University of Chinese Academy of Sciences

Impact of Big Data on SoC Design

Memory Design and HBM

Chun-Zhang Chen, Ph.D.

June 25-29, 2018

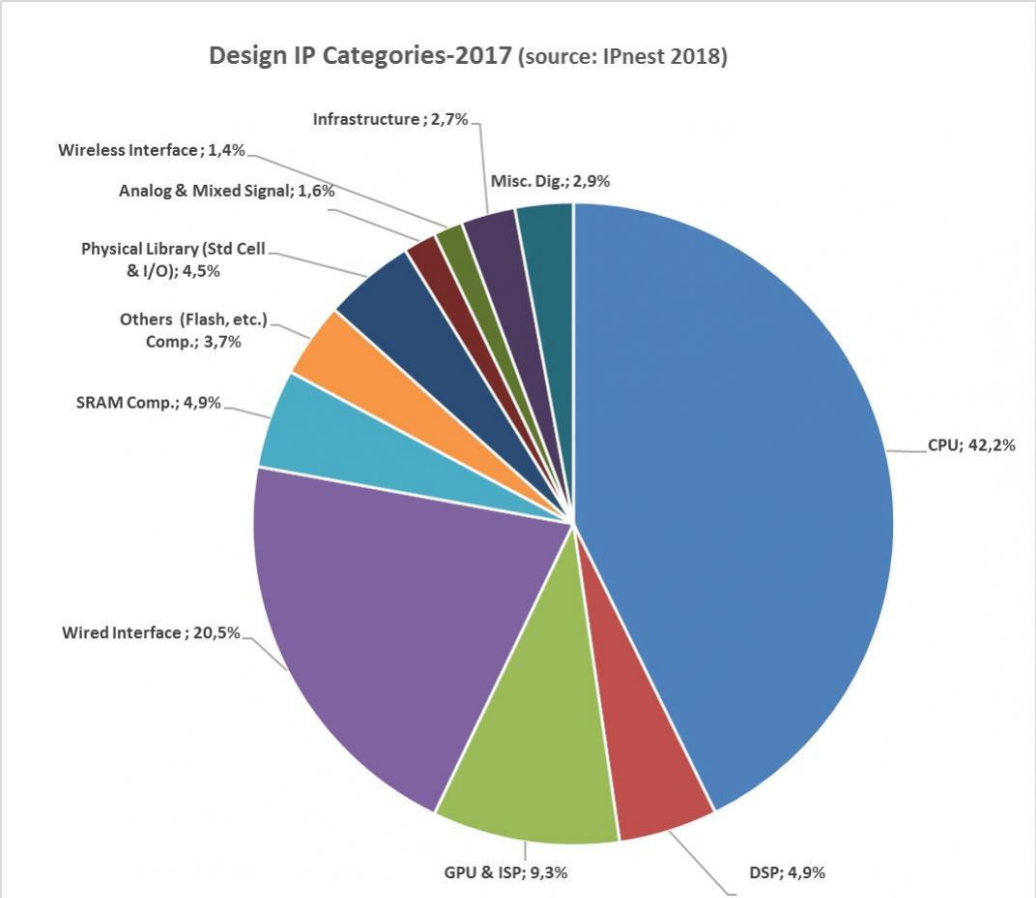
中国科学院大学2018年夏季

Memory Design and HBM

University of Chinese Academy of Sciences

- **Memory Designs**
- DDR/mDDR and GDDR
- RAM Cells
- HBM via 2.5D IC and 3D IC
- Discussion

Market of Memory Design



2018 Memory Design Grew

Prices of six generations of DRAMs

- DRAM 16kbit → 64Mbit (1978-2001)
  - $1/1977 ~ $2.95/2001
- DRAM 1Mbit

Amazingly

  - >$5000/1977 → $0.35/2000 → $0.08/2001 (≈1977)

- DRAM → NAND Flash

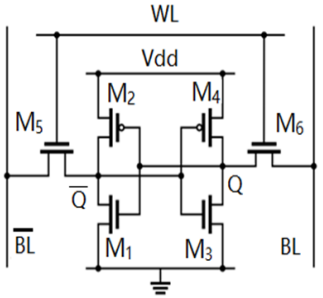
Memory Design Types

- RAM, ROM
- SRAM, DRAM
  - DRAM: SDRAM/DDR…
- OTP/MTP
  - MTP: PROM, EPROM, EEPROM (E^2PROM)
- Flash (NVRAM): NAND, NOR
  - 3D NAND, Xpoint
- SSD (NAND Flash)

SRAM, Static Random-Access Memory

中国科学院大学 University of Chinese Academy of Sciences

SRAM 6T-Cell

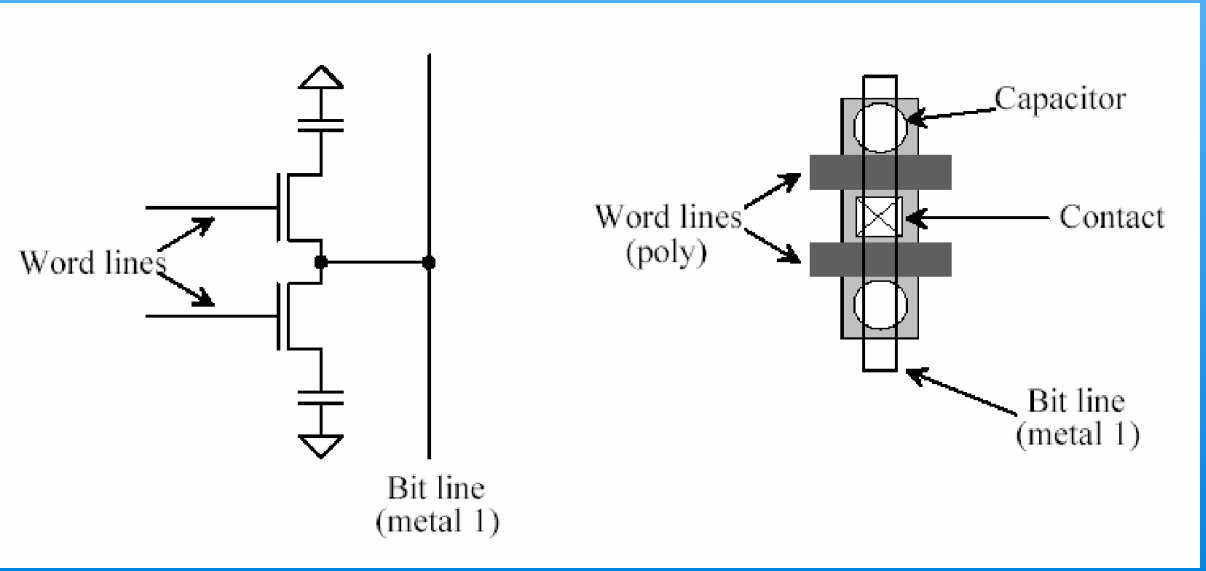


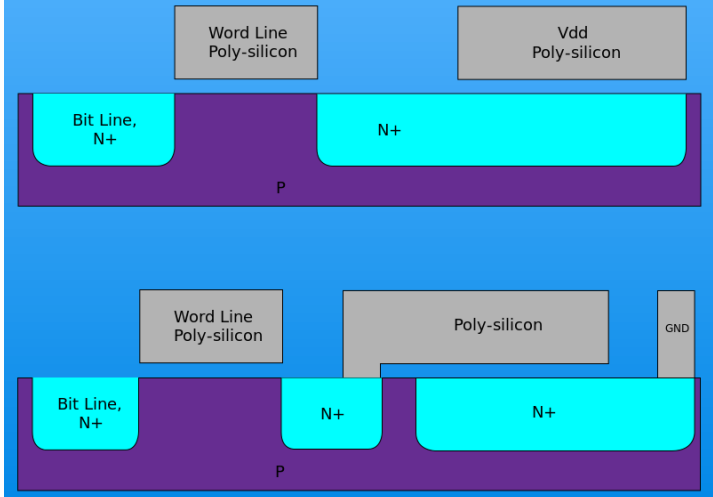
1T-SRAM Cell sizes (μm²/bit or mm²/Mbit)

Process node		250 nm	180 nm	130 nm	90 nm	65 nm	45 nm
6T-SRAM	bit cell	7.56	4.65	2.43	1.36	0.71	0.34
	with overhead	11.28	7.18	3.73	2.09	1.09	0.52
1T-SRAM	bit cell	3.51	1.97	1.10	0.61	0.32	0.15
	with overhead	7.0	3.6	1.9	1.1	0.57	0.28
1T-SRAM-Q	bit cell			0.50	0.28	0.15	0.07
	with overhead			1.05	0.55	0.29	0.14

SRAM is faster and more expensive than DRAM; it is typically used for CPU cache while DRAM is used for a computer's main memory.

DRAM, Dynamic Random-Access Memory





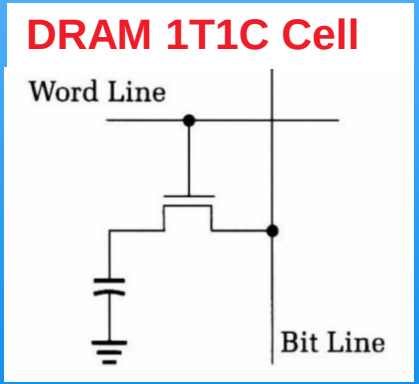
DRAM, SDRAM (synchronous) & ADRAM

Difference by the read style
  Sync. DRAM (SDRAM), Async. DRAM

Difference by the datarate, SDRAM has
  Single Data Rate (SDR)
  Double Data Rate (DDR)
  Quad Data Rate (QDR)

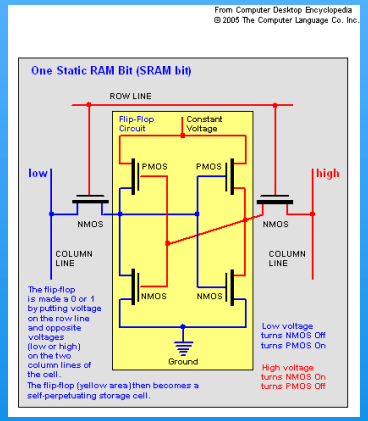
From applications Graphics (GDDR)
From power consumption,
  DDR, LPDDR (mDDR)

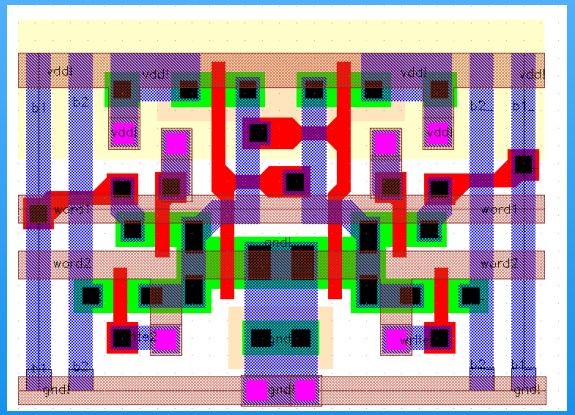
When DRAM is integrated with other modules (ex. CPU), it is called
  embedded DRAM (eDRAM)

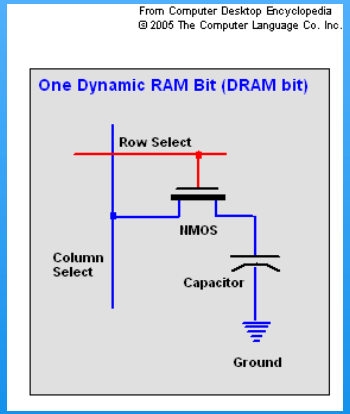


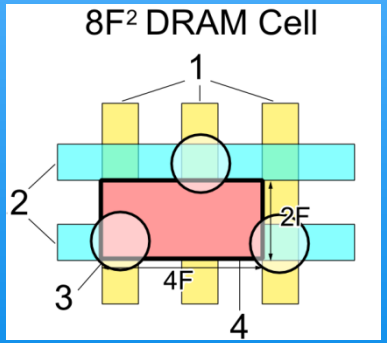
Layouts of SRAM and DRAM

- SRAM – 6T
- DRAM – 1T1C



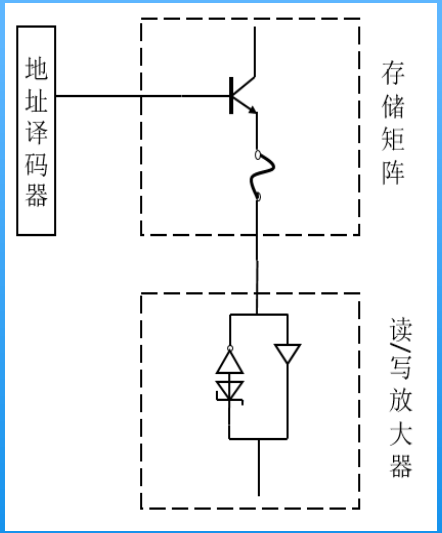


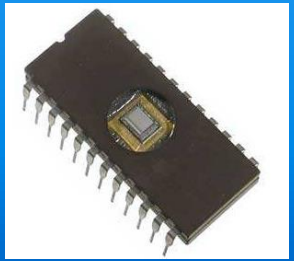




OTP - PROM

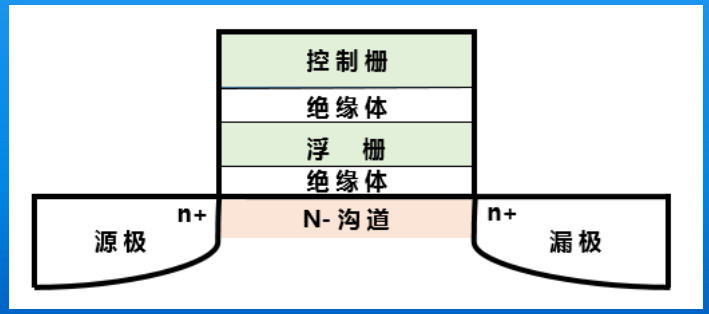
- **One-Time Programmable (OTP)**
- Typical OTP PROM
  - Wen Tsing Chow in 1956
- New OTP
  - Resistive RAM (RRAM | ReRAM)





MTP – EPROM & EEPROM

- **Multi-Time Programmable (MTP) Memory**
- Erasable Programmable Read-Only Memory，EPROM
- Electrically Erasable Programmable Read-only memory，EEPROM
  - EEPROM or E^2PROM, are Floating-Gate Transistors
  - Floating-Gate tech by Simon M. Sze & Dawon Khang, 1967
  - FRAM (or FeRAM) , MRAM are new



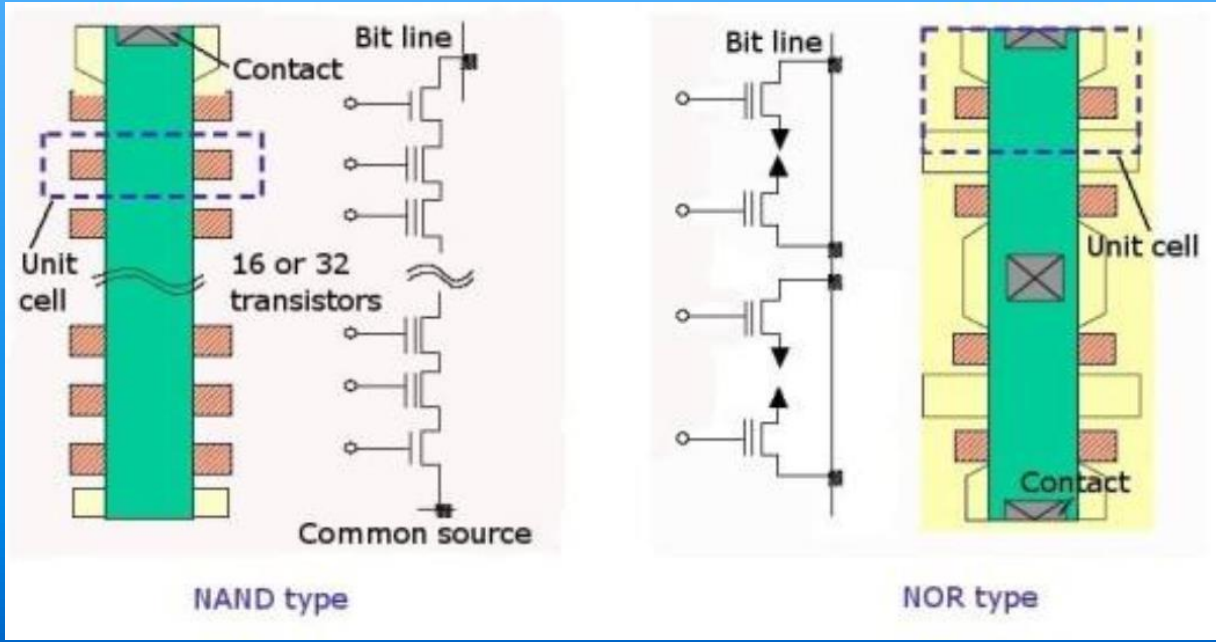
New Memory: NVRAM

- Ferroelectric RAM (FRAM or FeRAM) → E2PROM
  - popular high-κ gate dielectric HfO_2 (Sharp, Panasonic)
    - CBRAM (conductive-bridging RAM)
    - PCM (phase-change memory)
- Magnetic RAM (MRAM) → E2PROM
  - In production by Everspin Technologies, GF, Samsung

NAND Flash and NOR Flash

Both NAND & NOR types by Fujio Masuoka, at Toshiba 1985: US 4531203

- **NAND** Flash
  - Toshiba, 1989; ONFI
- **NOR Flash**
  - Intel, 1988; XIP Tech

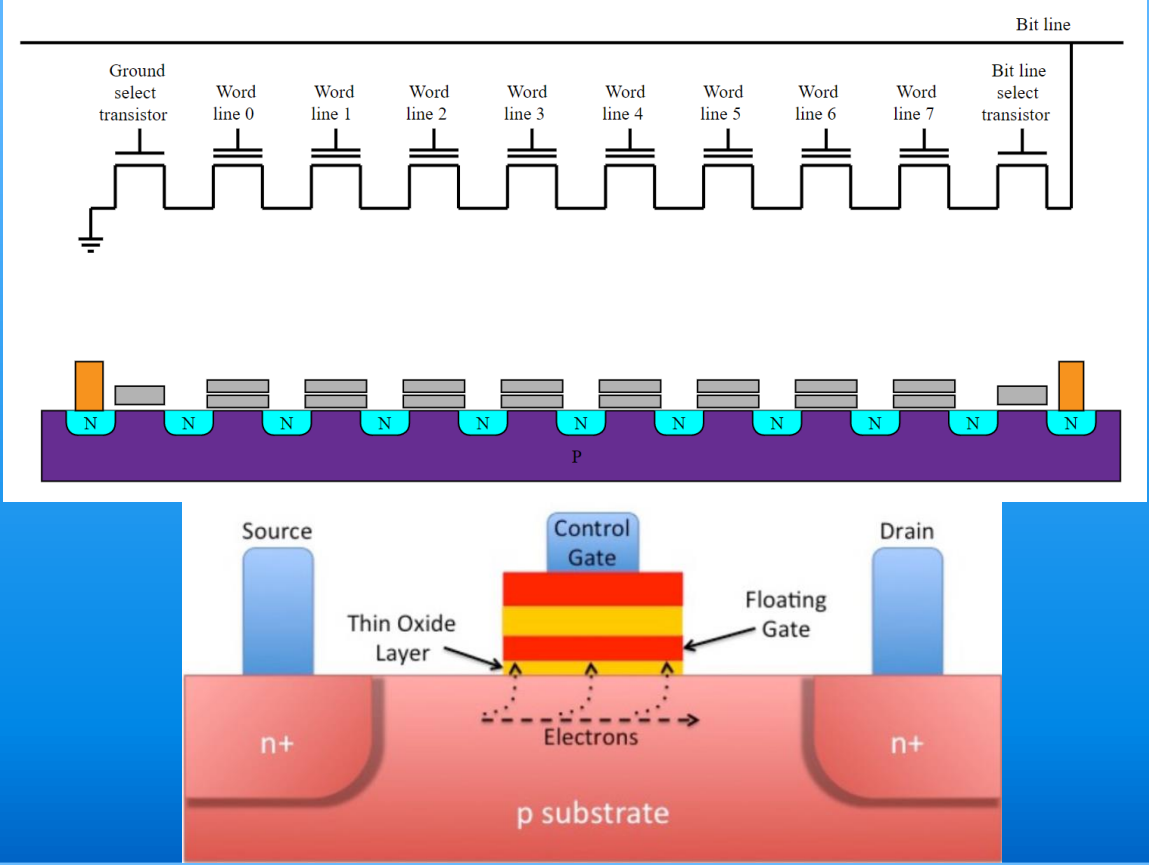


Multi-Level Cell (NAND Flash)

- SLC, Single-level cell, 1 bit/cell, 50k-100k times
- MLC, Multi-level cell, ECC needed
- TLC, Triple-level cell, used in SSD
- QLC!
- Comparison

The **floating-gate MOSFET** (**FGMOS**)

NAND Flash and Floating-Gate



SSD

- Application Areas
  - Notebooks, Desktops, Consumer Electronic
  - Embedded market (vehicle, industry control, commercial)
  - Servers, Storage in D. Ctr, Enterprise
  - Mil, Aero, Med
- Interfaces, indcluing
  - Fibre Channel (128 Gbit/s, in Servers)
  - **PCI Express** (PCIe Gen3 x 4, 31.5 Gbit/s)
  - SAS (Serial Attached SCSI, 12.0 Gbit/s, Servers)
  - **USB** (10 Gbit/s)、**Serial ATA** (SATA3.0, 6.0 Gbit/s) etc.

UFS (*Universal Flash Storage*) and UFS Card

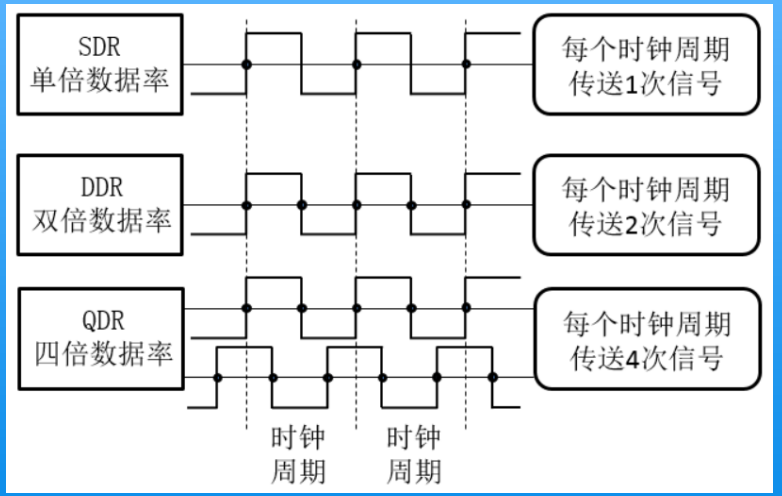
- Consumer Electronics
  - digital cameras, mobile phones & consumer electronic devices
  - *Nokia*, *Sony Ericsson*, *TI*, *STM*, *Samsung*, *Micron*, *SK Hynix*.
- SCSI Archi. Model; MIPI Alliance/M-PHY
- To replace eMMCs and SD cards; also eUFS
- JEDEC: 2011 V1.0 | 2013 V2.0 | 2018 V3.0
  - Total Bandwidth: 0.3GB/s, 1.2GB/s, 2.9GB/s
  - max. 5.8Gbit/s per lane …
- JEDEC: UFS Card: 2016 V1.0, 2018 V1.1

Memory Design & HBM

- Memory Designs: SRAM, DRAM, Flash
- **DDR/mDDR & GDDR: SDRAM, LPDDR, GDDR**
- RAM Cells: SRAM, DRAM, Flash
- HBM via 2.5D IC and 3D IC: Interposer, TSV
- Discussion: Applications

DDR

- Gen1 SDR (single datarate SDRAM)
- Gen2 DDR (SDRAM DDR)
- Gen3 DDR2
- Gen4 DDR3
- Gen5 DDR4
- 2017 DDR5



mDDR (LPDDR)

- Low Power DDR，LPDDR
  - Synchronous DRAM, LPDDR SDRAM, used in mobile, mDDR or Mobile DDR
- LPDDR/2/3/4

	LPDDR2	LPDDR3	LPDDR4
内部核心时钟 (MHz)	200	200	200
总线时钟 (MHz)	400	800	1600
数据速率 (Mbps)	800	1600	3200
带宽 (GB/s)	6.4	12.8	25.6
预取(prefetch)	4n	8n	16n
电压 VDD2/VDDQ/VDD1	1.2V/1.2V/1.8V	1.2V/1.2V/1.8V	1.1V/1.1V/1.8V
指令/地址总线	10 bits, DDR	10 bits, DDR	6 bits, SDR
Bank 数	4/8	8	8/ch (16)
容量	64Mb~8Gb	4Gb~32Gb	8Gb~32Gb
接口	HSUL_12	HSUL_12(可选ODT)	LVSTL
I/O 类型	x16/x32	x16/x32	2ch x16
封装	MCP/PoP	MCP/PoP	MCP/PoP

注：预取(prefetch)代表缓存大小(buffer size)。例如上表中 LPDDR3，其 prefetch 为 8n，代表每个预取的“数据字数(datawords)”为 8。

GDDR

- Graphics Double Data Rate, GDDR

性能	GDDR3	GDDR5
电压 VDD， VDDQ	1.5V, 1.35V	1.5V, 1.35V
时钟(MHz)	800/900	1750
数据倍率	2	4
数据传输率(Gbps)	1.6/1.8	6, 7, 8
I/O 宽度	(4)， 8, 16	32/16
存储块(Bank)数	8	16
预取	8n	8n
突发长度	4(Burst Chop), 8	8
典型颗粒容量(Gb)	1~2	2~8
循环冗余检验(Cyclic Redundancy Check， CRC)	~~N/A~~ 不适用	~~Yes~~ 适用
封装	BGA-78/96	BGA170

Memory Design & HBM

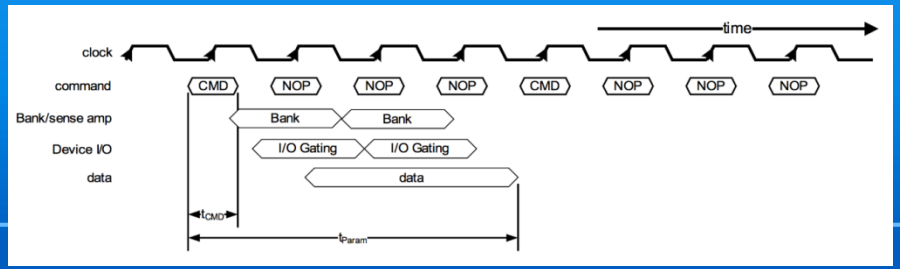
- Memory Designs: DRAM, SRAM, Flash
- DDR/mDDR & GDDR: SDRAM, LPDDR, GDDR
- **RAM Cells: SRAM, DRAM, Flash**
- HBM via 2.5D IC and 3D IC: Interposer, TSV
- Discussion: Applications

Digital Standard Cells and Memory Cells

- Combinational (or non-regenerative) logic
  - AND, OR, BUF; NAND, NOR, INV
  - Decoder/Encoder; Mux; XOR/XNOR; Add/Sub/ALU
  - PUN (pull-up network), PDN (pull-down network)
- ***Sequential*** (or regenerative) logic
  - Latches/FFs; registers (D/FF), counters, oscillators,
  - State-Machine; **Memory**
    - **Memory:** ROM**, RAM (SRAM, DRAM)**

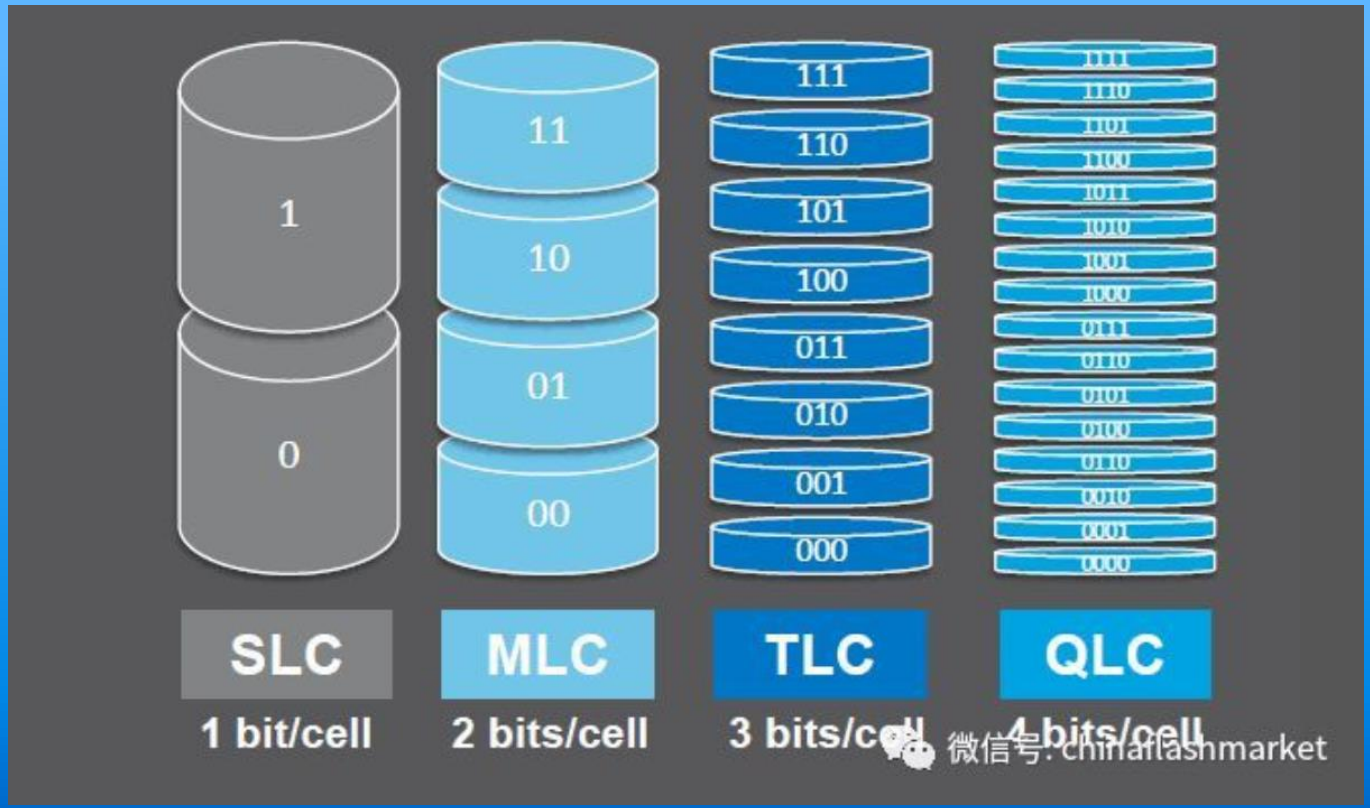
RAM Cell Structure and Timing

- SRAM
  - Inputs/Outputs
  - Structure (RAM Cell): add (decoder), bit-line, word-line
  - **Timing**
    - For read: t_AA, t_ACS, t_OE, t_OZ, t_OH
      - (Access from add/chip_sel/out_en/out_dis/out_hold)
    - For write: t_AS, t_AH, t_CSW, t_WP, t_DS, t_DH
      - (Access_setup_before/acc_hold_after/chip_sel_before/write_pulse/ data_setup_before/data_hold_after)
- DRAM



3D NAND Cell

- QLC



3D NAND Controller IC

	型号	制程	DRAM	支持 Flash	Flash 通道	支持接口	接口协议	ECC
Marvell	88NV1160	28nm	N	2D/3D TLC	4	PCIe3.0 x2	AHCI & NVMe1.3	LDPC
	88NV1120	28nm	N	2D/3D TLC	2	SATA 3.0	AHCI	LDPC
	88SS1084	28nm	Y	3D TLC/QLC	4	PCIe3.0 x4	NVMe1.3	LDPC
	88SS1100	28nm	Y	3D TLC/QLC	8	PCIe3.0 x4	NVMe1.3	LDPC
慧荣	SM2263XT	28nm	N	3D TLC/QLC	4	PCIe3.0 x4	NVMe1.3	LDPC
	SM2258XT	40nm	N	3D TLC	4	SATA 3.0	AHCI	LDPC
	SM2262EN	28nm	Y	3D TLC/QLC	8	PCIe3.0 x4	NVMe1.3	LDPC
群联	PS5008-E8T	40nm	N	2D/3D TLC	4	PCIe3.0 x2	NVMe1.2	SmartECC
	PS3111-S11	40nm	N	2D/3D TLC	2	SATA 3.0	AHCI	LDPC
	PS5012-E12	28nm	-	3D TLC/QLC	-	PCIe3.0 x4	NVMe1.3	LDPC

来源：中国闪存市场网 www. chinaflashmarket.com

Memory Design & HBM

- Memory Designs: DRAM, SRAM, Flash
- DDR/mDDR & GDDR: SDRAM, LPDDR, GDDR
- RAM Cell: SRAM, DRAM, Flash
- **HBM via 2.5D IC and 3D IC: Interposer, TSV**
- Discussion: Applications

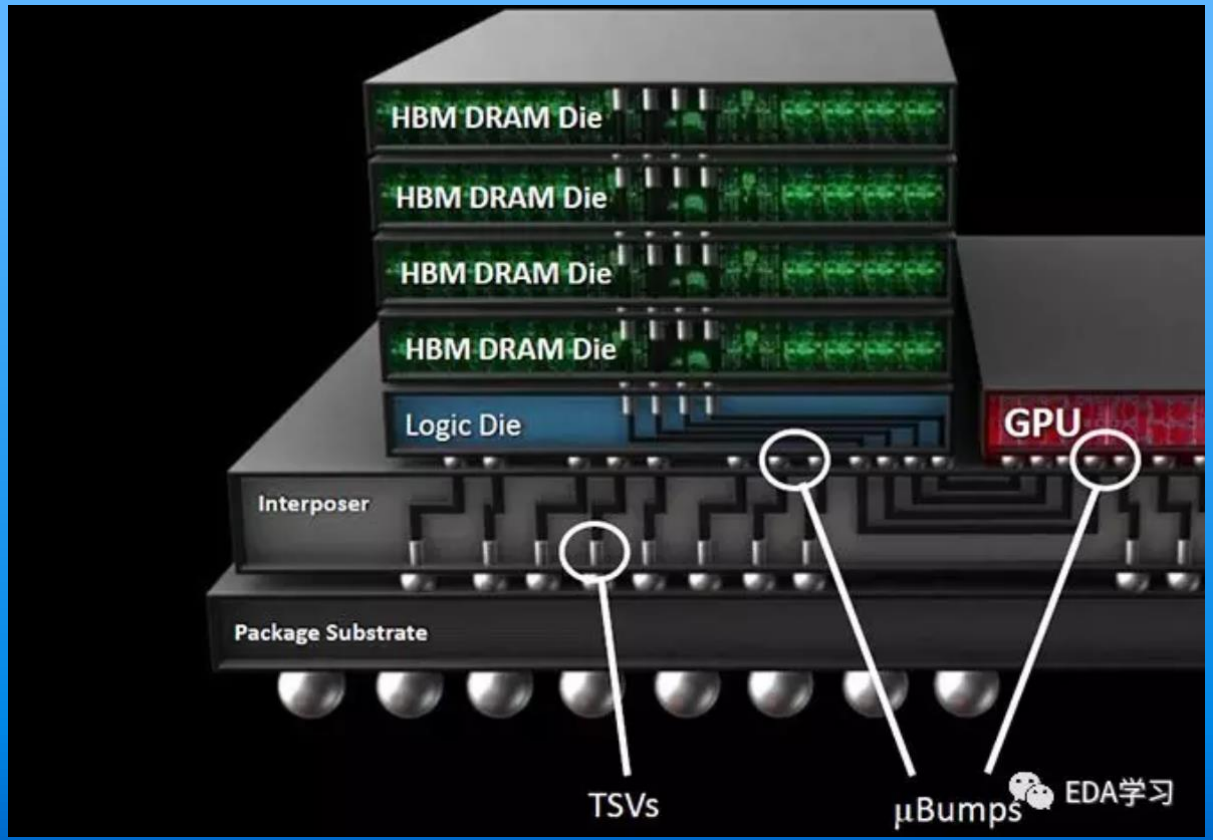
HBM via 3D IC

- Existing 4 Packaging Standards
  - Wide-IO (Samsung); HBM (AMD, Nvidia, Hynix);
  - HMC (Micron); CoWoS & InFO (TSMC)
- HBM via 3D IC and HBM via 2.5D IC
  - 3D IC (TSV + µBumps); 2.5D IC (Interposer, w/ KGD)
  - Components: DRAMs (DDR, SDRAM), GPU, …
  - FC and BGA
- HBM via FPGA (ex. Xilinx)

HBM for AI (RAM+GPU/CPU) Available

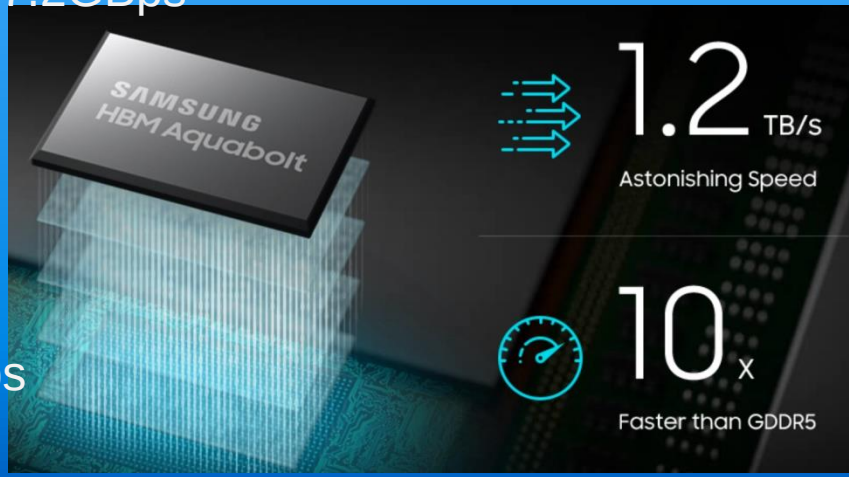
- HBM for CPU/GPU
- Gen1
- Gen2
  - 1024-bit word? 256 Gbps band
- Gen3? X4
  - 4096-bit word? 1024 Gbps band (>1 Tbps)

HBM via 3D IC and 2.5D IC



HBM at Samsung

- Samsung HBM2 Flarebolt's performance: 2.0Gbps at 1.35V
- HBM2 - GDDR5 data transmission speed calculation
  - 8GB HBM2 package's data bandwidth:
    - 2.4Gbps per pin x 1024bit bus = 307.2GBps
- 4 HBM2 package in a system:
  - 307.2GBps x 4 = appr. 1.2TBps
  - 8Gb GDDR5 die's data bandwidth:
    - 8Gbps per pin x 32bit bus = 32GBps



Memory Design & HBM

- Memory Designs: DRAM, SRAM, Flash
- DDR/mDDR & GDDR: SDRAM, LPDDR, GDDR
- RAM Cell: SRAM, DRAM, Flash
- HBM via 2.5D IC and 3D IC: Interposer, TSV
- **Discussion: Applications**

Applications of Memory Cells

- DRAM Market
  - Old: PC 10% decrease
  - New: Mobile, Server, Auto
  - Future: AI Edge Computing (Video Surveillance)
    - MEC (Mobile Edge Computing)
    - Related: Grid Computing
- NAND Market
  - SSD
  - Mobile

3D NAND

- Global Number of Leading Inventors on 3D NAND Flash