Direct Memory Access

Direct memory access (DMA) is a method that allows an input/output (I/O) device to send or receive data directly to or from the main memory, bypassing the CPU to speed up memory operations.

The process is managed by a chip known as a DMA controller (DMAC).

A defined portion of memory is used to send data directly from a peripheral to the motherboard without involving the microprocessor, so that the process does not interfere with overall computer operation.

In older computers, four DMA channels were numbered 0, 1, 2 and 3. When the 16-bit industry standard architecture (ISA) expansion bus was introduced, channels 5, 6 and 7 were added.

ISA was a computer bus standard for IBM-compatible computers, allowing a device to initiate transactions (bus mastering) at a quicker speed. The ISA DMA controller has 8 DMA channels, each one of which associated with a 16-bit address and count registers.

ISA has since been replaced by accelerated graphics port (AGP) and peripheral component interconnect (PCI) expansion cards, which are much faster. Each DMA transfers approximately 2 MB of data per second.

A computer's system resource tools are used for communication between hardware and software. The four types of system resources are:

- I/O addresses.
- Memory addresses.
- Interrupt request numbers (IRQ).
- Direct memory access (DMA) channels.

DMA channels are used to communicate data between the peripheral device and the system memory. All four system resources rely on certain lines on a bus. Some lines on the bus are used for IRQs, some for addresses (the I/O addresses and the memory address) and some for DMA channels.

A DMA channel enables a device to transfer data without exposing the CPU to a work overload. Without the DMA channels, the CPU copies every piece of data using a peripheral bus from the I/O device. Using a peripheral bus occupies the CPU during the read/write process and does not allow other work to be performed until the operation is completed.

With DMA, the CPU can process other tasks while data transfer is being performed. The transfer of data is first initiated by the CPU. The data block can be transferred to and from memory by the DMAC in three ways.

In burst mode, the system bus is released only after the data transfer is completed. In cycle stealing mode, during the transfer of data between the DMA channel and I/O device, the system bus is relinquished for a few clock cycles so that the CPU can perform other tasks. When the data transfer is complete, the CPU receives an interrupt request from the DMA controller. In transparent mode, the DMAC can take charge of the system bus only when it is not required by the processor.

However, using a DMA controller might cause cache coherency problems. The data stored in RAM accessed by the DMA controller may not be updated with the correct cache data if the CPU is using external memory.

Solutions include flushing cache lines before starting outgoing DMA transfers, or performing a cache invalidation on incoming DMA transfers when external writes are signaled to the cache controller.