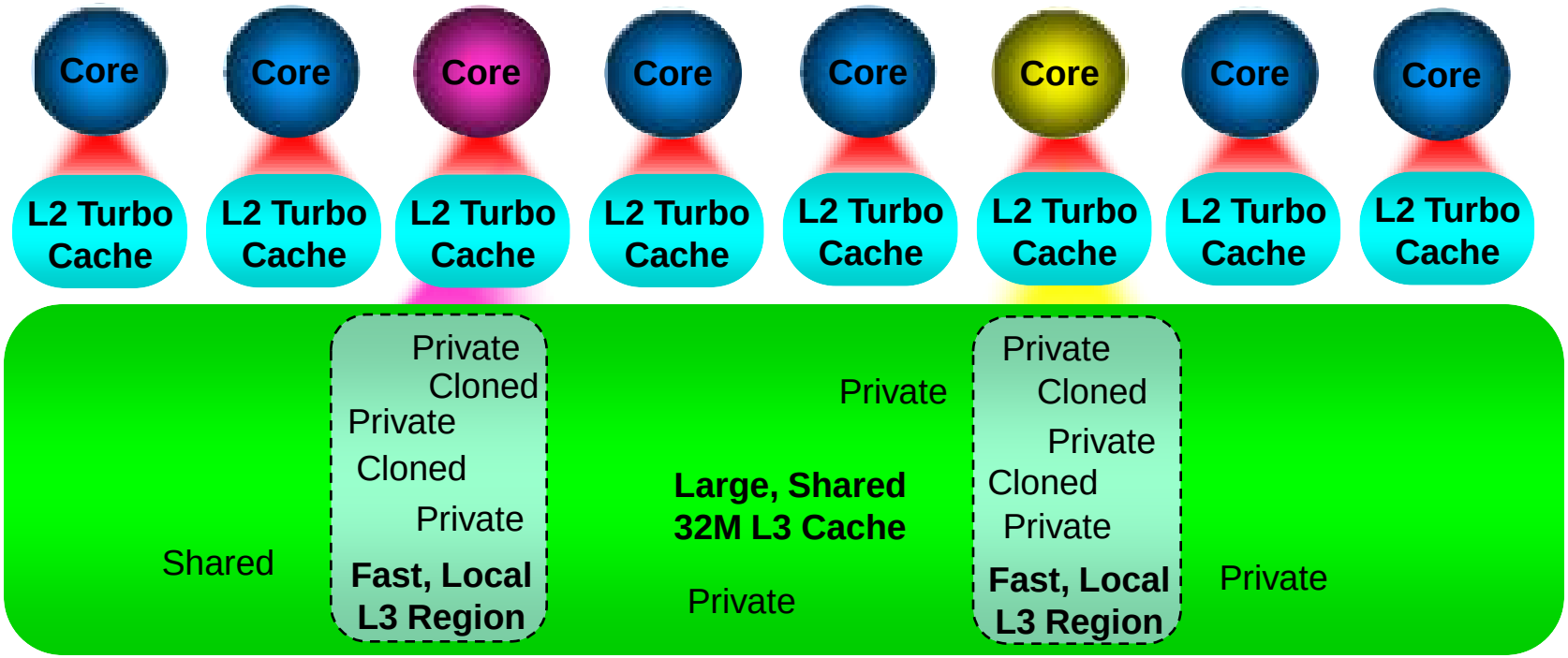
On-processor
30+ MB Cache



Private core
Sub-MB Cache

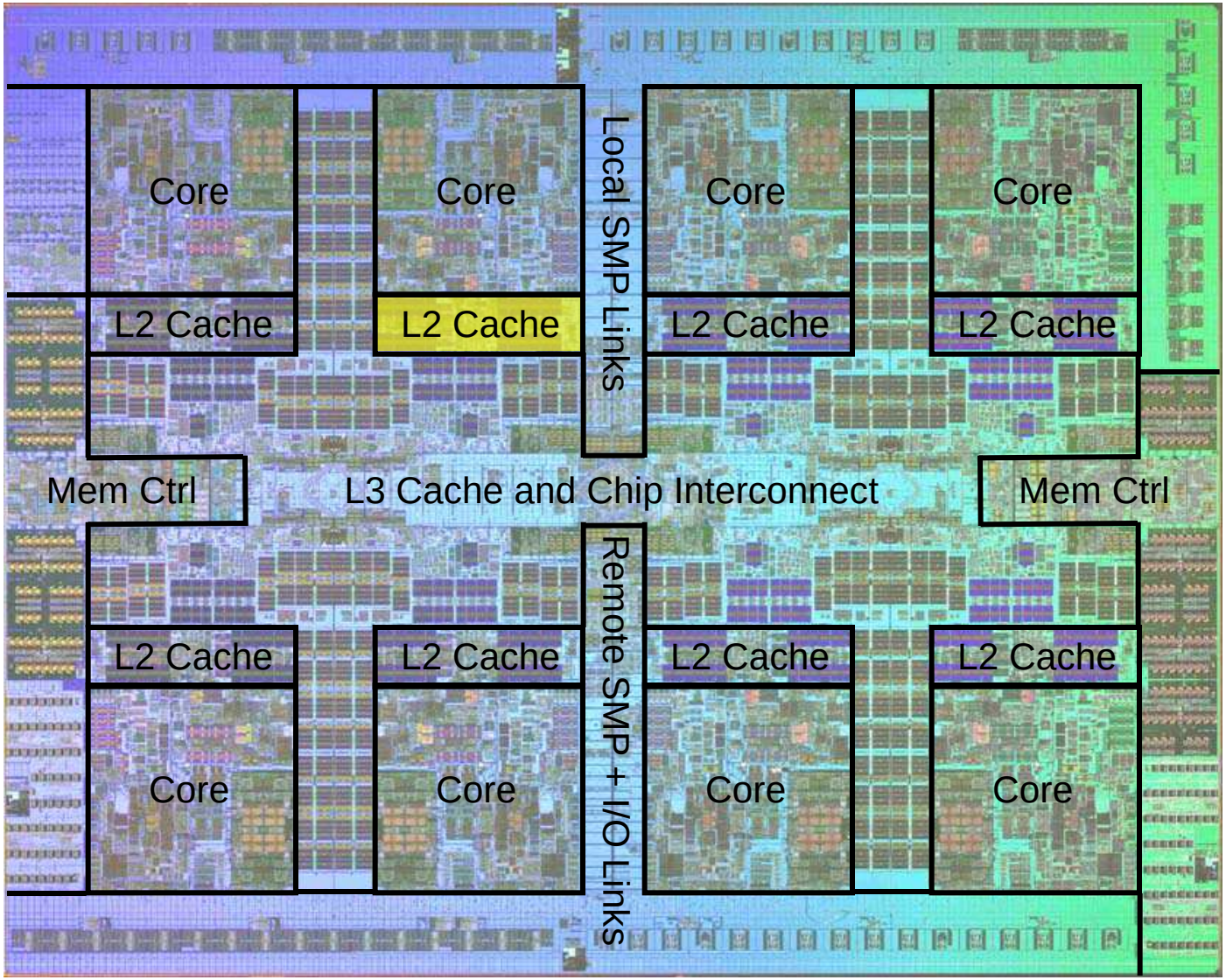
Cache Hierarchy Technology and Innovation

Solution: L2 "Turbo" Cache



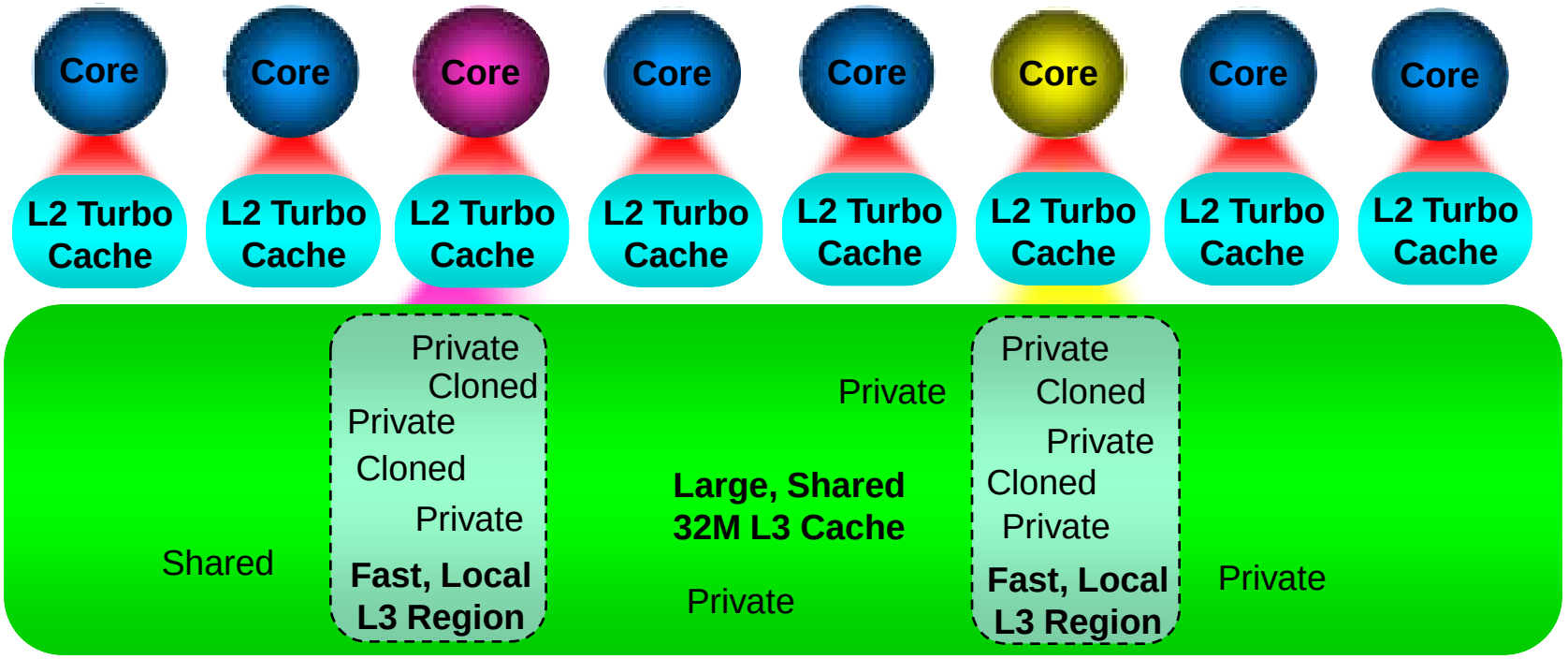
- L2 "Turbo" cache keeps a tight 256K working set with extremely low latency (~3X lower than local L3 region) and high bandwidth, reducing L3 power and boosting performance.

Cache Hierarchy Technology and Innovation



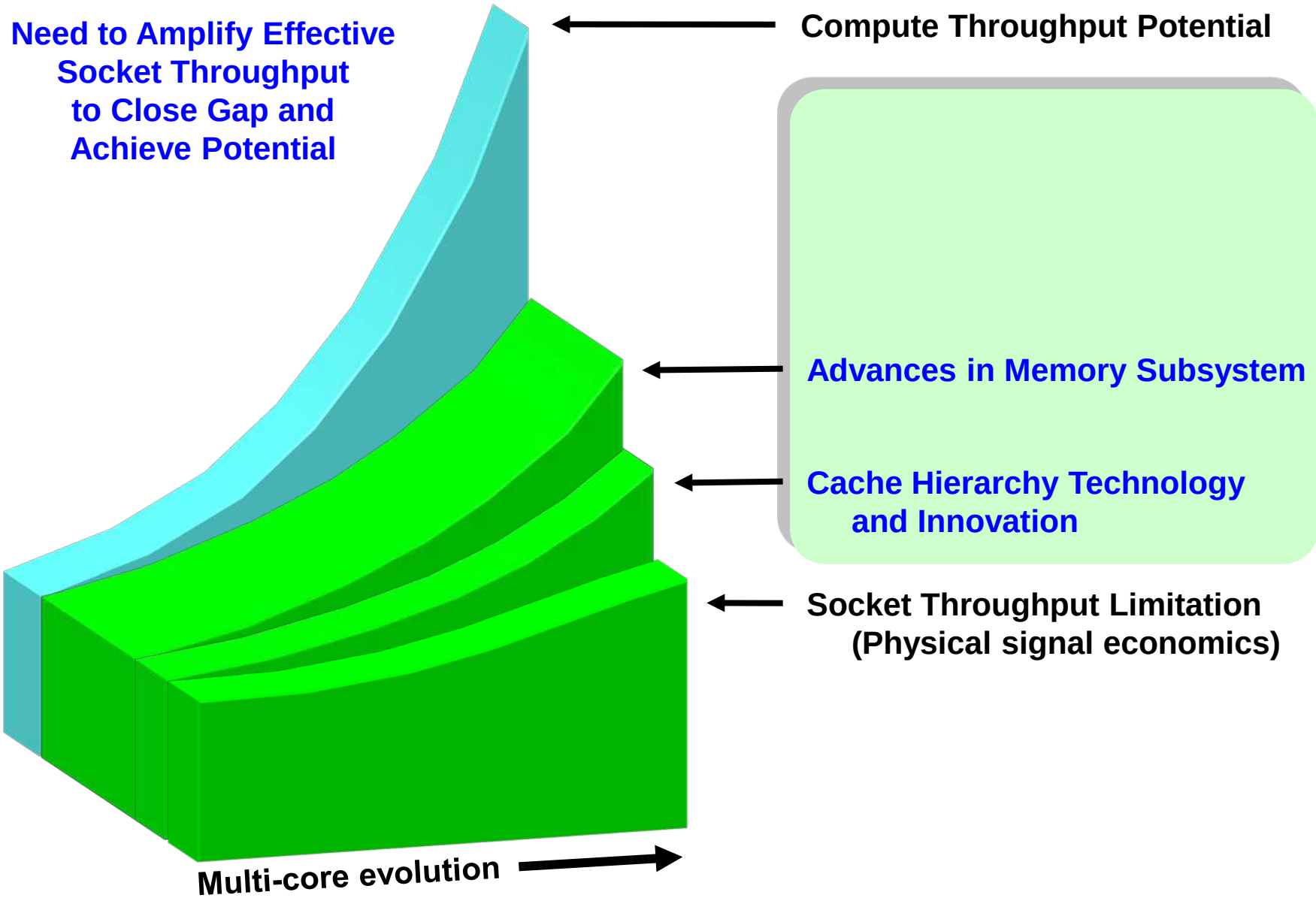
Cache Hierarchy Technology and Innovation

Cache Hierarchy Summary



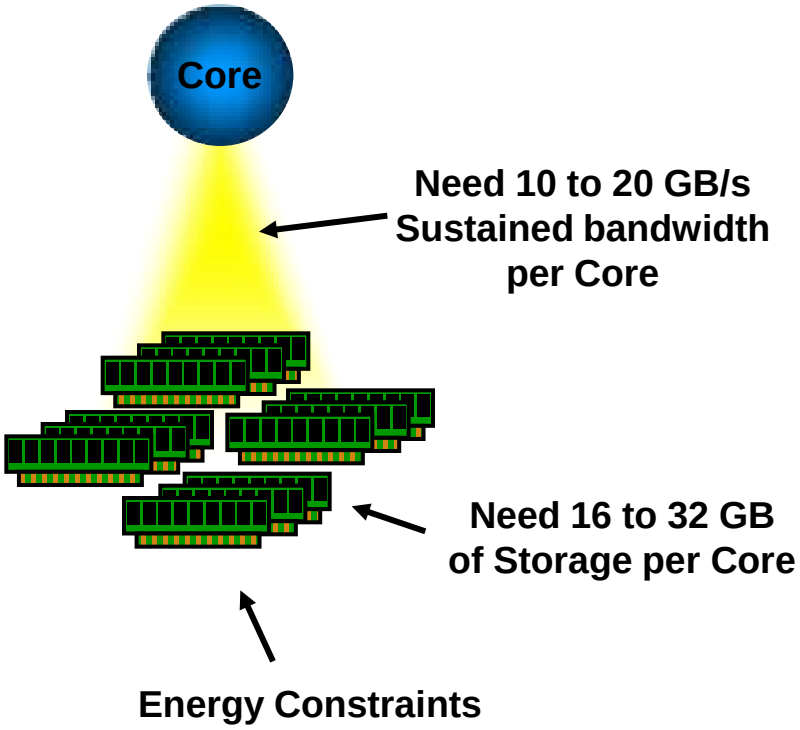
Cache Level	Capacity	Array	Policy	Comment
L1 Data	32K	Fast SRAM	Store-thru	Local thread storage update
Private L2	256K	Fast SRAM	Store-In	De-coupled global storage update
Fast L3 Region	Up to 4M	eDRAM	Partial Victim	Reduced power footprint (up to 4M)
Shared L3	32M	eDRAM	Adaptive	Large 32M shared footprint

Challenge: How does POWER7 maintain the Balance?



Advances in Memory Subsystem

Memory Subsystem Rqmt for POWER Servers



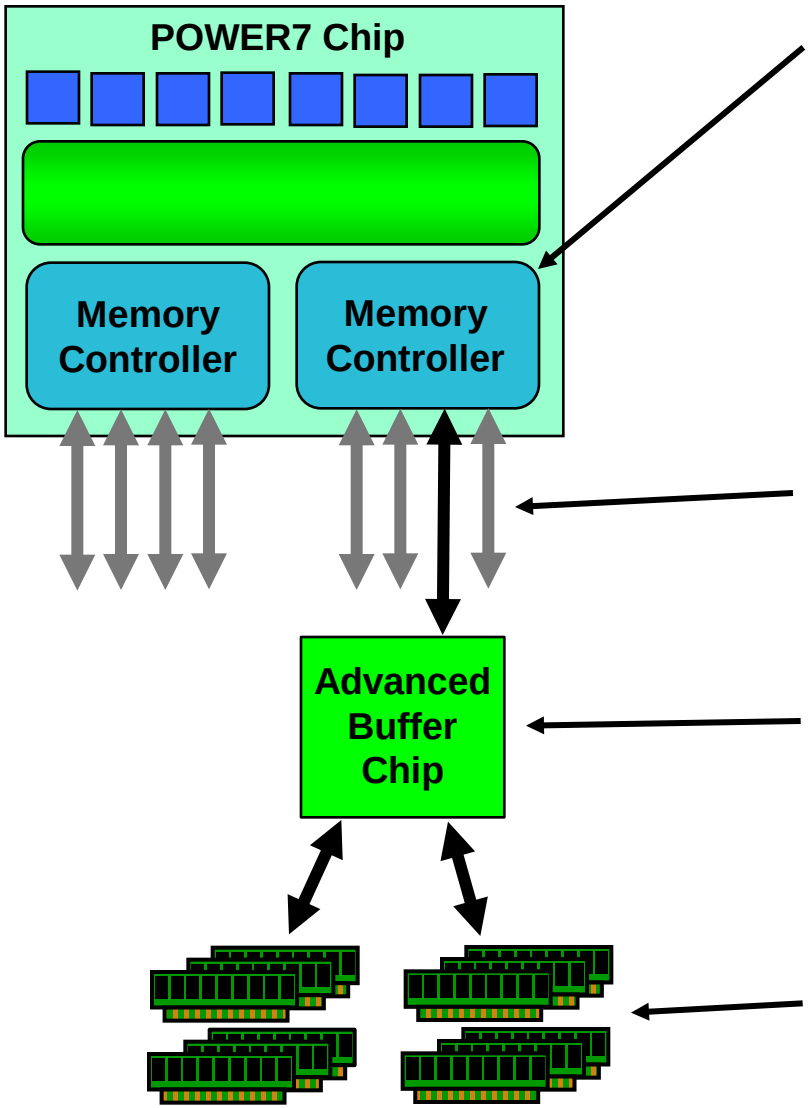
Challenge for Multi-core POWER7

Socket Challenge:
4x growth in memory bandwidth and capacity needed per socket.

System Challenge:
Packaging more memory into similar volume with similar energy and cooling constraints.

Advances in Memory Subsystem

Multi-faceted Solution



1) Dual Integrated DDR3 Controllers
- Massive 16KB scheduling window per POWER7 chip insures high channel and DIMM utilization
- Sparse access acceleration
- Advanced Energy Management
- Numerous RAS advances

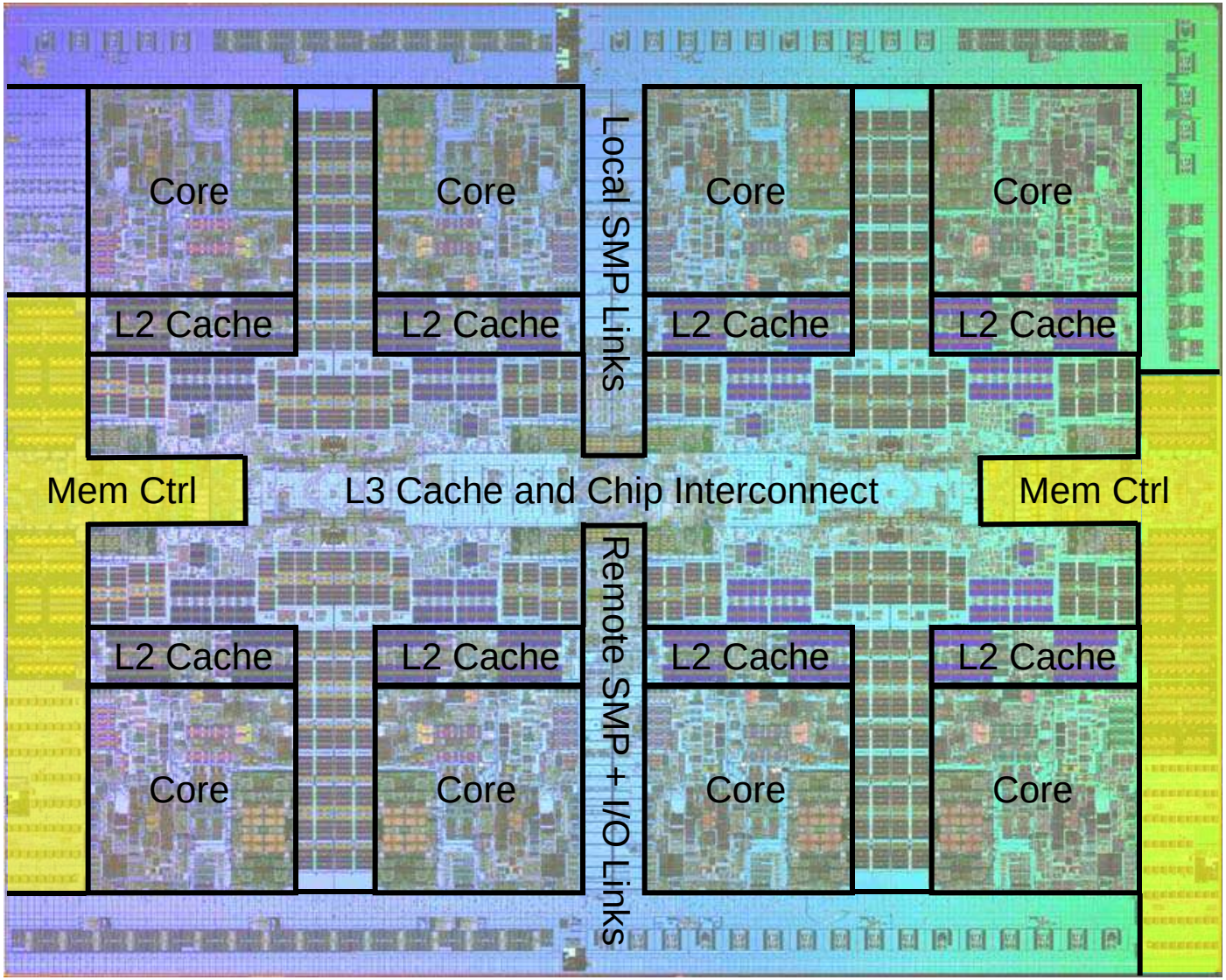
2) Eight high speed 6.4 GHz channels
- New low power differential signaling
- Sustained 100+ GB/s per socket

3) New DDR3 buffer chip architecture
- Larger capacity support (32 GB / core)
- Energy Management support
- RAS enablement

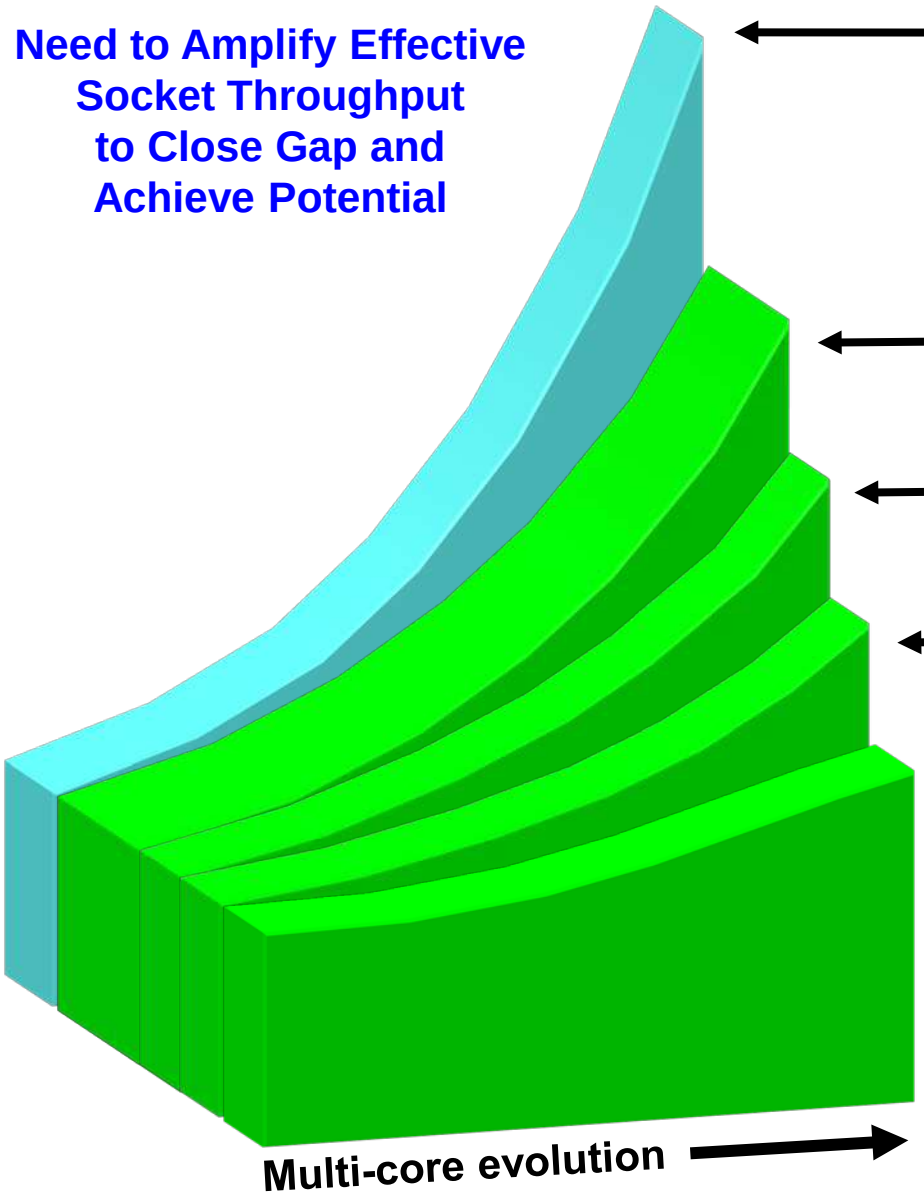
4) DDR3 DRAMs
- Supports 800, 1066, 1333, and 1600

* Statements regarding memory subsystem features do not imply that IBM will introduce a system with these capabilities.

Advances in Memory Subsystem



Challenge: How does POWER7 maintain the Balance?



Advances in Off-chip Signaling Technology

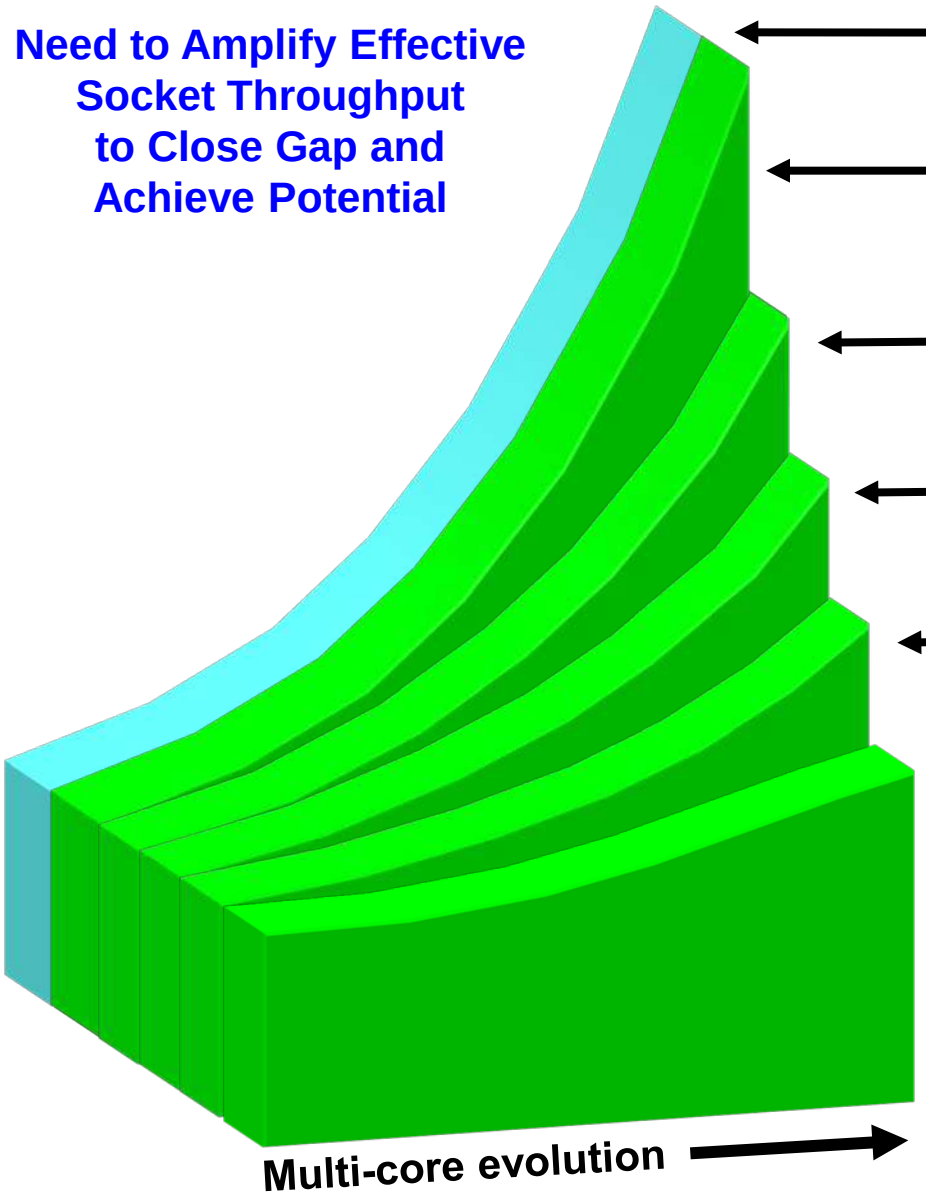
1) Enhanced Signal-ended "Elastic Interface" Technology
2) New high speed, low power Differential Technology

Interface	Signal Type	Info Width	Frequency	Bandwidth
Off-chip Cache	none	none	none	none
Memory Channels	Differential	28 bytes	6.4 Ghz	180 GB/s
I/O Bridge	Single-ended	20 bytes	2.5 Ghz	50 GB/s
SMP Interconnect	Single-ended	120 bytes	3.0 Ghz	360 GB/s
Total Bandwidth				590 GB/s

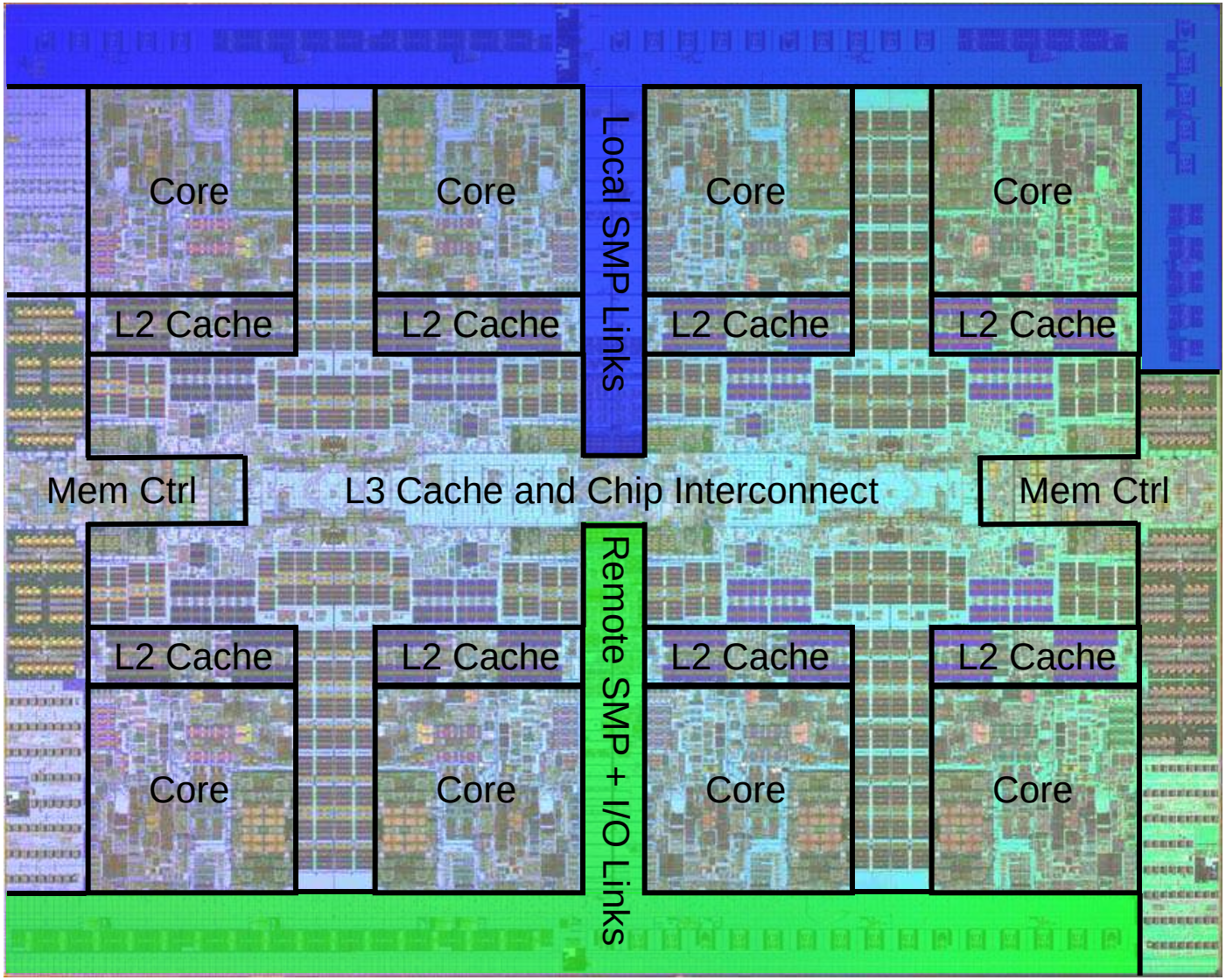
(Note that bandwidths shown are raw, peak signal bandwidths)

- Moving L3 onto POWER7 along with advances in signaling technology enables significant raw bandwidth growth for both memory and I/O subsystems. Note that advanced scheduling improves POWER7's ability to utilize memory bandwidth.

Challenge: How does POWER7 maintain the Balance?

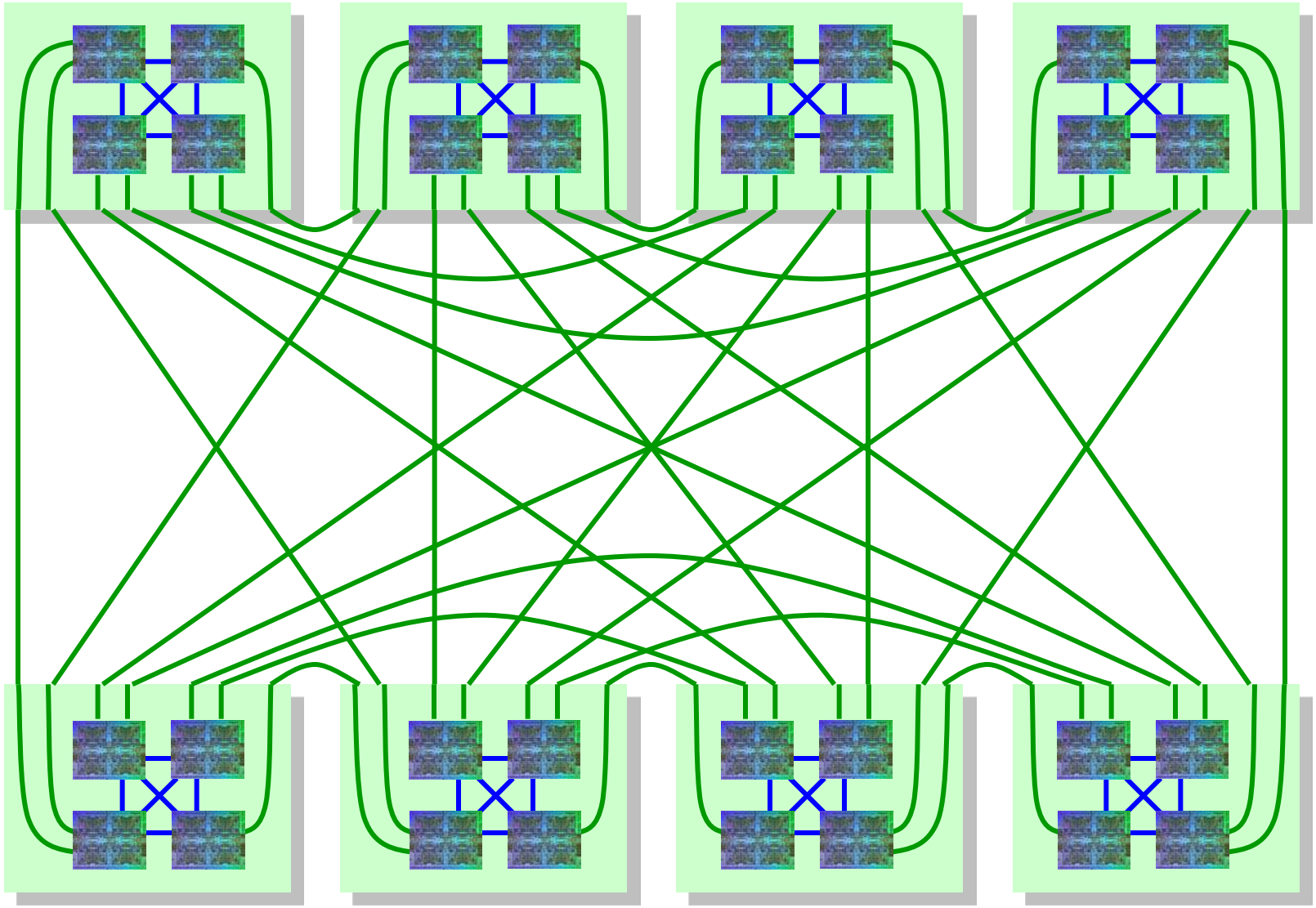


Exploit Long Term Investment in Coherence Innovation



Using local and remote SMP links, up to 32 POWER7 chips are connected

Exploit Long Term Investment in Coherence Innovation



Up to 32 POWER7 chips form a massive SMP system.

* Statements regarding SMP servers do not imply that IBM will introduce a system with this capability.

Exploit Long Term Investment in Coherence Innovation

Coherence Protocol Features	POWER7 Exploitation
- POWER storage Architecture enables decoupled global storage updates. Updates can be reordered and are effectively "deserialized". - Decentralized coherence resolution, and bounded latency broadcast transport layer.	- POWER Servers can drive massive coherence throughput. A 32-chip POWER7 system can manage over 20,000 concurrently reordered coherent storage operations (~4X more than POWER6 systems), with minimal tracking overhead per operation.
- Decentralized coherence resolution, advanced cache states, optimized on-chip transport, and broadcast free barriers.	- Low latency intervention, high performance locking constructs, and robust scaling.

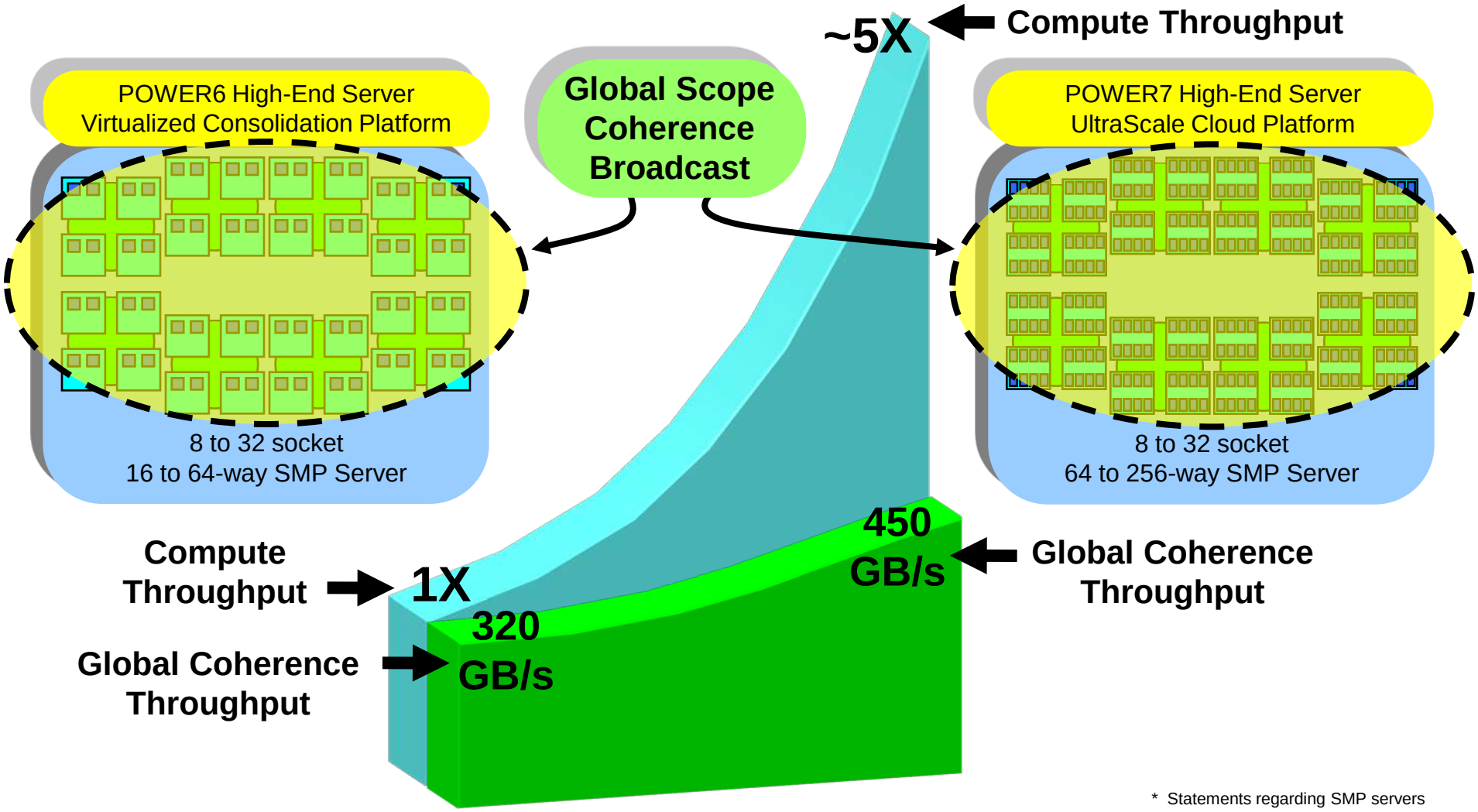
Key Ingredients for Balanced Scaling in Traditional POWER Servers:

- **Architecture enables re-ordered, decoupled storage updates**
- **Decentralized coherence resolution**
- **Broadcast transport layer**

* Statements regarding SMP servers do not imply that IBM will introduce a system with this capability.

Exploit Long Term Investment in Coherence Innovation

Challenge: As system size grows, Coherence broadcast traffic increases

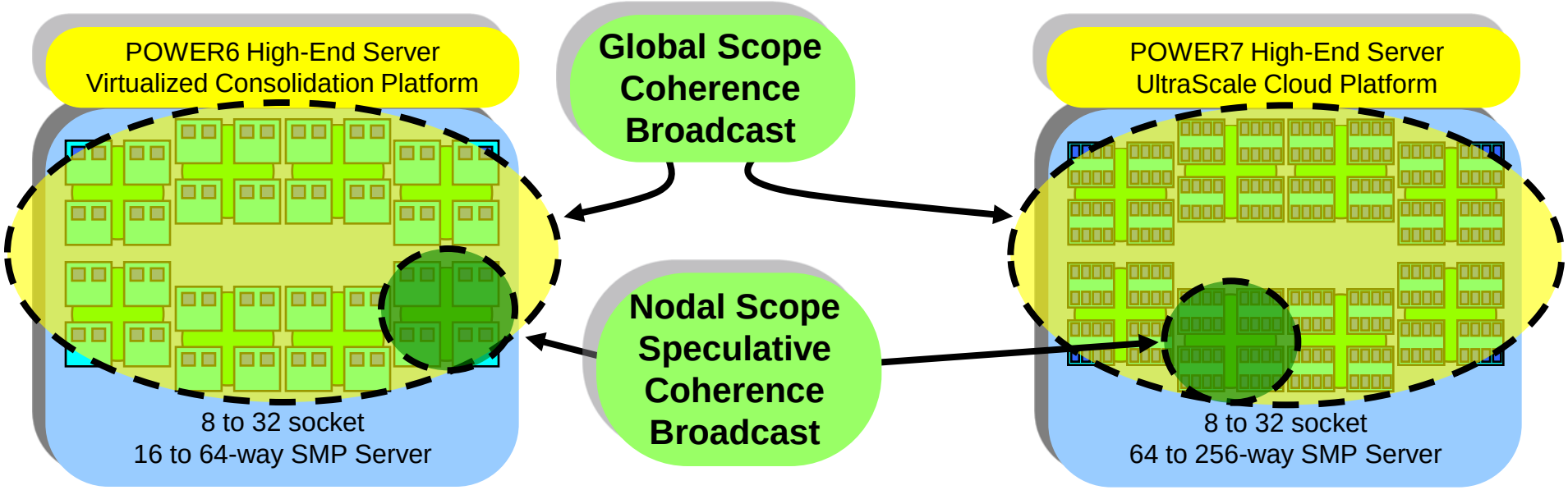


* Statements regarding SMP servers do not imply that IBM will introduce a system with this capability.

Exploit Long Term Investment in Coherence Innovation

Solution: Speculative limited scope Coherence broadcast

- **In 2003, recognized emerging trend**
- **Developed Dual-Scope Broadcast Coherence Protocol for POWER6**
- **Utilizes 13 cache states and integrated scope indicator in memory**



Provides value for POWER6

- Latency reduction
- Near Perfect Scaling for extreme memory intensive workloads
- Ultra-dense packaging (Power 575)

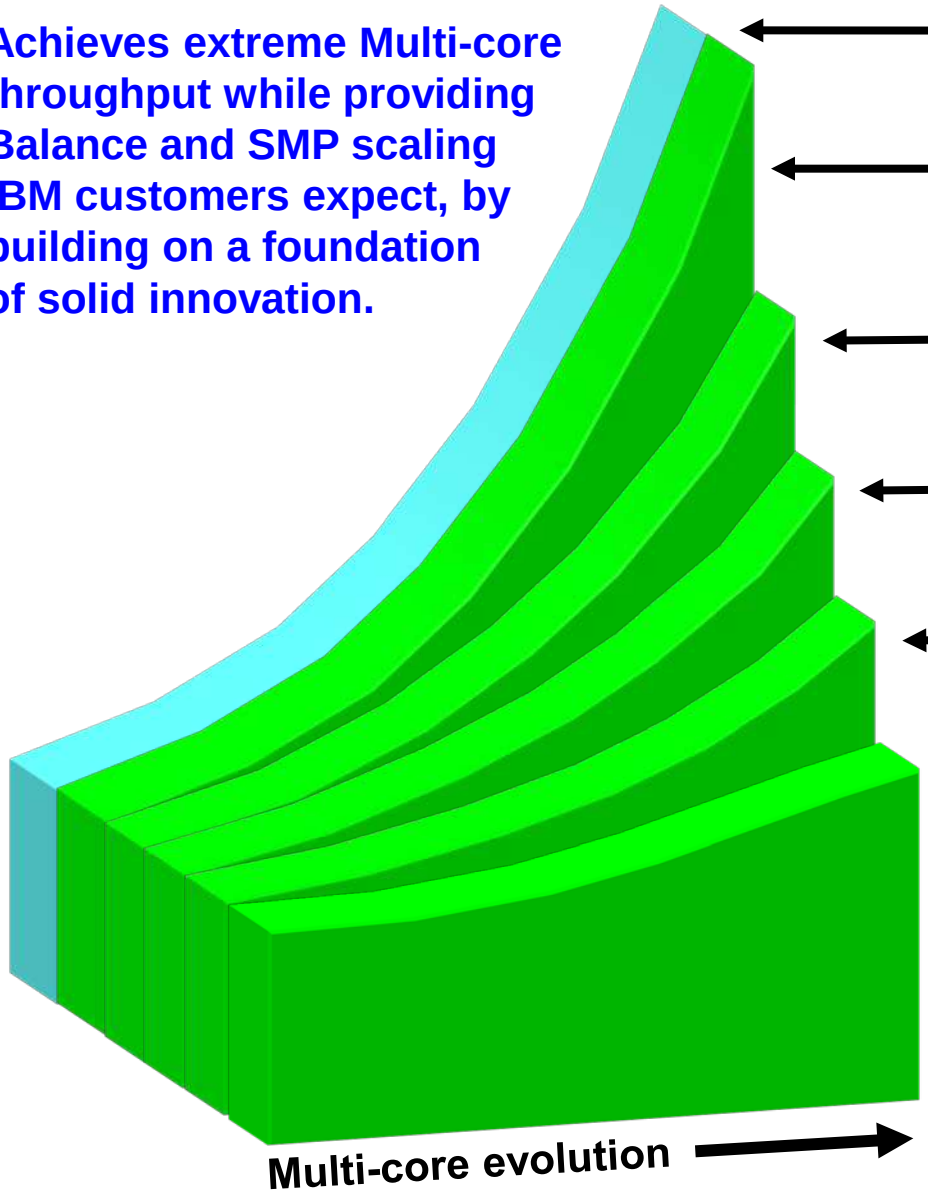
Necessity for POWER7

- 450 GB/s must grow to 1.6 TB/s to match POWER6 scaling
- 450 GB/s → 3.6 TB/s theoretical peak
- 3.6 TB/s → 14.4 TB/s with chip scope

* Statements regarding SMP servers do not imply that IBM will introduce a system with this capability.

Conclusion: POWER7 maintains the Balance

Achieves extreme Multi-core throughput while providing Balance and SMP scaling IBM customers expect, by building on a foundation of solid innovation.



Compute Throughput Potential

Exploit Long Term Investment in Coherence Innovation

Advances in Off-Chip Signaling Technology

Advances in Memory Subsystem

Cache Hierarchy Technology and Innovation

Socket Throughput Limitation (Physical signal economics)

IBM POWER chips uniquely positioned to excel given the emerging trends:

1) History of large SMP leadership
2) Storage Architecture economics
3) High density packaging leadership