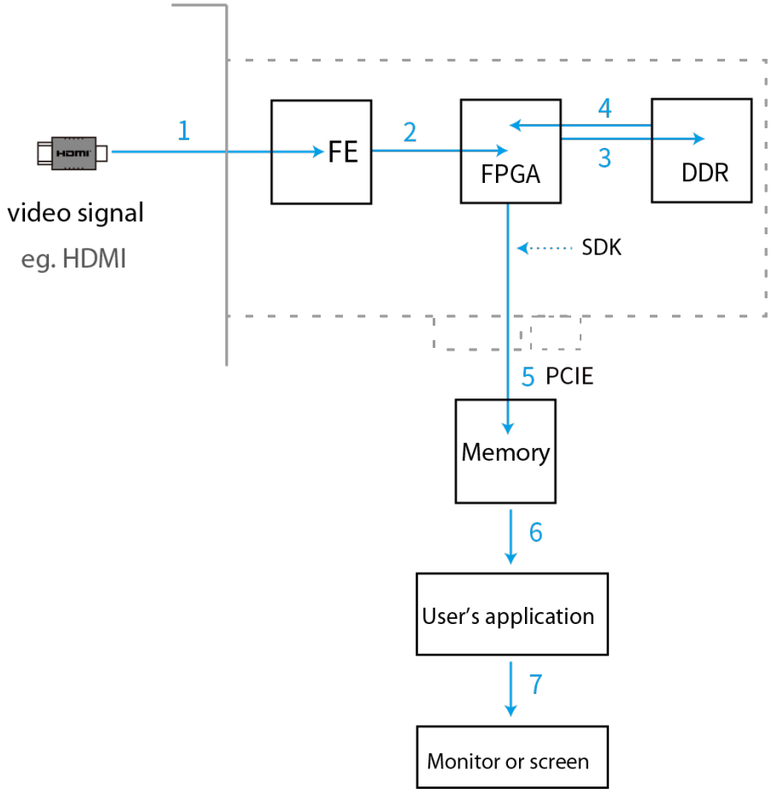
Low Latency & Partial Notification Function

Magewell devices has a low-latency mode which enables data transfer to the computer memory before one frame is completely stored in the on-board memory. To further reduce the total latency, users can integrate SDK with their application so that partial completion notification function can be used.

> Example: LowLatency
> Compatible hardware: Pro Capture Family
> Location: MWCaptureSDK\SDK\V3\Examples\Applications\LowLatency
> What can this example demonstrate:
> It demonstrates capturing video data using normal mode and low-latency mode with partial completion notification function. Users can see that low-latency mode can effectively reduce the capture latency while partial completion notification function reduces the processing latency caused by the user's application, further reducing the total latency.

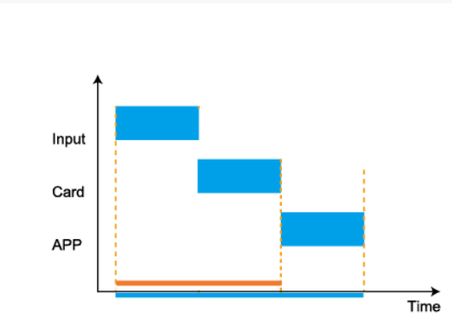
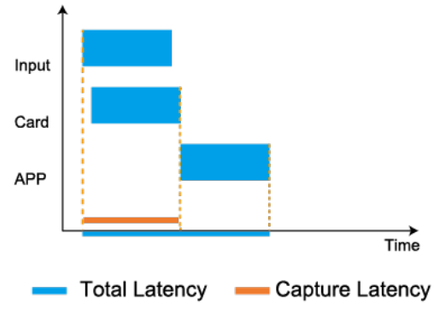


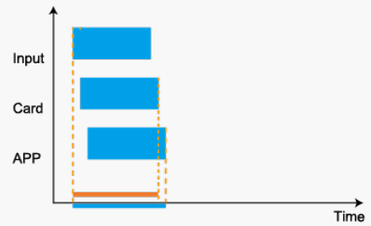
Definition of Different Latencies

- Buffer latency: The time for one frame to enter the card and to be stored in the on-board memory (DDR). See Step 1-3 in the flow chart.
- Transfer latency: The time for one frame to be transferred from FPGA and PCIe to the computer memory. See Step 4-5 in the flow chart.
- Capture latency: The time for one frame to enter the card, to be stored in the on-board memory and to be transferred to the computer memory. See Step 1-5 in the flow chart.
- Processing latency: The time for one frame to be obtained by user's application from the computer memory and use this frame for processing. See Step 6-7 in the flow chart.
- Total latency: The time from when one frame enters the capture card and until the user's utility finishes processing it. See Step 1-7 in the flow chart.

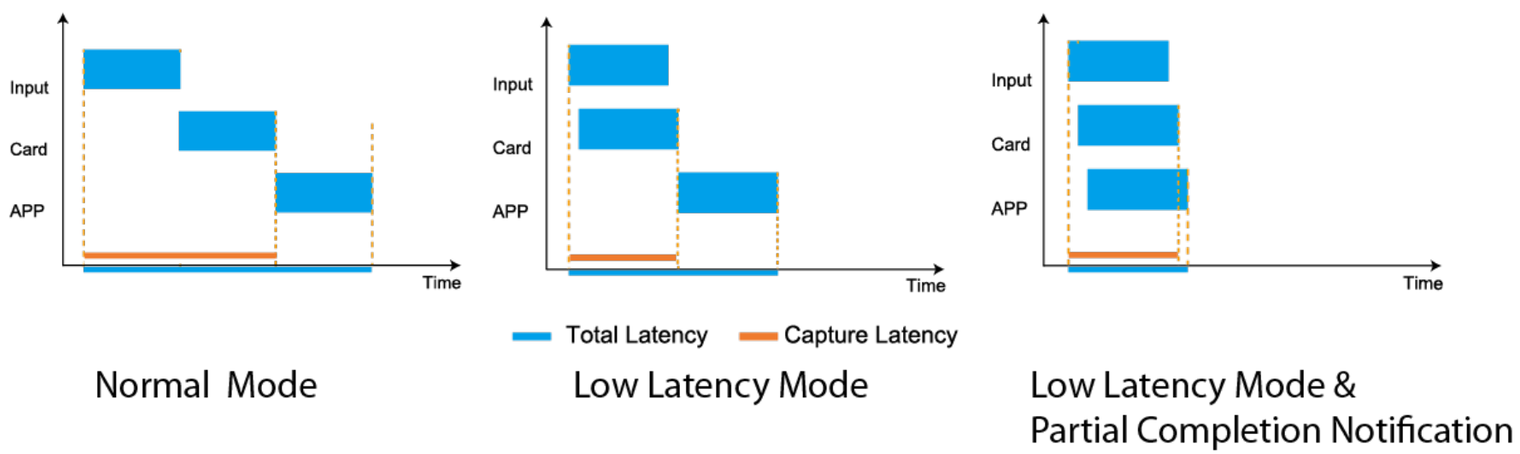
Three Transfer Modes

Magewell devices have three modes of data transfer, which are normal mode, low-latency mode and low-latency mode with partial completion function.

Mode Name	Principle	Diagram
Normal mode	1. One frame enters and is stored the buffer in the on-board memory. 2. The frame is transferred to the computer memory via PCIe. 3. User's application processes this frame.	
Low-latency mode	1. Some part of one frame, for example at least 64 lines, are transferred to the on-board memory and then will be immediately transferred to the computer memory via PCIe. 2. When one frame is fully stored in the on-board memory, the PCIe transfer of this frame will be almost finished at the same time. Due to factors like system coordination, there will be 1-2 ms delay. 3. User's application obtains this frame from the system memory and process it.	

Low-latency mode and partial completion notification mode	1. Some part of one frame, for example at least 64 lines, are transferred to the on-board memory and then will be immediately transferred to the computer memory via PCIe. 2. If the user's application has integrated MWCapture SDK of Magewell, partial completion event notification will be triggered after that part of data is transferred via PCIe. User's application can use this part of data for processing, for example GPU rendering. Compared to the first two models, user's application starts to process the video earlier. 3. In this mode, PCIe transfer and application processing start before one frame has been completely transferred. Therefore, it can effectively reduce the total latency for the users.	

Comparison of the three modes:



Theoretical Calculation

1. Buffer latency: when the frame rate of the input signal is 60 fps, the buffer latency (from one frame enters the card to it is completed stored in the on-board memory) is 1 / 60s = 16.7ms
2. Transfer latency (a frame is transferred from the on-board memory to the system memory):
   1920 x 1080 x 4 / 1024 / 1024 / 1600 x 1000 ≈ 4.94ms

The latency in real situation might be affected by the available bandwidth of the PCIe slot and the system coordination. The calculation result is for reference only.

Pro Capture HDMI 4K Plus Tested Results

*latency caused by application processing is not included

Mode Name	Video source	Capture latency	Processing latency	Total latency
Normal mode	1080P 60FPS RGB32	5.41ms	15.98ms	38.84ms
Low-latency mode and partial completion notification mode	1080P 60FPS RGB32	0.82ms	1.71ms	18.53ms

Processing latency: time needed for user's application to process one frame

Total latency: from one frame to enter the capture device to the frame is completed processed by the user's application

User's application used in this test: libjpeg-turbo. Use libjpeg to compress the captured image as jpeg files.