The future of graphic and mobile memory for new applications

August 21st, 2016 I JIN KIM I Samsung Electronics

Disclaimer

This presentation is intended to provide information concerning memory industry. We do our best to make sure that information presented is accurate and fully up-to-date. However, the presentation may be subject to technical inaccuracies, information that is not up-to-date or typographical errors. As a consequence, Samsung does not in any way guarantee the accuracy or completeness of information provided on this presentation. Samsung reserves the right to make improvements, corrections and/or changes to this presentation at any time.

The information in this presentation or accompanying oral statements may include forward-looking statements. These forward-looking statements include all matters that are not historical facts, statements regarding the Samsung Electronics' intentions, beliefs or current expectations concerning, among other things, market prospects, growth, strategies, and the industry in which Samsung operates. By their nature, forward-looking statements involve risks and uncertainties, because they relate to events and depend on circumstances that may or may not occur in the future. Samsung cautions you that forward looking statements are not guarantees of future performance and that the actual developments of Samsung, the market, or industry in which Samsung operates may differ materially from those made or suggested by the forward-looking statements contained in this presentation or in the accompanying oral statements. In addition, even if the information contained herein or the oral statements are shown to be accurate, those developments may not be indicative developments in future periods.

Contents

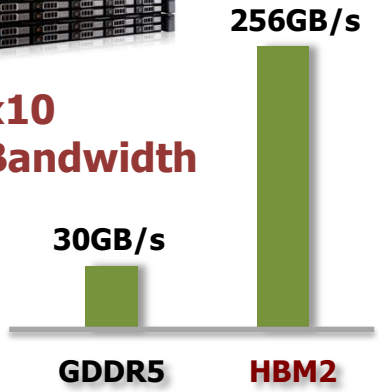
- Memory technology trend
- High speed graphic technology (>10Gbps)
- Low power mobile technology (>20%)
- Conclusion

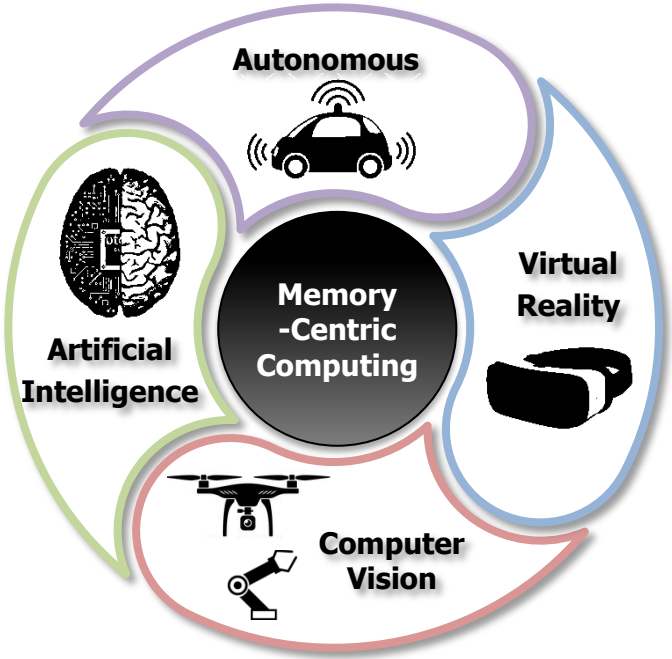
Memory technology trend

Memory is at the core of new applications

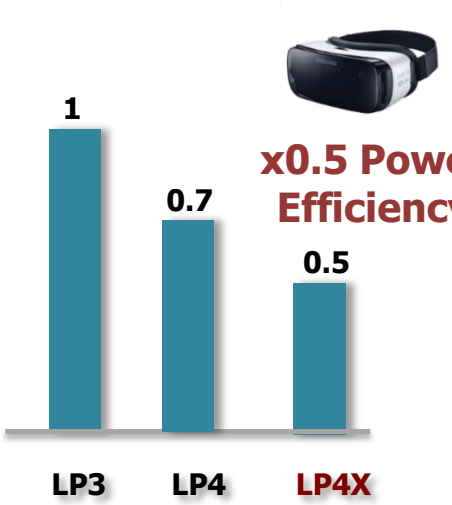
Higher Performance

x10 Bandwidth





Lower Power

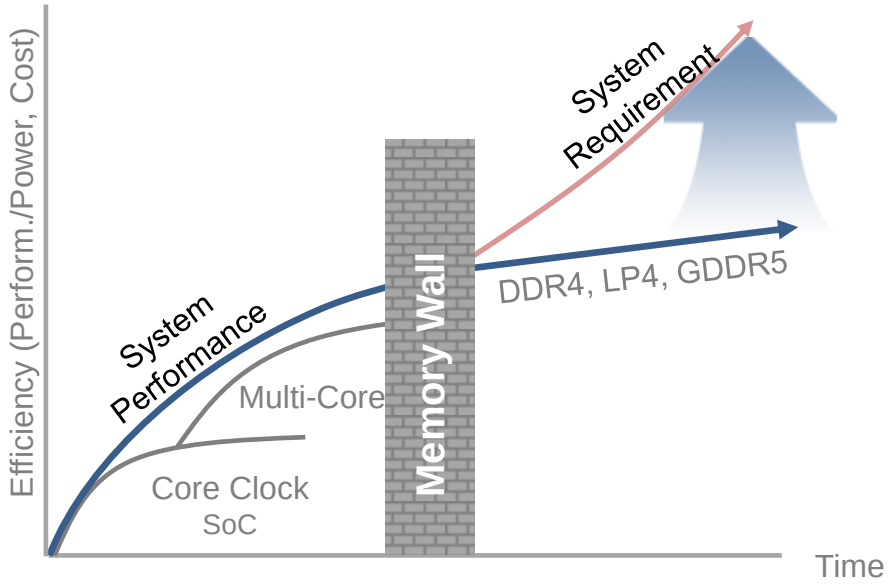


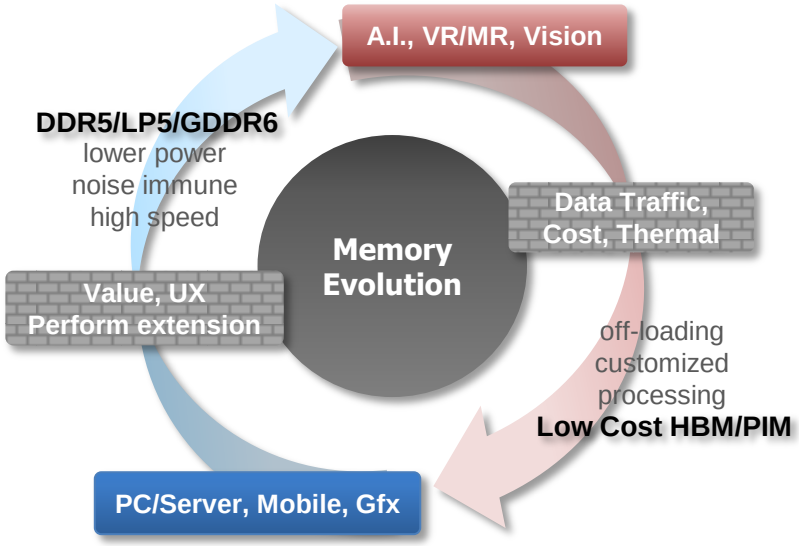
x0.5 Power Efficiency

Source: Samsung

Memory-centric system evolution

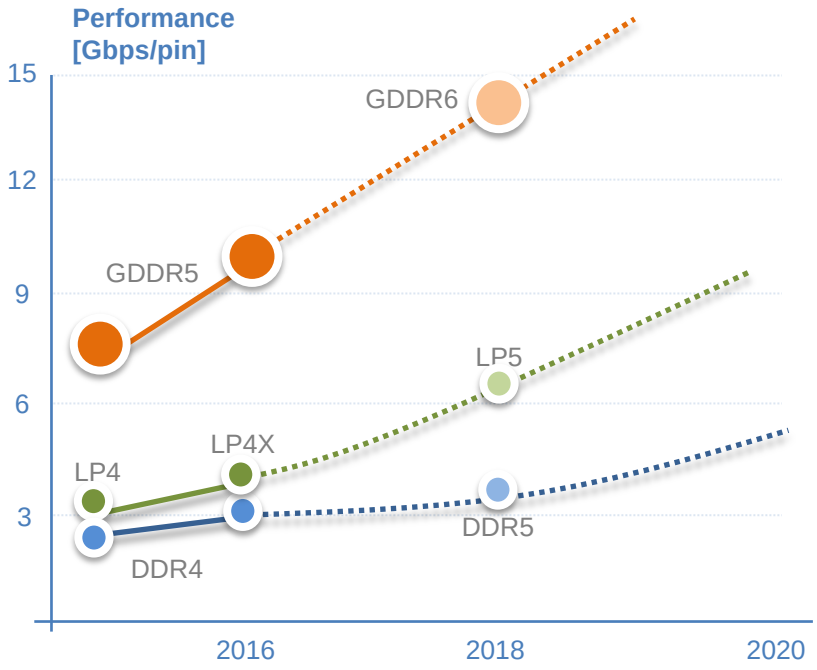
- **Extreme B/W, performance/power, data processing, cost effective solutions**

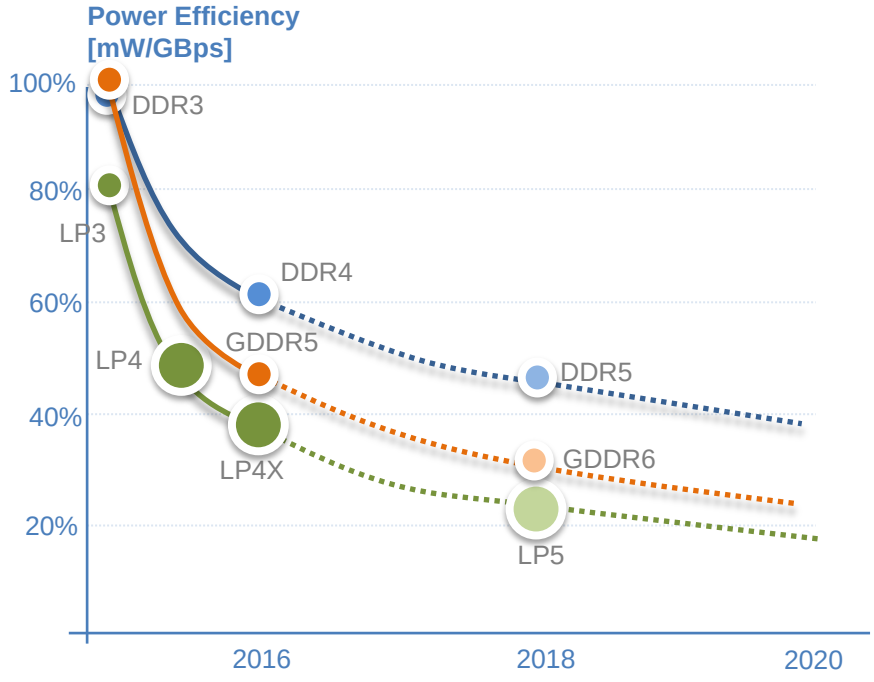




Memory technology trend

- GDDR6 with over 14Gbps, beyond 10Gbps GDDR5
- LP5, 20% more power-efficient than LP4X

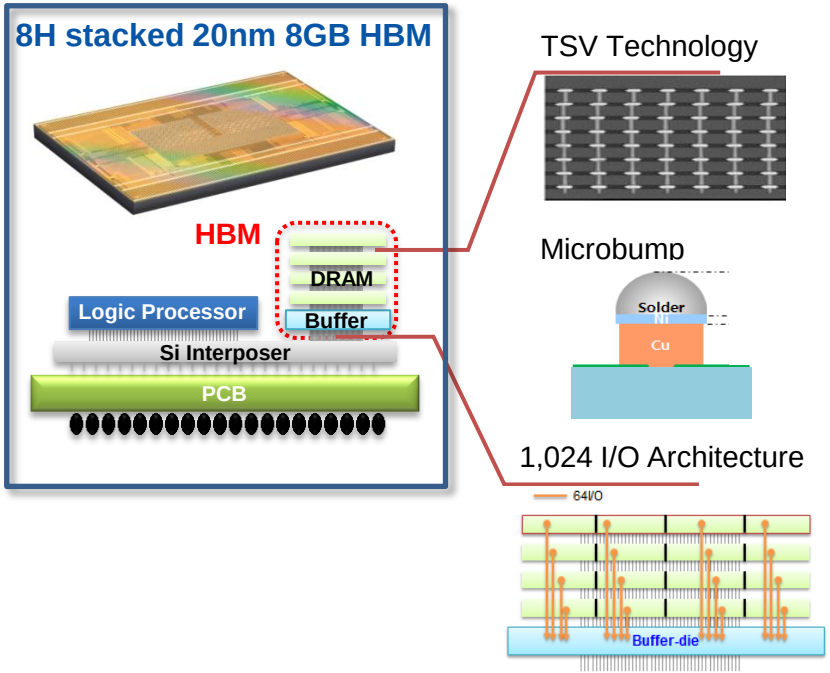




Source: ISCA2016, Samsung

High Bandwidth Memory: HBM

1TB/s **High Bandwidth**



Benefits

Performance

HBM — X 2.7

GDDR5

Power Efficiency

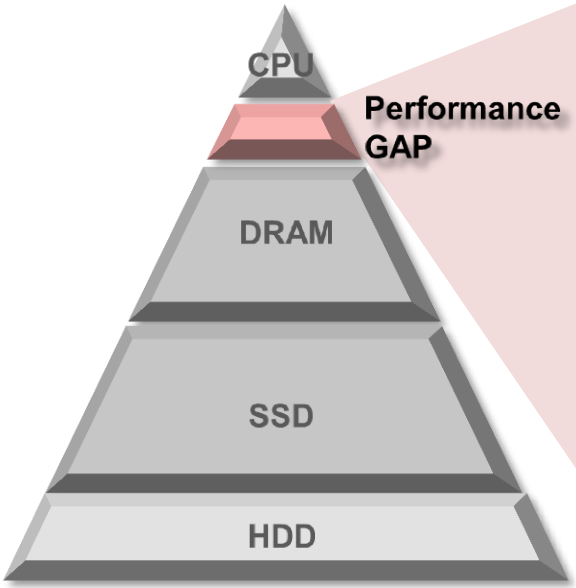
HBM — X 0.8

GDDR5

Source: Samsung

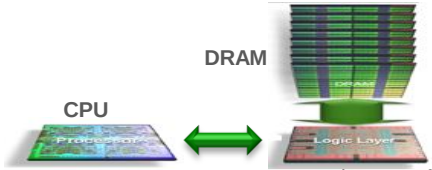
Processing In Memory: PIM

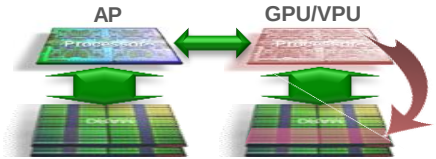
- Fill the performance gap and deliver energy-efficient solutions



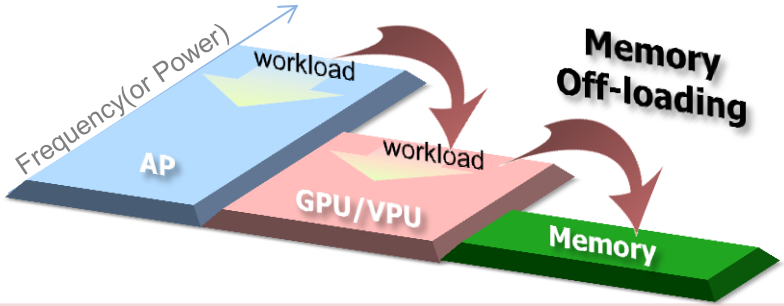
Processing In-Memory

Better parallelism and lower bus traffic





Memory off-loading for lower frequency and power



Source: Samsung
SAMSUNG

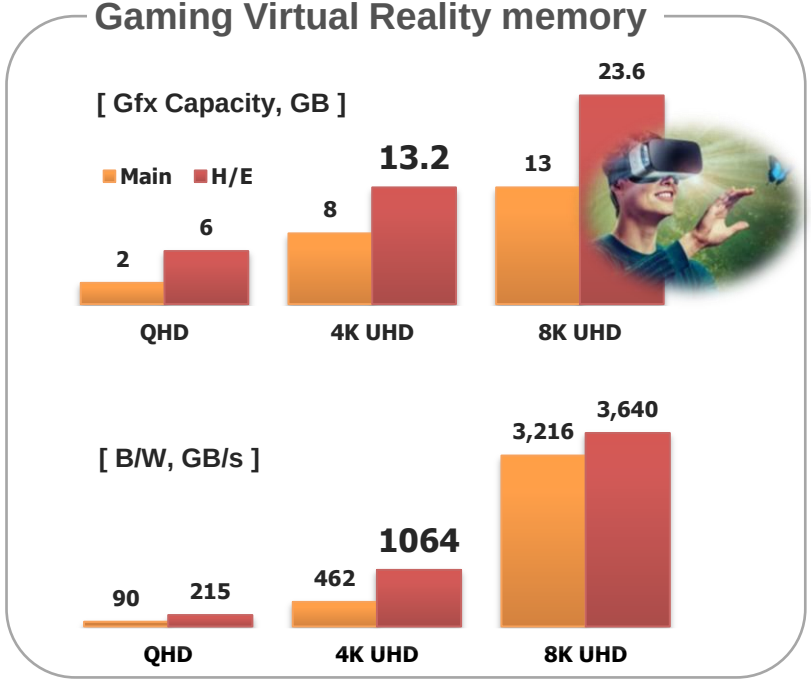
High speed graphic technology (>10Gbps)

- Graphic application requirement
- Asymmetric System, Crosstalk, EQ tuning
- GDDR6, Low cost HBM, PIM

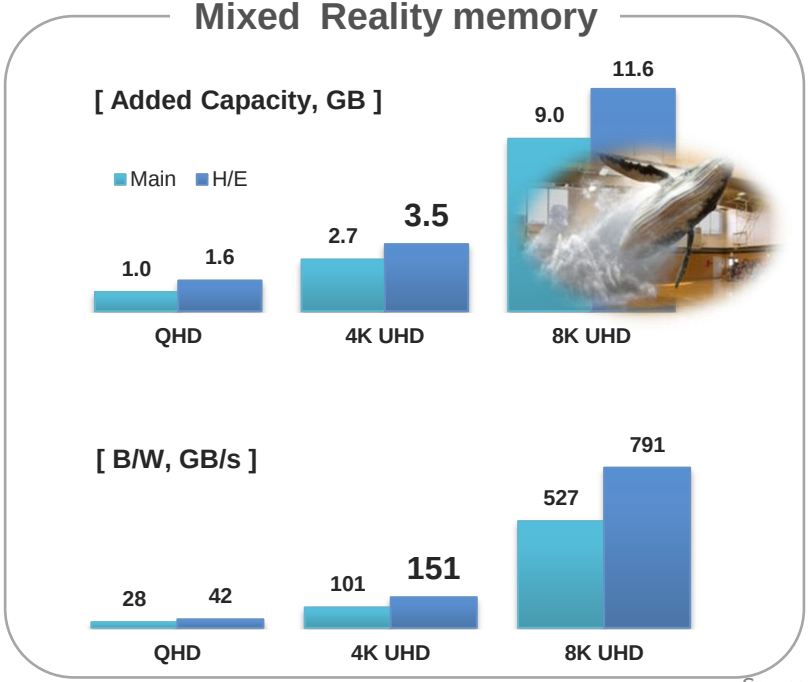
High speed memory requirement

- For 4K real infographic virtual reality, 13.2GB, 1TB/s memory needed
- For 4K 3D mixed reality, +3.5GB, 151GB/s memory needed

Gaming Virtual Reality memory



Mixed Reality memory



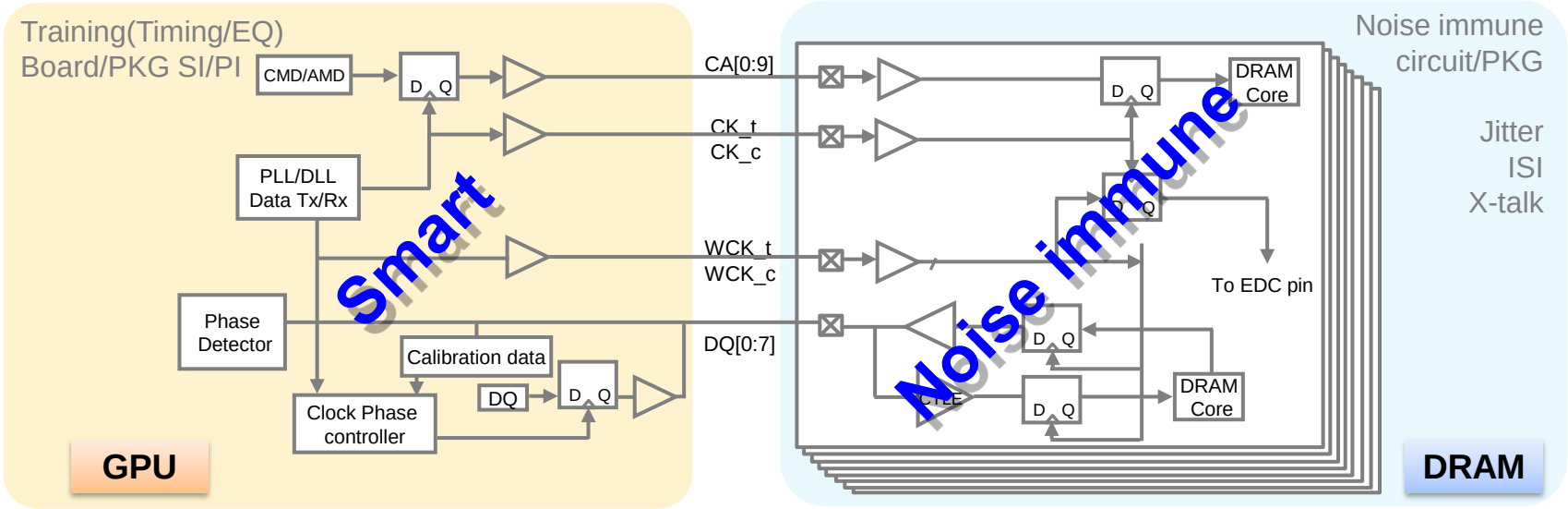
Source: Samsung

Variable Assumption

Poly count, fps, # of texture per fragment, cache hit rate, tri-linear filtered, # of virtual light source, Reflection/refraction ratio, ray bounce depth

Asymmetric system for higher data rate

- **Focus on the respectively dedicated features to maximize data rate**
  - Smart GPU : Training (Per-bit Timing/EQ) for minimizing static offset/noise
  - Noise immune DRAM : minimizing dynamic noise (Jitter, ISI/x-talk, clock duty/skew)

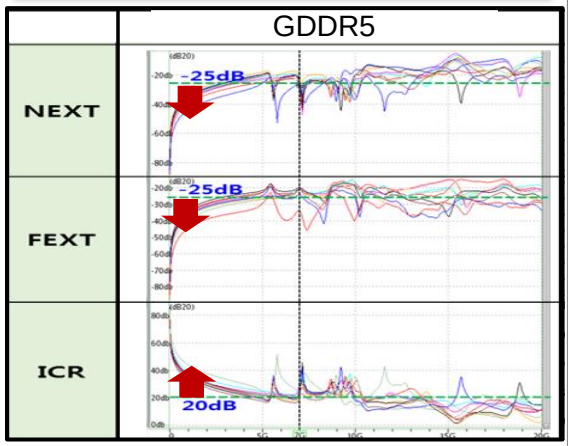


Source: Samsung

X-talk reduction for Board/PKG design

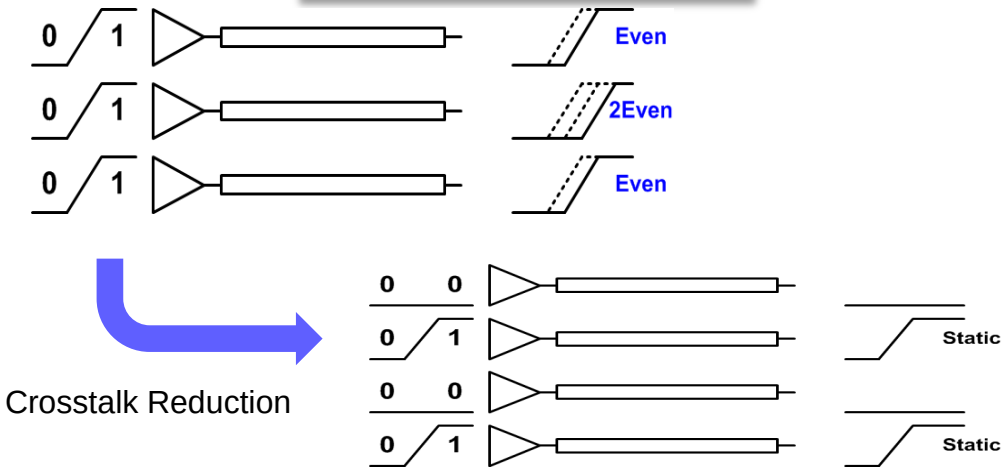
- Small X-talk Package : reduction of X-talk with better return path
- Crosstalk Reduction with coding : 3B4B, 8B9B

Small X-talk PKG requirement



ICR: Insertion loss to Crosstalk Ratio

3B4B encoding

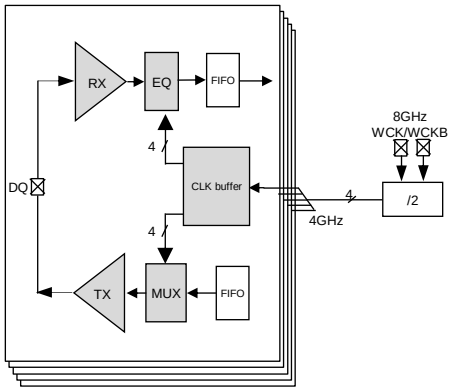


Source: Samsung

DFE for return-loss reduction on system

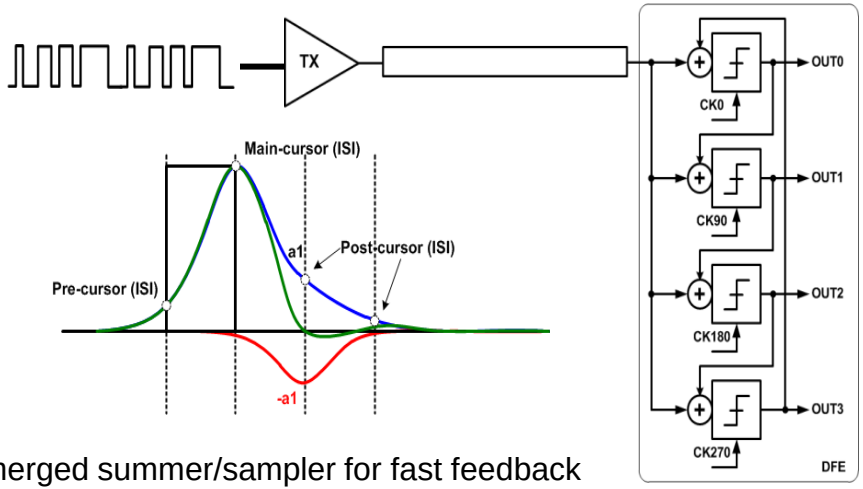
- **Single ended signaling requires noise immune equalizer**
  - DFE* is more suitable than CTLE**

CTLE & DFE



CTLE and DFE
Periodically Calibrated by GPU

Quarter rate DFE with summer in sampler



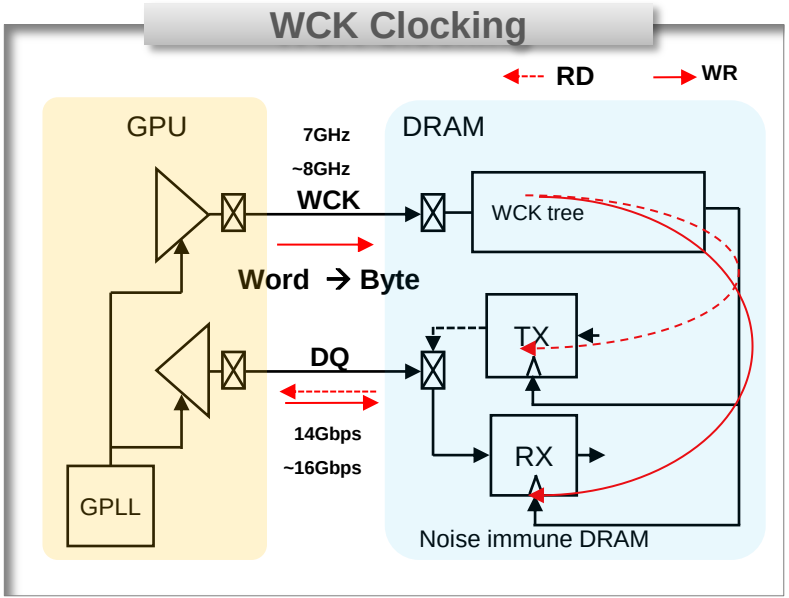
Adopt merged summer/sampler for fast feedback

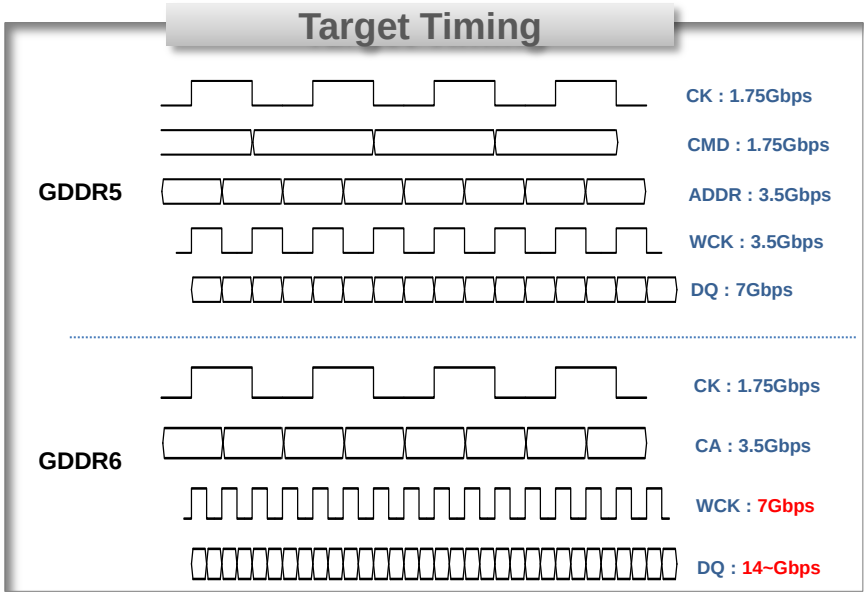
* Decision Feedback Equalization
** Continuous Time Linear Equalization

Source: Samsung

GDDR6 ideas

- **High Speed Signaling, 14Gbps ~ 16Gbps, 1.35V**
  - Low jitter clocking with WCK/byte, Per-bit RX/TX equalizer training, X-talk reduction
  - 2 channel with BL16, same Clock/ADD freq., twice of WCK/DQ freq.



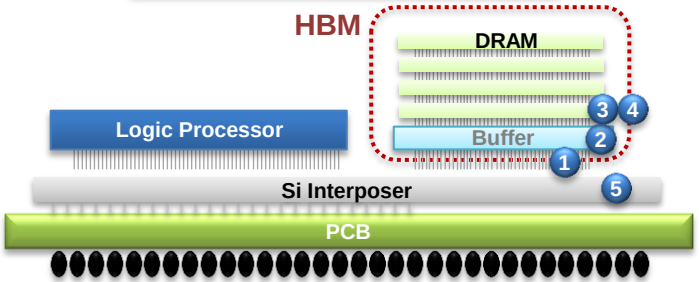


Source: Samsung

Low cost HBM for consumer segment

- **~ 200GBps with smaller # of TSV compared to HBM2**
  - Cost competitiveness ; remove buffer die, reduce # of TSV, organic interposer, etc..
  - Need inputs from Client segment for specific features

Challenge for HBM



Challenges

1. IO reduction, Smaller # of TSV
2. Remove buffer die
3. Master/Slave structure
4. Remove ECC
5. Si or organic Interposer

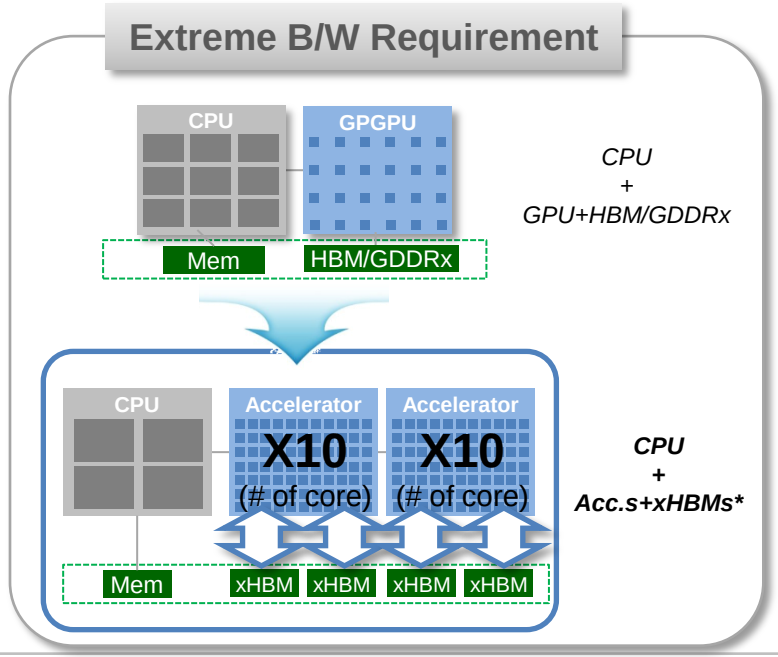
Comparison

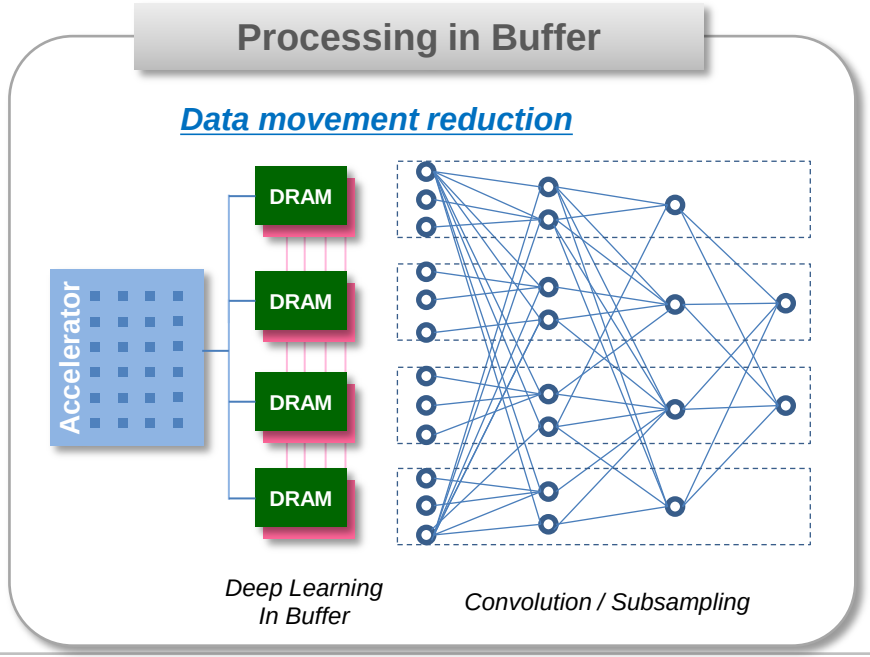
	HBM2	Low cost HBM
I/O	1024	~512
Pin speed	2Gbps	3Gbps ~
BW (GB/s)	256	~ 200
Cost/GB	1	**0.X**

Source: Samsung

PIM, Deep Learning in DRAM

- **Parallel processing in buffer to reduce extreme-bandwidth**
  - convolution, subsampling, matrix calculation
- **Collaborate with accelerator for performance/cost**





* xHBM: Extreme HBM

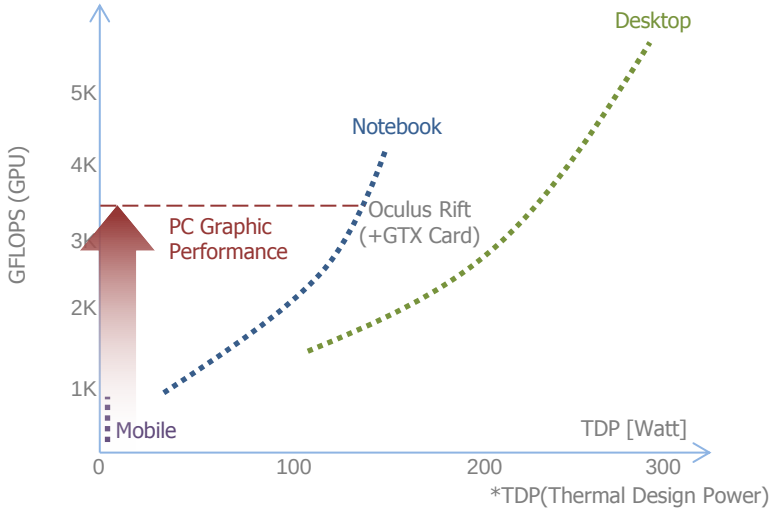
Low power mobile technology (>20%)

- Motivation for low power mobile
- LP4X / LP5
- PIM

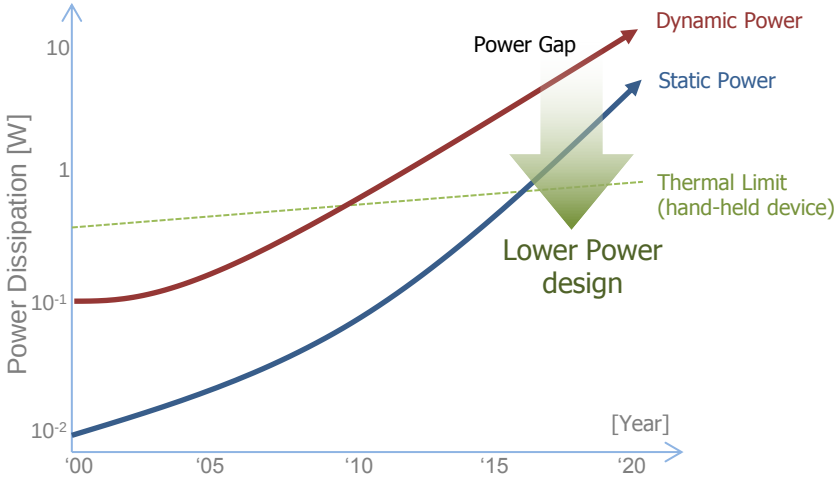
Motivation for low power mobile

- PC-level graphic performance and mobile power budget
- Power is continuously increasing with limited thermal budget

Performance vs. TDP



Power Dissipation Trend

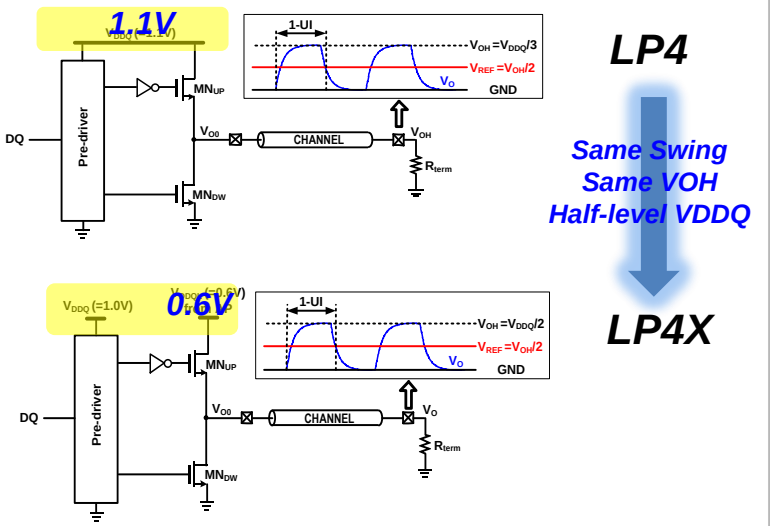


Source: Samsung

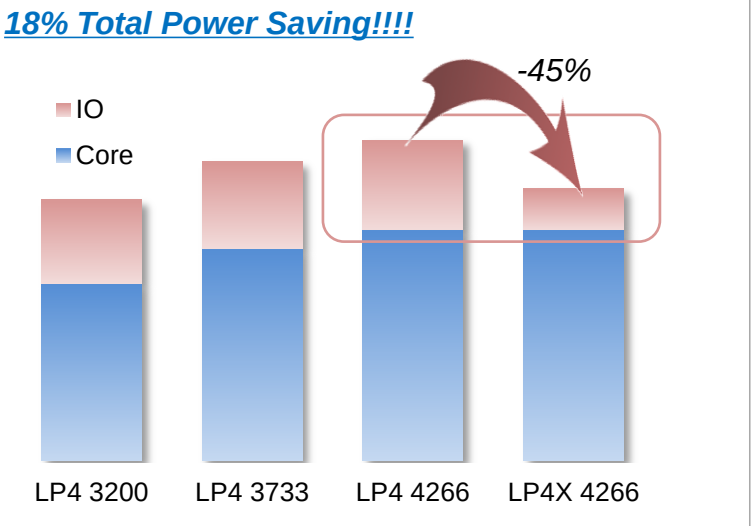
Lower power solution, LP4X

- **LP4X : 4266Mbps, VDDQ/VDD = 0.6V/1.1V**
  - IO power reduction with 0.6V VDDQ, Good example of small change but big gain

LP4X Idea



LP4X Power Reduction



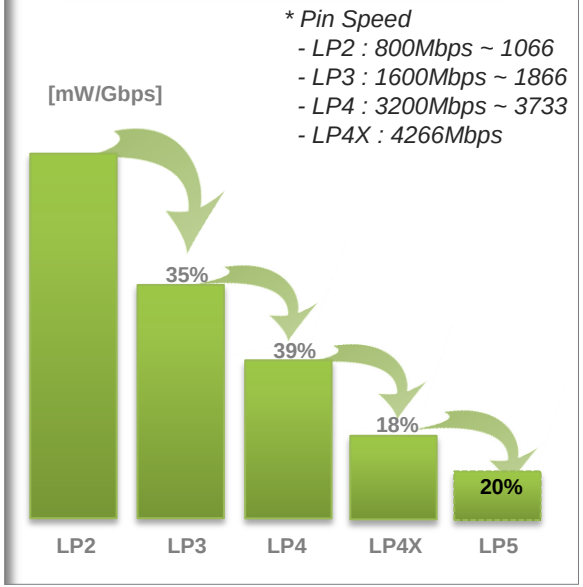
•Conditions : IDD4R(VDDQ+VDD2) Spec Value / 50% Data change each burst transfer / Included process node contribution

Source: Samsung

LP5 target & ideas

- **LP5 : 6400Mbps, VDDQ/VDD < 0.6V/1.1V**
  - Extremely high band-width(~6.4Gbps) and smart power reduction(~20%)

Power Efficiency Trend



LP5 ideas

CMD Based Data CLK(WCK)

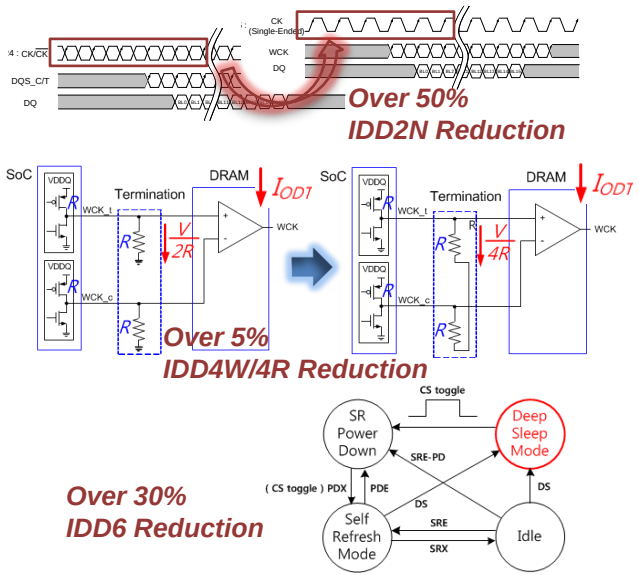
- IDD2N reduction

WCK Center-tap term

- IDD4W/R reduction

Deep Sleep Mode

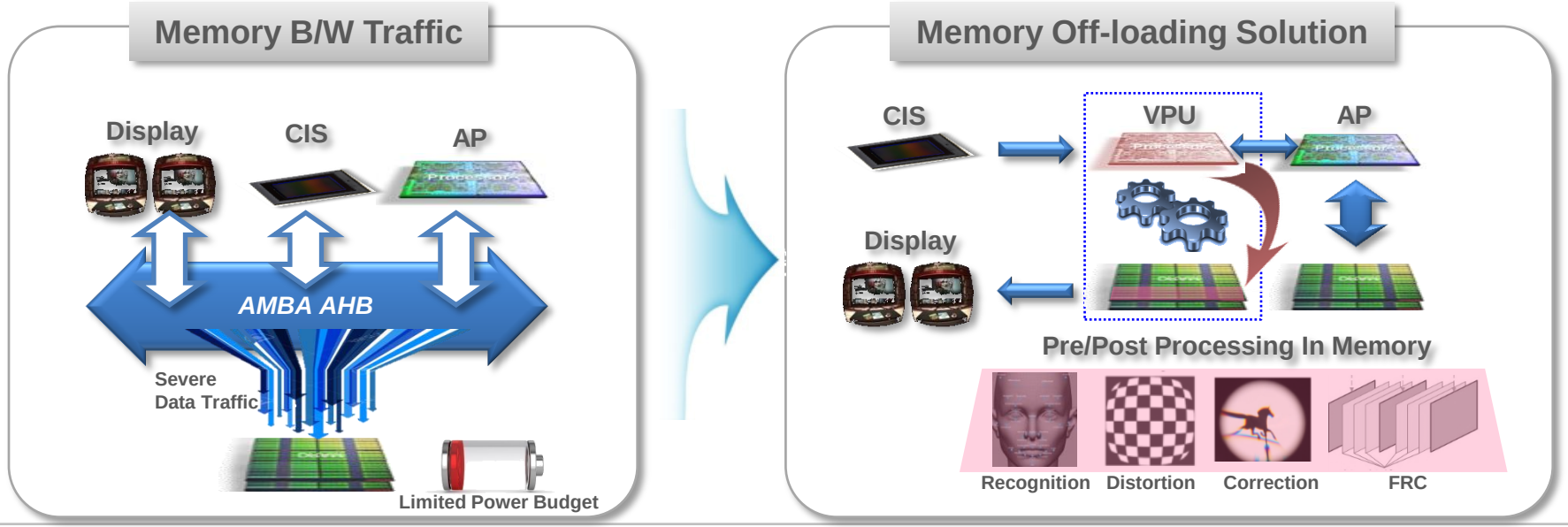
- IDD6 reduction



Source: Samsung

PIM, Lower power processing

- **Memory off-loading for reduced power consumption**
  - Reduce the unnecessary data transfer and frame rate control
- **Collaborate with SoC/AP for performance/power**
  - PoC with special memory for post/pre-processing



Conclusion

Conclusion

- **Memory requirements have become more strict in time with respect to performance, power, and cost**

- **Keeps innovating technology to correspond to those requirements**
  - Make efforts to extend the value of traditional memory
  - Figure out innovative memory solution

- **Close collaboration with partners is essential for delivering the right memory solution.**

kjh5555@samsung.com

Thank You!