JEDEC STANDARD

GRAPHICS DOUBLE DATA RATE (GDDR6) SGRAM STANDARD

JESD250

JULY 2017

JEDEC SOLID STATE TECHNOLOGY ASSOCIATION

Published by
©JEDEC Solid State Technology Association 2015
3103 North 10th Street
Suite 240 South
Arlington, VA 22201-2107

Contents

GRAPHICS DOUBLE DATA RATE 6 (GDDR6) SGRAM

(From JEDEC Board Ballot JCB-17-16, formulated under the cognizance of the JC-42.3C Subcommittee on DRAM Parametrics.)

1 SCOPE

This document defines the Graphics Double Data Rate 6 (GDDR6) Synchronous Graphics Random Access Memory (SGRAM) specification, including features, functionality, package, and pin assignments.

The purpose of this Specification is to define the minimum set of requirements for 8 Gb through 16 Gb x16 dual channel GDDR6 SGRAM devices. System designs based on the required aspects of this standard will be supported by all GDDR6 SGRAM vendors providing compatible devices. Some aspects of the GDDR6 standard such as AC timings and capacitance values were not standardized. Some features are optional and therefore may vary among vendors. In all cases, vendor data sheets should be consulted for specifics. This document was created based on some aspects of the GDDR5 Standard (JESD212).

2 GDDR6 SGRAM STANDARD OVERVIEW

8 Gb	=	2 Channels 256Mb x 16	2 x (16Mb x 16 x 16 banks)	/	2 Channels 512Mb x 8	2 x (32Mb x 8 x 16 banks)
12 Gb	=	2 Channels 384Mb x 16	2 x (24Mb x 16 x 16 banks)	/	2 Channels 768Mb x 8	2 x (48Mb x 8 x 16 banks)
16 Gb	=	2 Channels 512Mb x 16	2 x (32Mb x 16 x 16 banks)	/	2 Channels 1Gb x 8	2 x (64Mb x 8 x 16 banks)
24 Gb	=	TBD				
32 Gb	=	TBD				

2.1 FEATURES

- 2 separate independent channels with point-to-point interface for data, address and command
- Half CA data rate differential clock inputs CK_t/CK_c for CMD/ADD (CA) per 2 channels
- Four half data rate or quarter data rate differential clock inputs WCK_t/WCK_c, each associated with a data byte (DQ, DBI_n, EDC) or Two quarter data rate or half data rate differential clock input WCK_t/WCK_c, each associated with the two bytes in the channel (Vendor specific)
- Double Data Rate (DDR) or Quad Data Rate (QDR) data (with regards to the WCK) (Vendor specific)
- Double Data Rate (DDR) Command Address (with regards to the CK)
- 16 internal banks
- 4 bank groups for t_{CCDL} = 3 t_{CK} and 4 t_{CK}
- 16n prefetch architecture: 256 bit per array read or write access per channel
- Burst length: 16 only
- Programmable READ latency: 9 to 36 t_{CK}
- Programmable WRITE latency: 5 to 8 t_{CK}
- WRITE Data mask function via CA bus (single/double byte mask)
- Data bus inversion (DBI) & Command Address bus inversion (CABI)
- Input/output PLL/DLL on/off mode
- Command Address training: command address input monitoring by DQ/DBI_n/EDC signals
- WCK2CK clock training with phase information by EDC signals
- Data read and write training via READ FIFO (depth 6)
- READ FIFO pattern preload by LDFF command
- Direct write data load to READ FIFO by WRTR command
- Consecutive read of READ FIFO by RDTR command
- Read/Write data transmission integrity secured by cyclic redundancy check using either a half or full data rate CRC
- READ/WRITE EDC on/off mode
- Programmable EDC hold pattern for CDR
- Programmable CRC READ latency = 1 to 4 t_{CK} and CRC WRITE latency = 10 to 16 t_{CK}
- Low Power modes
- On-chip temperature sensor with read-out
- Auto precharge option for each burst access
- Auto refresh & self refresh modes
- 32ms, auto refresh (16k cycles)
- Temperature sensor controlled self refresh rate and Partial Array Self Refresh
- Per-Bank / Per-2-Bank Refresh
- Optional digital t_{RAS} lockout
- On-die termination (ODT)
- ODT and output driver strength auto-calibration with external resistor ZQ
- Programmable termination and driver strength offsets (40 ohm to 60ohm)
- Internal VREF for data inputs and CA inputs with programmable levels
- Separate internal VREF for CA (Command / Address) inputs
- Vendor ID1 and ID2 for identification
- x16/x8 mode configuration set at power-up with EDC
- Pseudo-channel mode configuration set at power up with CA6
- 1.35V +/- 0.0405V supply for device operation (V_{DD})
- 1.35V +/- 0.0405V supply for I/O interface (V_{DDQ})
- 1.8 + 0.108V / - 0.054V supply for V_{PP}
- 180 ball BGA package with 0.75mm pitch

2.2 FUNCTIONAL DESCRIPTION

The GDDR6 SGRAM is a high-speed dynamic random-access memory designed for applications requiring high bandwidth. GDDR6 devices contain the following number of bits:

8 Gb has 8,589,934,592 bits
12 Gb has 12,884,901,888 bits
16 Gb has 17,179,869,184 bits
24 Gb has 25,769,803,776 bits
32 Gb has 34,359,738,368 bits

The GDDR6 SGRAM's high-speed interface is optimized for point-to-point connections to a host controller. On-die termination (ODT) is provided for all high-speed interface signals to eliminate the need for termination resistors in the system.

GDDR6 uses a 16n prefetch architecture and a DDR or QDR interface to achieve high-speed operation. The device's architecture consists of two 16 bit wide fully independent channels.

GDDR6 operates from a differential clock CK_t and CK_c. CK is common to both channels. Command and Address (CA) are registered at every rising edge of CK and every falling edge of CK. There are both single cycle and multi cycle commands. See command truth table for details.

GDDR6 uses a free running differential forwarded clock (WCK_t/WCK_c) with both input and output data registered and driven respectively at both edges of the forwarded WCK. See Clocking section for details.

Read and write accesses to GDDR6 are burst oriented; accesses start at a selected location and consists of a total of sixteen data words. Accesses begin with the registration of an Activate command, which is then followed by a Read, Write (WOM) or masked Write (WDM, WSM) command.

The row and bank address to be accessed is registered coincident with the Activate command. The address bits registered coincident with the Read, Write or masked Write command are used to select the bank and the starting column location for the burst access.

This specification includes all features and functionality required for GDDR6 SGRAM devices. In many cases the GDDR6 specification describes the behavior of a single channel.

2.3 DEFINITION OF SIGNAL STATE TERMINOLOGY

GDDR6 SGRAM will be operated in both ODT Enable (terminated) and ODT Disable (unterminated) modes. For highest data rates it is recommended to operate in the ODT Enable mode. ODT Disable mode is designed to reduce power and may operate at reduced data rates. There exist situations where ODT Enable mode can not be guaranteed for a short period of time, i.e., during power up.

Following are four terminologies defined for the state of a device (GDDR6 SGRAM or controller) signal during operation. The state of the bus will be determined by the combination of the device signal connected to the bus in the system. For example, in GDDR6 it is possible for the SGRAM pin to be tristated while the controller signal is HIGH or ODT. In both cases the bus would be HIGH if the ODT is enabled. For details on the device's signals and their function see Sections 9.1 and 9.2.

Device pin signal level:

- HIGH: A device signal is driving the Logic "1" state.
- LOW: A device signal is driving the Logic "0" state.
- Hi-Z: A device signal is tristate.
- ODT: A device signal terminates with ODT setting, which could be terminating or tristate depending on Mode Register setting.

Bus signal level:

- HIGH: One device on bus is HIGH and all other devices on bus are either ODT or Hi-Z. The voltage level on the bus would be nominally V_{DDQ}.
- LOW: One device on bus is Low and all other devices on bus are either ODT or Hi-Z. The voltage level on the bus would be nominally VOL(DC) if ODT was enabled, or V_{SS} if Hi-Z.
- Hi-Z: All devices on bus are Hi-Z. The voltage level on bus is undefined as the bus is floating.
- ODT: At least one device on bus is ODT and all others are Hi-Z. The voltage level on the bus would be nominally V_{DDQ}.

2.4 DEFINITION OF CLOCKING TERMINOLOGY

- Data refers to the signal being clocked (e.g. DQ by WCK and CA by CK
- Half rate: clock is running at half of the data rate (e.g. WCK 4GHz and DQ at 8Gbps, or CK 1GHz and CA at 2Gbps)
- Quarter rate: clock is running at a quarter of the data rate (e.g. WCK 2GHz and DQ at 8Gbps)
- Eighth rate: clock is running at one eighth of the data rate (e.g. WCK internal 1GHz and DQ at 8Gbps)
- DDR (Double Data Rate): complement to half rate, referring to data relative to clock
- QDR (Quad Data Rate): complement to quarter rate, referring to data relative to clock
- ODR (Octa Data Rate): complement to eighth rate, referring to data relative to clock

2.5 CLOCKING

The GDDR6 SGRAM supports two operating modes for WCK frequency which differ in the DQ/DBI_n pin to WCK clock frequency ratio. The GDDR6 SGRAM supports DDR and QDR operating modes for WCK frequency which differ in the DQ/DBI_n to WCK clock frequency ratio.

Figure 1 illustrates the difference between a DDR WCK and a QDR WCK. Figure 60 illustrates a WRITE command with a DDR WCK clock while Figure 61 illustrates a WRITE command with a QDR WCK clock. Figure 74 illustrates a READ command with DDR WCK clocking and Figure 75 illustrates a READ command with QDR WCK clocking. Other figures in the specification are shown only with the DDR WCK for simplicity unless otherwise noted.

GDDR6 SGRAM also supports 2 granularities for the WCK data clock in the device. GDDR6 SGRAM devices can be designed with either a WCK/byte or a WCK/word. The ball-out has provisions for a WCK/byte but also supports WCK/word with the unused WCK balls as NC; the host must turn the unused WCK off.

The DRAM info bits for WCK Granularity, WCK Frequency and Internal WCK can be read by the host during the initialization process to determine the WCK architecture for the device and for devices that support multiple frequencies, MR2 OP11 allows for the mode to be set. For the frequencies for each mode see Table 68.

In both WCK QDR and DDR modes the GDDR6 device operates from a differential clock CK_t and CK_c. Command and Address (CA) are registered at every rising and falling CK edge. For both WCK DDR and QDR ratio the GDDR6 device can support either a full data rate EDC or a half data rate EDC. See EDC section for more details.

A rising CK (or WCK) edge is defined as the crossing of the positive edge of CK_t (or WCK_t) and the negative edge of CK_c (or WCK_c). A falling CK (or WCK) edge is defined as the crossing of the negative edge of CK_t (or WCK_t) and the positive edge of CK_c (or WCK_c).

Table 1 — Example Clock and Interface Signal Frequency Relationship

<table>
<tr><th>PIN</th><th colspan="2">DDR WCK</th><th colspan="2">QDR WCK</th><th>UNIT</th></tr>
<tr><td>CK_t, CK_c</td><td colspan="2">1.5</td><td colspan="2">1.5</td><td>GHz</td></tr>
<tr><td>CA</td><td colspan="2">3.0</td><td colspan="2">3.0</td><td>Gbps/pin</td></tr>
<tr><td>WCK_t, WCK_c</td><td colspan="2">6.0</td><td colspan="2">3.0</td><td>GHz</td></tr>
<tr><td>DQ, DBI_n</td><td colspan="2">12.0</td><td colspan="2">12.0</td><td>Gbps/pin</td></tr>
<tr><td>EDC</td><td>6.0</td><td>12.0</td><td>6.0</td><td>12.0</td><td>Gbps/pin</td></tr>
</table>