# A Comparison of current SDRAM types: SDR, DDR, and RDRAM

**B. Klehn and M. Brox**

Infineon Technologies AG, Munich, Germany

**Abstract.** The ever increasing demand for bandwidth of computer-systems lead to several standards of SDRAMs. This article compares SDR, DDRI, DDRII, and RDRAM systems. Besides the overall basic innovations, differences will be discussed. Topics like architecture, interfaces, and modules are described.

## 1 Introduction

Over the last years the clock rate of a microprocessor in a typical desktop PC has exceeded the 2 GHz number. To feed the processor and the other active components in such a system with data, the memory sub-system has to be able to provide the data sufficiently fast. To achieve this goal, multiple innovations had been introduced and integrated into ever more powerful devices. On the other hand, important basic concepts have remained in use over various generations. In this article, we intend to describe differences and commonalities of the various main-memory types and discuss reasons for the changes. In this respect, we will start with a discussion of the basic concept of a DRAM cell.

## 2 Basic DRAM functionality

The memory cell of a DRAM consists of a capacitor and a select transistor (Fig. 1). The binary information is stored as a charge on the capacitor. Cells are oriented along a matrix of parallel bitlines and perpendicular wordlines. The read-out of a memory cell is shown in Fig. 2. To prepare for the read-out, bitlines are pre-charged to an intermediate level, which generally is set at half of the array operation voltage. The read-out starts with decoding the externally supplied row-address into a physical wordline. Asserting this wordline opens the select transistor. As the voltage on the storage capacitor is either lower (logical '0') or higher (logical '1') the

*Correspondence to:* B. Klehn
(bernd.klehn@infineon.com)

voltage on the BL changes during this process. As an example, Fig. 2 shows the read-out of a '1' causing the bitline voltage to increase (trace BLt). Typical values for a storage capacitor are in between 20 fF and 40 fF, which is much lower than the value for the parasitic bitline capacitance (between 100 fF and 200 fF). Therefore, the voltage change on the bitline is relatively small – often in the order of some 10 mV, only. To be able to securely read the cell, the voltage change is amplified in the differential sense-amplifier which compares voltage on the bitline BLt with the voltage on a neighbouring, floating bitline BLc. At the end of a sense-amplification, all cells of one wordline are read out and read-out information has been amplified to digital values. Selection of the sense-amplifier to be read is performed through the column-decoder. Dependent on the column-address, one sense-amplifier gets connected to the data-lines over the array, re-amplified in a second amplifier stage and driven towards the output buffers (Fig. 3).

A common method to achieve a continuous data-stream from a DRAM is the so-called burst-mode. Here, multiple data are read-out from one line of sense-amplifiers one after another. The memory controller has to supply only the address of the first read access; the DRAM internally generates all the subsequent column-addresses. In its simplest implementation, however, this continuous stream would have to stop the moment a new wordline would have to be activated. Activation of a different wordline requires bringing down the old wordline, precharging the bitlines and asserting the new wordline.

In order to minimize this severe access penalty, all modern DRAM types are internally arranged as multiple quasi-independent memories. Fig. 3 shows a typical block-level layout for a 256 Mb-SDRAM. Memory cells are organized into four independent memory banks of 64 Mbit, each. Organization into multiple banks allows e.g. a wordline activation in bank 1 to appear concurrently to a read-operation in bank 0. Thus, as long as the row-address sequence supplied from the memory controller is well behaved, a continuous read data-stream can be provided. Yet, in typical applica-
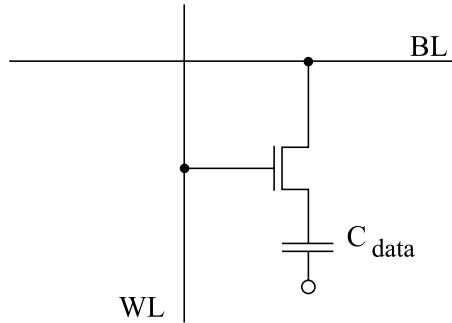
**Fig. 1.** Schematics of a memory cell build from a storage capacitor and a select transistor. The select transistor connects the capacitor with the bitline BL. Activation of the select transistor is performed through wordline WL.
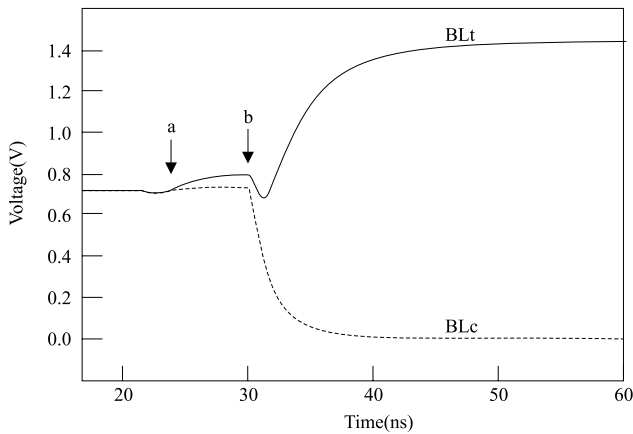


**Fig. 2.** Bitline voltage as a function of time during read-out of a memory cell. Example shows read-out of a '1' from a cell attached to BLt. At a, wordline gets asserted and charge from the memory cell flows onto the bitline. The neighbouring bitline BLc gets capacitively coupled high. At b, charge transfer from the cell is complete and the sense-amplifier can get activated. After sense-amplification at around 40 ns, digital levels are present on BLt and BLc.

tions, peak bandwidth cannot be sustained. Often word-lines need to be activated in the same bank. In this case the delay encountered by the array pre-charge cannot be hidden. Activation of wordlines, sensing and pre-charging are dominated by parasitic delays as heavily loaded, yet narrow lines in the array need to be driven. Shrinking device dimensions yielded only a performance improvement of a factor around 3 (see Fig. 4). This is rooted in the fact that devices as implemented in DRAMs need to follow other device design guide lines such as e.g. devices in logic processes. As the number of DRAMs in a memory system is often very large, the standby current has to be as low as possible, which forces one to use transistors with some 100 mV of threshold voltage. To close the gap between the processor and the memory, circuit techniques have to be employed, which will be discussed in the following section.

## 3   Clock synchronization

Older memory devices up to 66 MHz used an external asynchronous interface. The memory itself did not have an external clock input. Any activity was purely edge-triggered and commands could be issued at any arbitrary time as long as the internal time constants of the memory were met. For higher frequency operation, to help ensure reliable input of data, addresses, and commands, an external clock has been used. Typically, the memory samples the input lines at the rising edge of the clock. Commands can, thus, be issued to the memory at only multiples of the clock cycle time. Legal command combinations are described in detail in the datasheet of the manufacturer. Use of an external clock marked the introduction of the so-called SDRAM.

### 3.1   SDR

Standardized SDRAMs input commands and addresses at the rising edge of the clock. In a first development step, input and output of data was also possible at the rising clock-edge, only (Fig. 5). This concept permits a single data transmission per clock cycle and consequently led to the name of Single-Data-Rate-SDRAM (SDR-SDRAM). The clocksignal itself is single-ended. Output of the data itself was not specifically synchronized to this clock. The only requirement is that following the clock edge for a minimum time of $t_{OH}$ the previous data is held valid, while after $t_{AC}$ new data is available. To give typical values for a 133 MHz-SDR-SDRAM, $t_{OH} = 3$ ns, and $t_{AC} = 5.4$ ns. In between, no valid data is present on the bus.

### 3.2   DDR

The natural evolution of this concept towards higher bandwidth, is to allow input and output of data at twice the rate. This scheme as shown in Fig. 5 is called Double-Data-Rate (DDR). DDRI and DDRII-SDRAMs derive their name from this data transmission approach.

As illustrated in Fig. 5 for the same clock-frequency, the so-called data-eye (the amount of time data is valid) is only half as wide in DDR as in SDR. Increased precision can be gained through an improved clock system architecture as will be discussed in more detail below. An important device level change, however, is the introduction of an active alignment of the data-output to the clock. For this purpose, Delay-Locked-Loop (DLL) circuits were added to the DDR-devices. In a DLL, the received external clock is purposely delayed such that $t_{AC}$ can be made as close to 0ns as possible. Comparing to SDR-SDRAMs, a typical value for a 133 MHz-DDR-SDRAM is $-0.75$ ns $< t_{AC} < 0.75$ ns. In a further change, the clock itself is improved. The single-ended clock of the SDR is replaced by a fully-differential clock. As a timing reference, the clock cross-point is defined which is more precisely to determine than the clock edge especially in the presence of significant system noise.

(a)



(b)
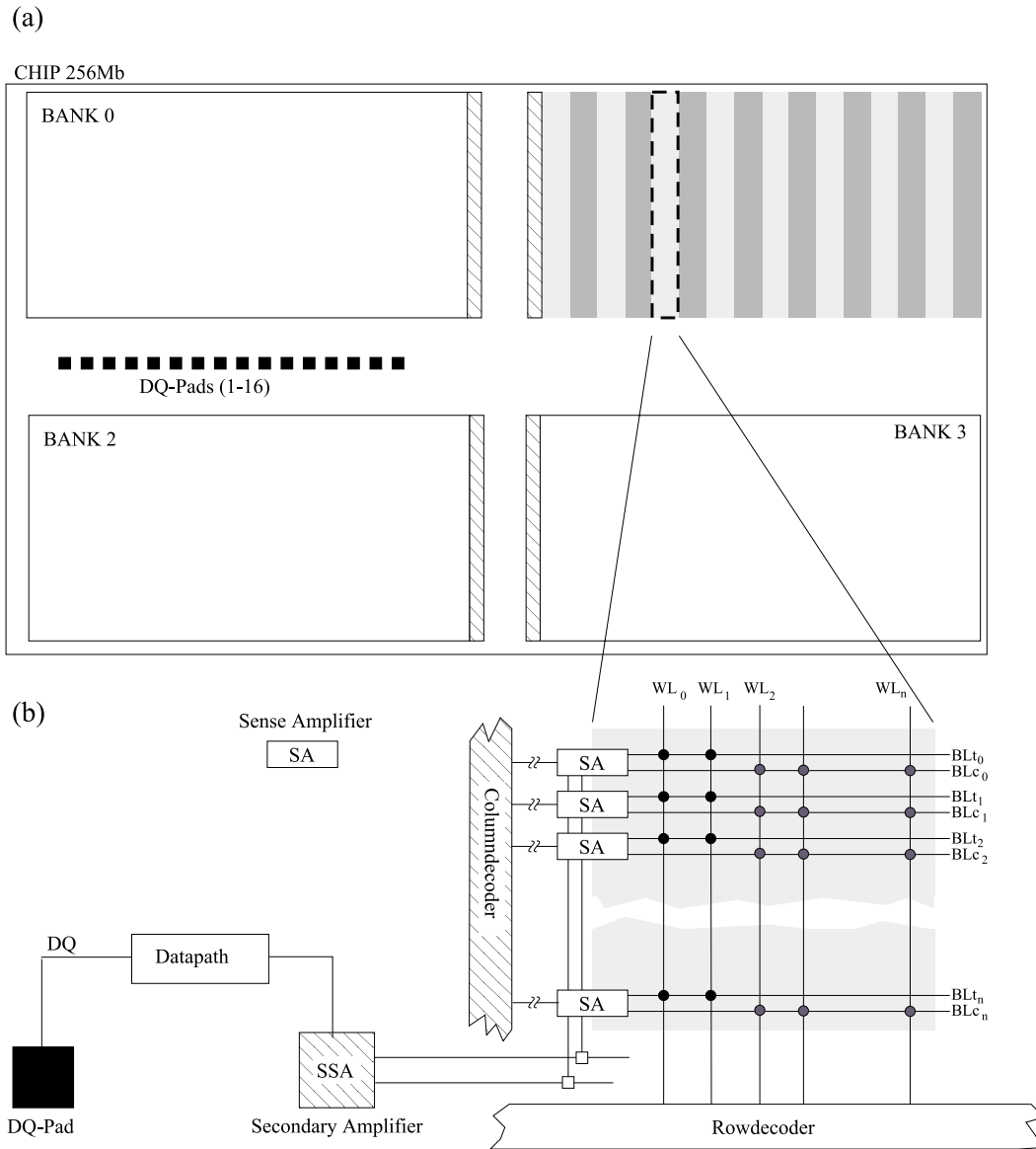
Fig. 3. (a) Sketch of the architecture of a typical 256 Mb-DRAM. Memory cells are laid out in four memory banks of 64 Mb. Each bank can be treated as a partially independent memory. (b) Sub-division of one 64 Mb-bank into 4 Mb-units. The row-decoder selects one WL. Through column-decoding, one sense-amplifier gets selected to drive data towards the data-path.

## 4 Prefetch

A memory using the DDR output-scheme has to deliver up to 16bit for each half clock cycle. This could be implemented by reading out the array twice per clock-cycle. At a clock rate of 200 MHz this would be equivalent to a read-out rate of 400 MHz which could be facilitated if the array is partitioned into very small sub-units with short bitlines and short wordlines. Short bitlines, however, require a high number of sense-amplifiers and short wordlines a high number of row-decoders on the chip. A more typical solution is to prefetch $2 \times 16$ bit in one array-access for one DDR clock-cycle (more precise: $2\times$-prefetch). Under this condition, core frequency can be held constant at 200 MHz, while data-

rate achieves 400 Mb/s per pin. Figure 6 illustrates the principle, which also leads to a simple extension path towards even higher bandwidths while still keeping core frequency constant. RDRAM in contrast, uses an $8\times$-prefetch to reach a data-rate of 1066 Mb/s per pin while running the core at a relaxed 133 MHz-frequency. Time-multiplexing circuitry in the chip has to guarantee only that prefetched data is driven out in the correct order.

## 5 Differences in the concepts

Table 1 gives an overview over clock-rates, bit-rates and core frequency for the various architecture as presented above. While prefetching is effective in limiting the core frequency

**Table 1.** Performance summary of the relevant architectures. Typical values for frequency is given

| Typ | Clock(MHz) | Core(MHz) | Scheme | Datarate per pin (Mb/s) | Prefetch | Data-width |
|------|------------|-----------|--------|--------------------------|----------|------------|
| SDR | 133 | 133 | single | 133 | none | 64 |
| DDRI | 133 | 133 | double | 266 | 2 | 64 |
| DDRII | 266 | 133 | double | 533 | 4 | 64 |
| RDRAM | 533 | 133 | double | 1066 | 8 | 16 |
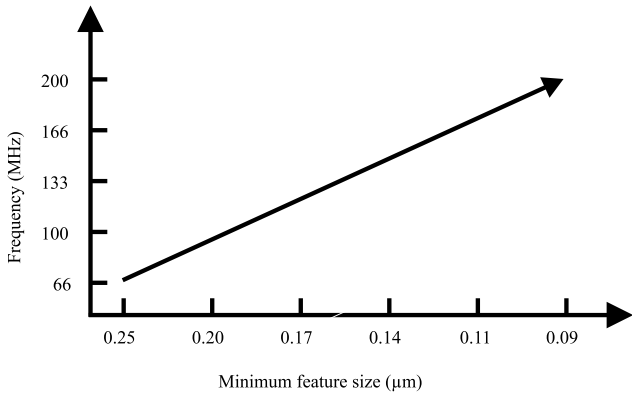


**Fig. 4.** Operation frequency improvement through smaller feature size. Improvement is around a factor of 3 going from $0.25\,\mu$m to $0.09\,\mu$m.
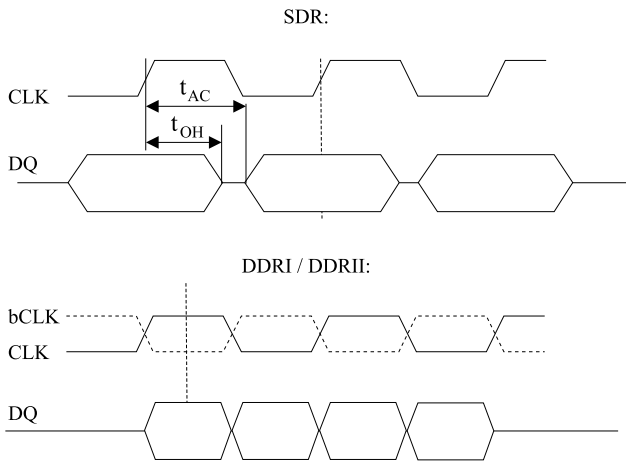


**Fig. 5.** The graph depicts the external clock inputs for SDR and DDRI/DDRII. The crossing of CLK and bCLK defines the point of time t1 much more fail-safe compared to the edge of a single ended clock.

to sustainable values for each architecture, the frequency of the data lines is differing strongly. Here, architectural differences between the different types are becoming relevant. The following section will touch upon these.
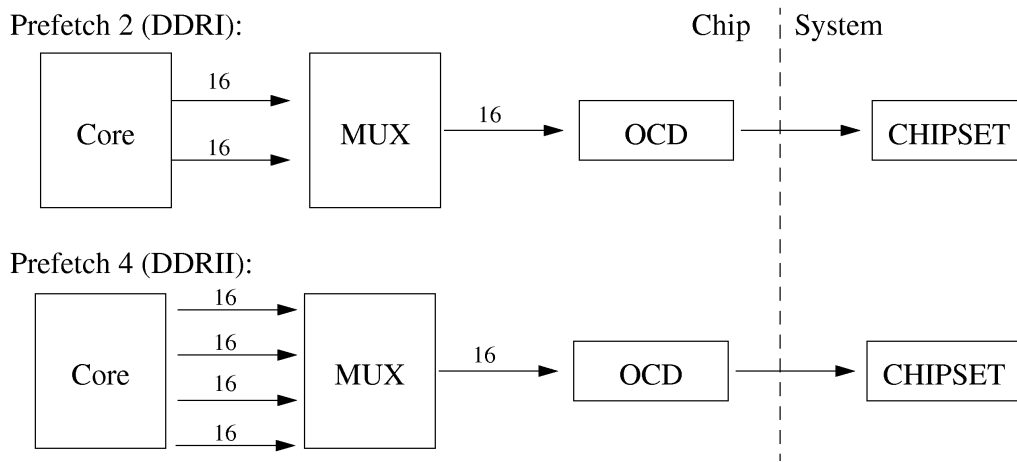
### 5.1 Commands and requests

The memory controller needs a way to initiate activity in the memory. The classical way is the simple command protocol
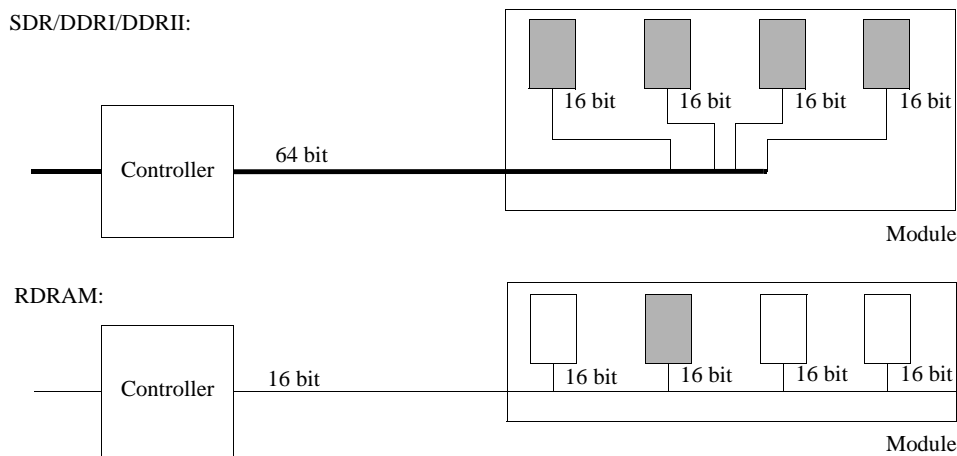
as implemented on the evolutionary SDR/DDR family of devices. All control and address information for one command in the DRAM is transmitted during a single clock cycle. This protocol requires typically 5 control- and 15 address-lines to encode all relevant commands (e.g. activate row, precharge row, read column, write column, ...). In this approach, the complexity of the address and control lines is limited because these lines are run in single-data-rate fashion, however, on the expense of a rather large number of lines. Another advantage is the small latency: complete decoding of the command can procede after only one clock cycle. RDRAMs follow a completely different approach. Here, e.g., a single set of eight lines for addresses and control may be used. Three of these lines transmit row oriented requests; five transmit column oriented requests. The three row-oriented lines handle row activation and row precharge. To be able to encode an e.g. 16bit wide address onto 3 lines, the information is time-multiplexed and transmitted in DDR-like fashion. On a first look this introduces multiple disadvantages. Four clock cycles are required to transmit all information required in one access, which increases access latency compared to the one-clock cycle approach of SDR/DDR devices. Furthermore, a capture-DLL needs to be implemented as the data-eye onto the shared control, data, and address lines (the bus) gets too narrow to be received with conventional means. As, however, the number of lines gets very small, the bus can be run at very high frequency while still keeping total system cost at a reasonable level. In addition, seperation of the bus into row-request and a column-request sections enables a very flexible protocol: row and column operations can be initiated independent from and parallel to each other. For example, it is possible to issue a read to bank 0 and an activate to bank 1 in parallel.

### 5.2 Module design

SDR and DDR systems are addressing multiple devices on a module in a parallel fashion. As an example, Fig. 7 shows the 64 bit data lines from the SDR/DDR controller separating into four 16 bit sets of lines running to four 16 bit wide memories in parallel. In RDRAM, devices are placed serially on the bus. At a given time, only one device can be actively reading or writing. To select one out of multiple devices, this approach requires an additional device identifier besides the conventional row or column address for selection within one memory. Here again, the distinction between wide, lower-frequency SDR/DDR data lines and the narrow,

Prefetch 2 (DDRI):

Chip | System

Core — 16 → 16 → MUX — 16 → OCD — → CHIPSET

Prefetch 4 (DDRII):

Core — 16, 16, 16, 16 → MUX — 16 → OCD — → CHIPSET

**Fig. 6.** Prefetch: Capturing of 2x resp. 4x the required data volume and time-multiplexing it onto the output leads to a simple multiplication of the external data-rate of the memory device.

SDR/DDRI/DDRII:

Controller — 64 bit — 16 bit | 16 bit | 16 bit | 16 bit

Module

RDRAM:

Controller — 16 bit — 16 bit | 16 bit | 16 bit | 16 bit

Module

**Fig. 7.** SDR- and DDR-modules are operating in desktop-systems with a data-width of 64 bit and address typically 8 devices in parallel. The RDRAM-bus is significantly narrower (16 bit) and only one device is addressed. RDRAM bandwidth is increased by running the bus at a very high clock-frequency.

higher-frequency RDRAM bus gets obvious. On the other hand, RDRAMs are forced to operate at higher clock-rate because for a given clock-rate a DDR-system would be able to deliver four times the RDRAM peak bandwidth simply because of the 64 bit data-width.

### 5.3 Termination, clocking and strobing

As discussed before, SDR-systems are using a simple, single-ended external clock signal. All data lines are non-terminated and no specific additional means exist to improve signal quality and data valid window. Output drivers are simple voltage source push-pull type drivers. On chip-level, DDR and RDRAM implement DLLs for better alignment of the data output. However, differences do not stop here but continue on the architectural level. Figure 8 shows the clocking and output schemes realized in the different advanced DRAM types.

DDRI uses a mid-level termination of all lines. Addition-

ally, stub-resistors are introduced to limit reflections. A further, important change is the introduction of the data-strobe signal (DQS). DQS is a bi-directional synchronization signal, which is generated by the controller during a write access to the memory and by the memory during a read-access to the controller. In this way, DQS travels in the same direction as the data signal, itself. DQS is generated in a way that each edge (rising and falling) indicates valid data. As the flight time is the same for DQS and DQ, both controller and memory can use the DQS signal to capture incoming data. Different arrival times of the uni-directional always-on clock signal can thus be effectively cancelled out.

DDRII uses the very same DQS-scheme. Only termination is moved onto the chip itself (on-die-termination, ODT). ODT further improves signal quality and reduces total system cost as discrete termination resistors are no longer required. To support multi-module systems, all associated devices need to be able to turn-on or turn-off ODT on-demand.

(a)

(b)
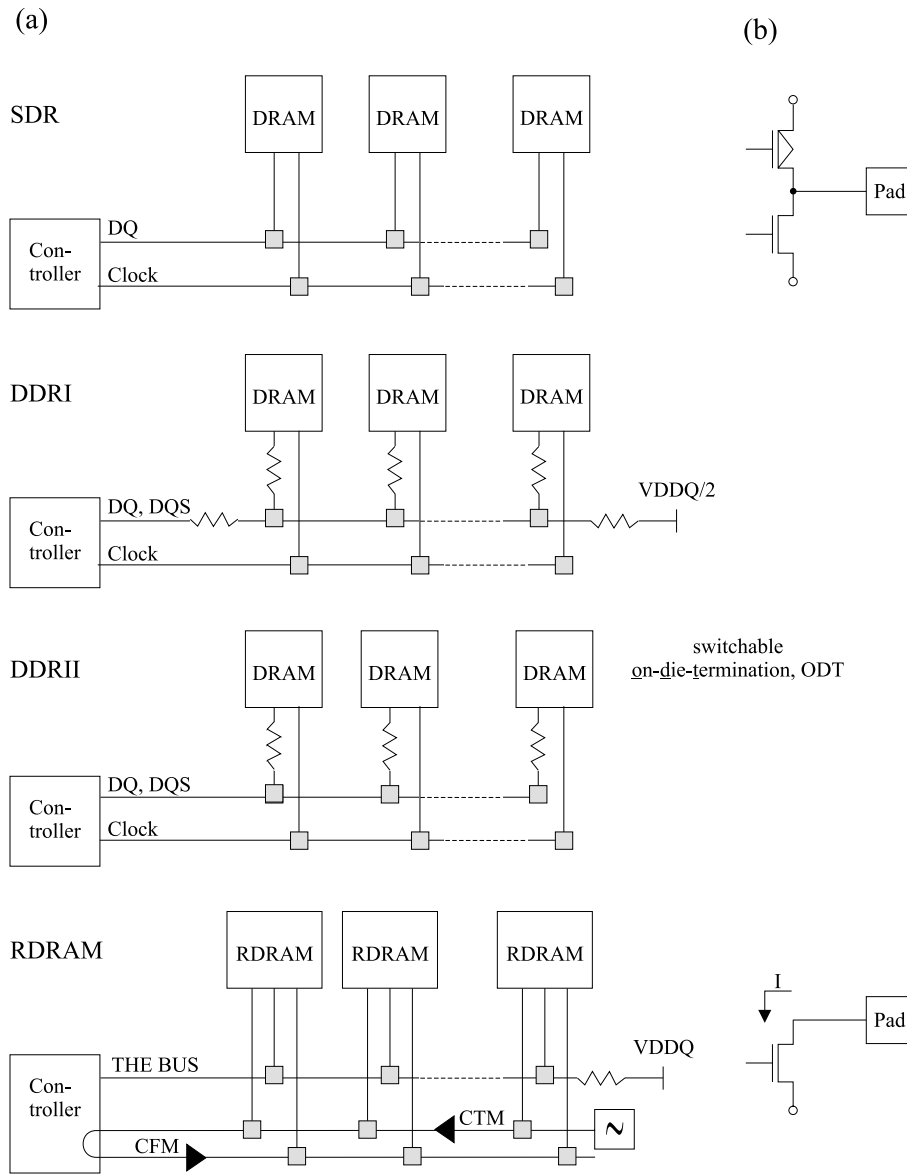


**Fig. 8.** Data line topologies for the different DRAM types **(a)**: SDR, DDRI, DDRII, and RDRAM. **(b)** Simplified final driver stage.

RDRAMs employ a very different scheme. A clock is generated at the far end of the bus. The CTM-clock travels in the direction of the controller, gets mirrored-back at the controller and afterwards travels as CFM-clock on its way back to the far end of the bus. These two clocks can be used to synchronize write data to the memory with CTM, and read-data from the memory with CFM. In this scheme, travel direction of relevant clock and data is the same such that flight time differences are cancelled out. A high level termination is present only at the far-end of the bus such that the controller represents an open end to the bus. At this open end signal reflection occurs, which has to be accounted for in the RDRAM system design. Especially, RDRAM drivers need to be designed with a high differential output resistance such that no second reflection occurs when a reflected wave passes a data-driving memory. This can be guaranteed by designing

the output-driver as a current source, which makes its design very much different than the SDR/DDR voltage mode driver. Signaling levels are currents of 0 mA for a logical '0' and typically 30 mA for a logical '1'. Only on the data lines, the travelling current wave translates to a travelling voltage wave which can be received by memory respective controller.

## 6 Chip area

If the same technology and number of storage cells are assumed the architecture influences the chipsize. Here, the number of banks is important. Also a higher prefetch needs more datalines running from the banks to the multiplex unit. The number is increased by a factor equal to the prefetch. The 16 data lines from a SDR-SDRAM are doubled for

DDRI and quadrupled for DDRII. The package-oriented communication between a controller and an RDRAM requires a complex decoder logic block in the I/O part. The reduced supply voltages of emerging products adds to the area because of complicated on-chip pump and generator systems.

## 7  Outlook

It is possible to run the core of an SDRAM with frequencies of 133, 166 or even more than 200 MHz. Combining it with a high prefetch and a large number of parallel devices, data rates in the region of several gigabytes per second are possible. Otherwise to have a noticeable speed up in a computer system all components of such a system have to be balanced. With these bandwidth image processing and simulations are possible at high level. To enter new sophisticated applications a higher operating speed of computer systems is still desirable.

## References

Itoh, K.: VLSI Memory Chip Design, Springer-Verlag, Berlin Heidelberg New York, 2001.

Infineon Technologies, SDRAM Datasheets, www.infineon.com/cgi/ecrm.dll/ecrm/scripts/prod_ov.jsp