DRAM scaling challenges and solutions in LPDDR4 context

Kishore Kasamsetty
Product Marketing Director, Cadence
MemCon 2014

cādence®

Agenda

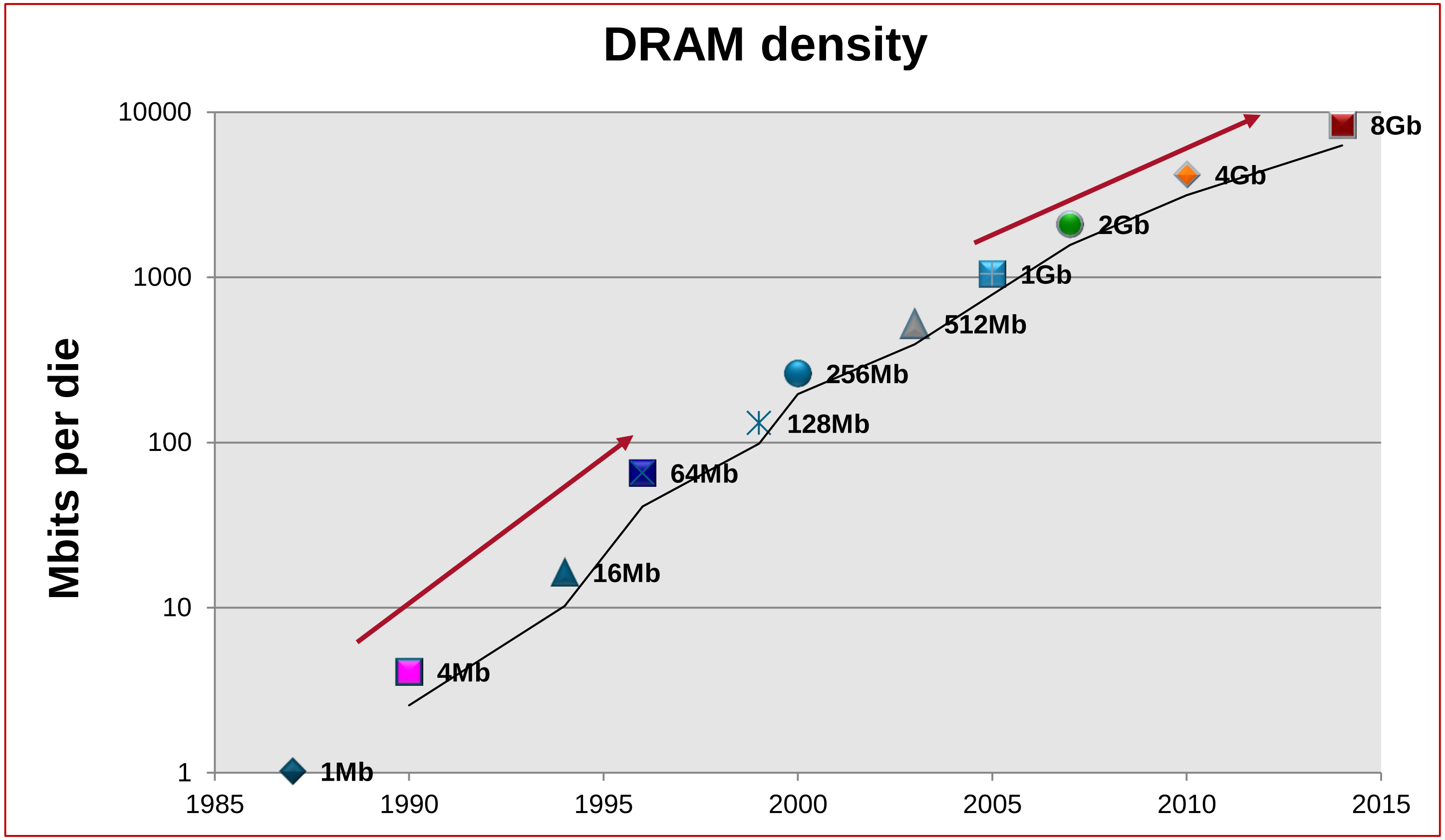
- Mobile DRAM drivers
- DRAM scaling trends/challenges
- LPDDR4 enhancements for density scaling
  - Non 2N devices
  - PPR (Post package Repair)
  - TRR (Target Row Refresh)
  - DDR IP implications
- LPDDR4 enhancements for bandwidth scaling
  - Multi command channels per die
  - DDR IP implications

Growth of Mobile Applications

Drivers for DRAM BW and capacity

- Relentless growth of mobile applications
- DRAM bandwidth drivers
  - Higher resolution displays (1080p/2K/4K), larger displays
  - Game console class gaming
  - Multi core processing
- DRAM capacity drivers
  - Sophisticated OS with larger footprint
  - Multi processing
  - Integrated radios/sensors in application processors
- DRAM solutions perennial challenges
  - Bandwidth, density per die, power
  - LPDDR4 is first generation that needs to innovate on all three metrics beyond what PC DRAM can deliver

DRAM density scaling history



- DRAM density growth flattening
  - 4x every 3 years not happening any more
- Maintaining storage capacitance at reduced feature size
  - Reliability challenges

DRAM bandwidth scaling history

- High yield DRAM column cycle (CAS frequency) has remained constant (200-250Mhz) over last 10 years
  - DRAM processes optimized for capacitance and not speed
- Higher bandwidth achieved by increasing prefetch size
  - Use fast IO for higher bit rates
  - Get more data from same address each cycle

Device	Pre-fetch/ Minimum access for x32 system	Typical Data rates
SDRAM	1 = 4bytes	200 Mbps
DDR1	2 = 8 bytes	400 Mbps
DDR2	4 = 16 bytes	800 Mbps
DDR3	8 = 32 Bytes	1600 Mbps
DDR4	16= 64 bytes ? ✗	3200 Mbps
LPDDR4	16 = 64 bytes ? ✗	3200 Mbps

Too big of access size. Effective useful bandwidth is low

cadence®

LPDDR4 offers highest density DRAM

- Mobile systems benefit from high density per die
  - Small form factor, BOM cost
  - LPDDR4 spec allows higher densities than DDR4
- DRAM yields challenging at higher densities
  - Non 2N density devices introduced in LPDDR4

Memory Density (per Die)	4Gb	6Gb	8Gb	12Gb	16Gb	24Gb	32Gb
Memory Density (per channel)	2Gb	3Gb	4Gb	6Gb	8Gb	12Gb	16Gb
Configuration	16Mb x 16DQ x 8 banks x 2 channels	24Mb x 16DQ x 8 banks x 2 channels	32Mb x 16DQ x 8 banks x 2 channels	48Mb x 16DQ x 8 banks x 2 channels	64Mb x 16DQ x 8 banks x 2 channels	TBD x 16DQ x TBD banks x 2 channels	TBD x 16DQ x TBD banks x 2 channels
Number of Rows (per channel)	16,384	24,576	32,768	49,152	65,536	TBD	TBD

cadence®

LPDDR4 offers highest density DRAM

- Mobile systems benefit from high density per die
  - Small form factor, BOM cost
  - LPDDR4 spec allows higher densities than DDR4
- DRAM yields challenging at higher densities
  - Non 2N density devices introduced in LPDDR4

Memory Density (per Die)	4Gb	6Gb	8Gb	12Gb	16Gb	24Gb	32Gb
Memory Density (per channel)	2Gb	3Gb	4Gb	6Gb	8Gb	12Gb	16Gb
Configuration	16Mb x 16DQ x 8 banks x 2 channels	24Mb x 16DQ x 8 banks x 2 channels	32Mb x 16DQ x 8 banks x 2 channels	48Mb x 16DQ x 8 banks x 2 channels	64Mb x 16DQ x 8 banks x 2 channels	TBD x 16DQ x TBD banks x 2 channels	TBD x 16DQ x TBD banks x 2 channels
Number of Rows (per channel)	16,384	24,576	32,768	49,152	65,536	TBD	TBD

cadence

Non-2N density devices implications

- Transparent to CPU/ host system, see total address space
- DRAM controllers today
  - Support flexible address mapping schemes (Bank/Row/Rank addresses)
  - DRAM controllers typically look at single bit to determine page/bank/rank changes/rollovers
- DRAM controllers with Non 2N devices
  - Need multi bit address compare to determine (Page/Rank/Device)
  - Should not impact performance/throughput with correct implementation

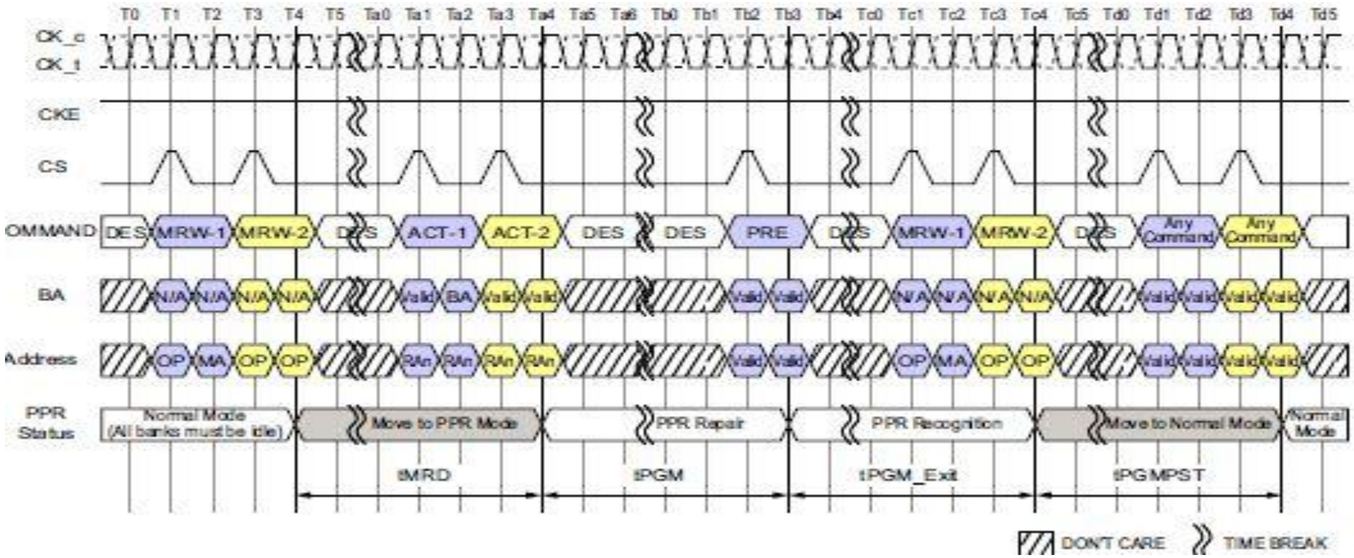
Memory Density (per Die)	4Gb	6Gb	8Gb	12Gb	16Gb	24Gb	32Gb
Memory Density (per channel)	2Gb	3Gb	4Gb	6Gb	8Gb	12Gb	16Gb
Number of Rows (per channel)	16,384	24,576	32,768	49,152	65,536	TBD	TBD

LPDDR4 introduces Post package repair (PPR)

- Higher density DRAM susceptible to increased single row failures
- DRAM devices historically have row redundancy circuits to address these
  - Improve yields at die sort, uses “efuse” technology
  - Bad rows remapped to built in redundant rows
  - Not exposed to host system
- LPDDR4 standard includes PPR
  - Repair scheme accessible to controller

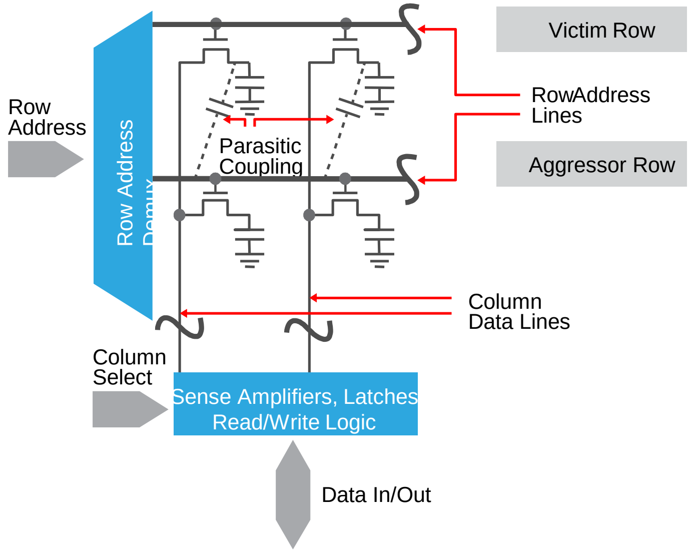
cadence®

LPDDR4 post package repair



- Simple command control repair protocol defined in LPDDR4 (~1000ms)
- <u>Applications</u>
  - *Multi die assembly*:   Do BIST check and repair failing rows
  - *System initialization:* MC can do BIST check and repair failing rows
  - *Field failures*: Need software tracking to accumulate ECC failures and determine failing rows
- Memory controllers should check for unintended PPR entry possibilities
  - Certified memory models (VIP) can check and flag these

Row Hammering / Target Row Refresh



"Row Hammering"
Frequently accessed rows (target rows) disturbs adjacent rows (victim)

- LPDDR4 DRAM requires controllers to do repair using Target row refresh mode (TRR) when a threshold of "victim" hits happen on adjacent rows
- Very expensive to track the activity for thousands and rows
- Statistical approaches and prior application knowledge may yield practical solutions[1]

1. ISCA 2014 "Flipping Bits in Memory without accessing them" Intel Labs and CMU

cadence®

DRAM bandwidth scaling history

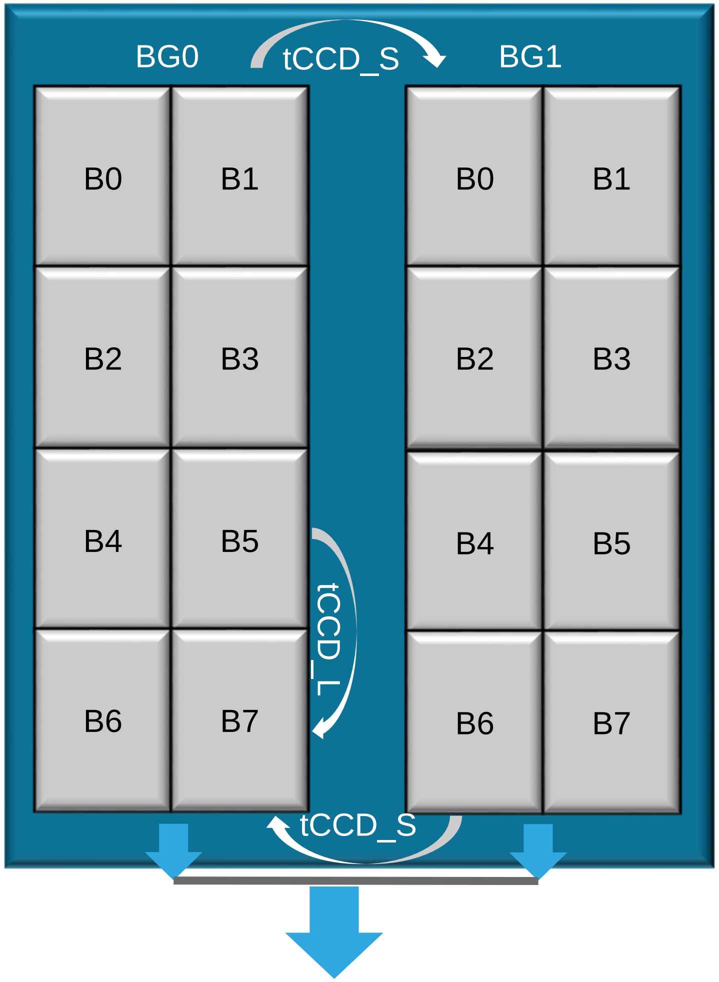
- High yield DRAM column cycle (CAS frequency) has remained constant (200-250Mhz) over last 10 years
  - $/bit reduction drives DRAM economics
- Higher bandwidth achieved by increasing prefetch size
  - Get more data each cycle and use fast IO to increase bandwidth

Device	Pre-fetch/ Minimum access for x32 system	Typical Data rates
SDRAM	1 = 4bytes	200 Mbps
DDR1	2 = 8 bytes	400 Mbps
DDR2	4 = 16 bytes	800 Mbps
DDR3	8 = 32 Bytes	1600 Mbps
DDR4	16= 64 bytes ? ✗	3200 Mbps
LPDDR4	16 = 64 bytes ? ✗	3200 Mbps

Too big of access size. Effective useful bandwidth is low

cadence®

DRAM bandwidth scaling : DDR4 solution

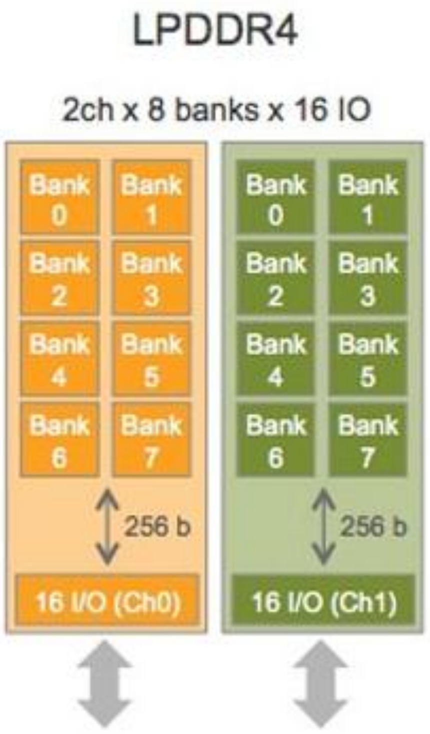


- Two Bank Groups
  - tCCD_L is longer than tCCD_S
  - Access size stays ***32*** bytes
  - Full bandwidth needs ping-pong access
- Continuous access to a single bank group maxes at 66% utilization

cādence®

DRAM bandwidth scaling : LPDDR4 solution

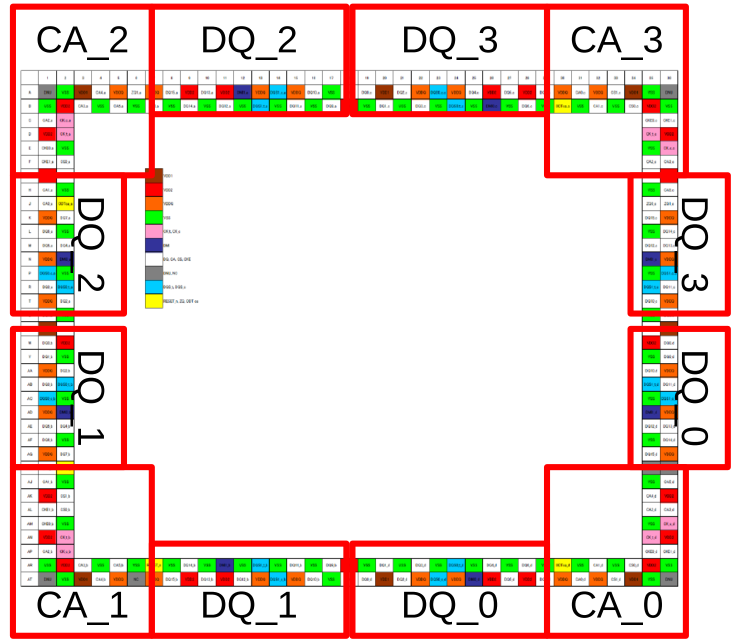
- Two command channels
  - 32 bit system will have 2 command channels
  - Minimum access size stays 32 Bytes
- Independent control allows better utilization for localized data
- Independent control allows for additional powerdown flexibility
- Down side
  - Potential for more pins (6 pin command helps)
  - Complicated PCB/PKG routing for dual mode memory systems



cadence

POP packaging differences in LPDDR4/3

LPDDR3 64 bit
CA & DQ on opposite side

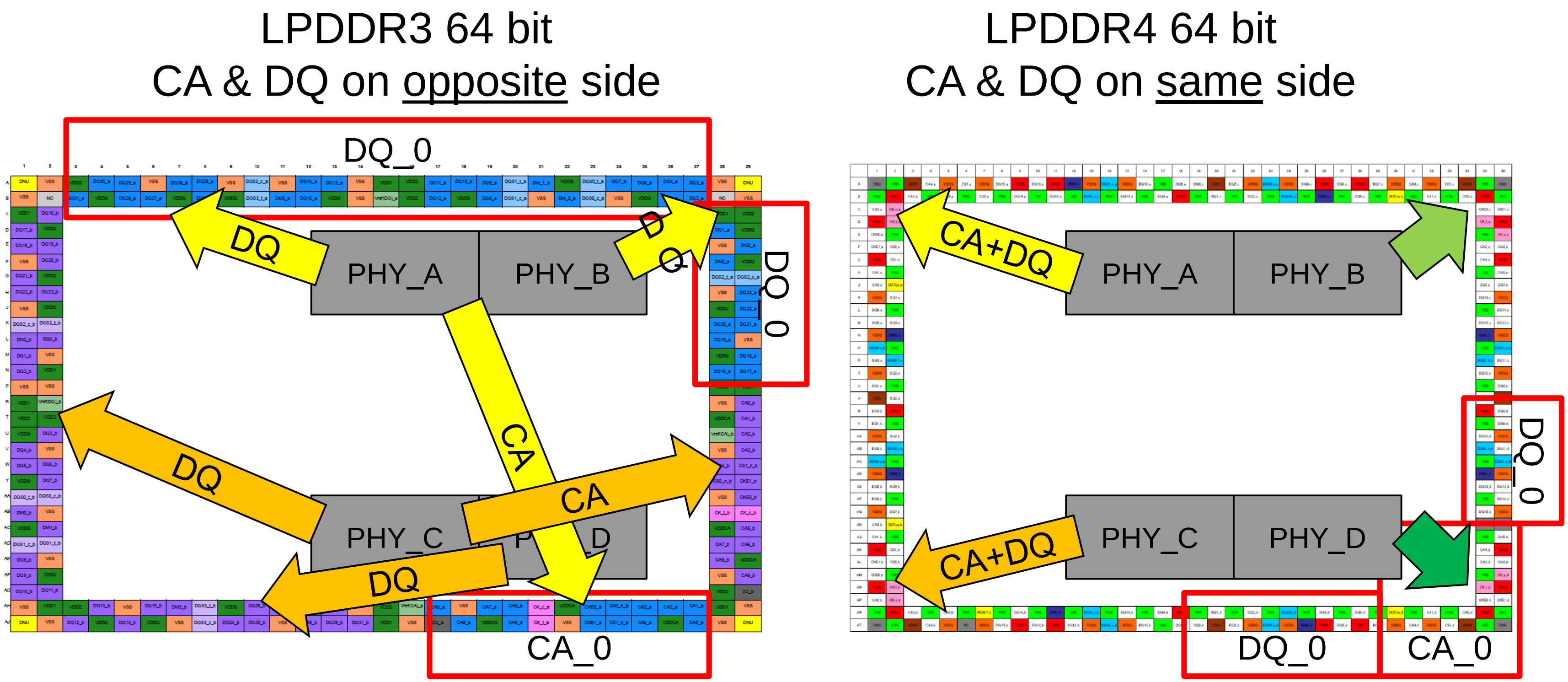


LPDDR4 64 bit
CA & DQ on same side

- Increased number of channels force changes to ballout
- Difficulties for doing dual mode channel systems

cadence®

LPDDR4 focused SoC PKG design



- LPDDR4 optimized placement can work for LPDDR3
- Still need long routes in package and Soc for LPDDR3
- PHY/Controller flexibility is needed to make it work

cādence®

Controller and PHY IP Techniques to ease PCB and Package routing

- DRAM Controller and PHY IP may employ techniques to ease the burden and provide package/PCB routing flexibility for multi-mode
  - Per bit deskew on CA bus
  - CA bit swapping
  - DQ bit swapping
  - Dual-mode (SDR and DDR) support for CA

cadence®

Summary

- LPDDR4 added PPR, TRR, non 2N density devices to meet the high per die density requirements
- LPDDR4 introduces dual channel systems to scale and meet bandwidth requirements
- Cadence offers Controller, PHY and VIP solutions need to optimally and reliably work with LPDDR4 based systems

cadence®