

How Intel® Agilex™ FPGA is Enabling Resource and Power Efficient 4K, 8K Video Processing Solutions

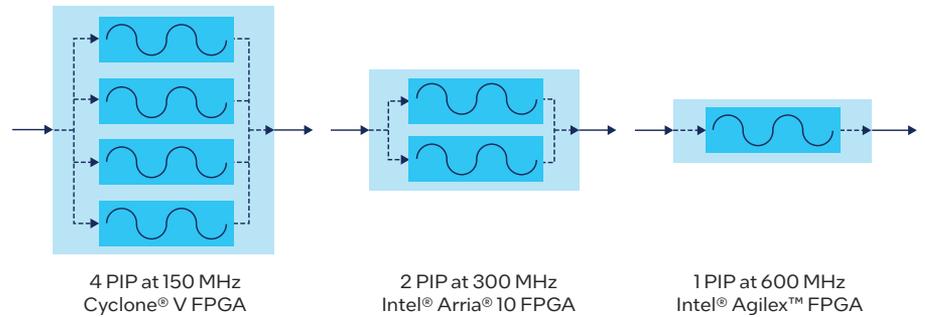
Authors Introduction

Neil Childs

Engineering Manager

Intel Programmable Solutions Group

The new Intel® Agilex™ family of 10 nm FPGAs is designed to operate at higher frequencies than previous families. This enables FPGA developers to minimize resource usage and power for a given logic function. The ability to reach 600 MHz, often without requiring extensive rewriting of existing register transfer level (RTL), is of particular interest to video designers as it enables “4K” video at 60 frames-per-second to be processed as 1 pixel-in-parallel (PIP).



The ability to run half the processing pipelines at double the clock frequency, compared to previous FPGA families, will become more important with the advent of “8K” video solutions. It means that video IP cores already capable of operating at 4 pixels-in-parallel will not require re-architecting with 8 pixels-in-parallel support for “8K”.

Table of Contents

- Introduction 1
- Video Clock Frequencies..... 1
- Real-World FPGA Clock Frequencies..... 2
- IP Core Resource Usage..... 2
- Power Savings..... 3
- Wider System Improvements..... 3
- Full System Case Study 3
 - Resource Usage 4
 - Pin Count 4
 - Device Selection 4
- Conclusions..... 5

Video Clock Frequencies

Video resolutions have increased over the decades from SD (720x486), through HD (1920x1080) and UHD “4K” (3840x2160) to “8K” (7,680x4,320) and beyond. The clock frequency required to handle this increasing bandwidth has likewise increased. The “pixel clock” for SD resolution video at 60 frames per second (fps) was a mere 13.5 MHz; easily accomplished today but challenging at its introduction in the early 90’s. High definition (HD) video resolutions required clock frequencies of 74.25 MHz or 148.5 MHz, which again were challenging but achievable for their era. Today’s “4K” resolution requires a pixel clock of 594 MHz, in excess of what FPGAs could realistically reach until very recently, while “8K” needs 2,376 MHz. These very high clock rates forced a different approach from video engineers.

When “4K”, or UHD TV, was first being widely developed around a decade ago, the typical FPGA families used were the Stratix® V FPGA or Intel® Arria® 10 FPGA families, which were not intended to reach 594 MHz. To cope with this limitation, video intellectual property (IP) cores such as scalars or color space converters, were redesigned to process multiple pixels on each clock cycle. In the majority of cases this meant duplicating the entire video pipeline within the IP core. Moving from 1 pixel-in-parallel to 2 pixels-in-parallel (PIP) for “4K” video could result in a

doubling of FPGA resources used. The current early adopters of “8K” video designs often rely on a similar technique of processing 8 pixels-in-parallel, with predictable increases in FPGA utilization.

An FPGA family such as Intel Agilex FPGA, which can truly handle real video designs at 594 MHz, therefore allows designers to halve the size of certain video IP cores when compared with previous families.

It should be noted that for video, unlike some other technology areas, resource utilization does not scale linearly with frequency. The device either reaches 594 MHz, or it does not. You either need 2X the resources, or you do not. An FPGA only capable of 550 MHz may allow extra headroom for routing and timing closure, but the video logic will likely remain clocked at 300 MHz. Such steps in clock frequencies do exist in other technology areas, for example PCIe has such steps at 62.5 MHz, 125 MHz, 250 MHz, and 500 MHz. Each frequency step you can achieve, halves the width of the datapath saving fabric and routing resources inside the FPGA.

It should further be noted that the resource saving is not uniform for all resource types. For some IP cores, such as a 3D LUT, halving the number of pixels processed in parallel will roughly halve all the resources required (typically adaptive logic modules (ALMs), M20Ks, and digital signal processors (DSPs)). However, for a core such as a video scaler, the linestore memories required for a vertical filter are required once per IP core regardless of the duplicated structures. Averaged over a whole design this means that the savings made in ALMs and DSPs are not completely matched by savings in M20Ks. While usage of these memory blocks will reduce, it is not typically on the same scale as other resource types. This leads to the interesting observation that faster FPGAs will benefit from a slightly different ratio of resource types, with more M20Ks being most useful.

Real-World FPGA Clock Frequencies

Few FPGA designs ever run close to the theoretical maximum frequency of the part used. To achieve anything close to the maximum, the designer would have to manually tune the RTL, effectively hand placing every DSP, M20K and ALM. Historically it would also mean keeping registers close together to minimize routing delays; long routing lines and high fanouts were a common cause of reduced overall performance. This issue led to the introduction of Intel® Hyperflex™ FPGA Architecture routing in Intel® Stratix® 10 FPGA, which alleviated the routing bottleneck on clock frequency, although often required extra pipelining registers to be added to the RTL.

Intel Agilex FPGAs include many changes intended to address common FPGA performance bottlenecks. The process gains of moving from 14 nm to 10 nm manufacturing, have enabled DSPs, M20Ks, and general FPGA fabric to run much closer to the maximum frequency for wider use cases. For example, provided the systolic registers are enabled using the chainin/chainout no longer causes the useable DSP frequency to decrease. It is now possible to run the DSPs at 676 MHz even in the slower -3V speed grade.

Second generation of Intel Hyperflex FPGA Architecture in Intel Agilex FPGAs includes an improved High-Speed Bypass Path, which improves default performance with RTL, which is not suitable for Intel Hyperflex FPGA Architecture. While the fine grain Hyper-Retiming allows the design tools to extract the fastest possible performance from the routing resources. These changes mean Intel Agilex FPGAs can deliver an up to 40% increase in core performance.

The overall effect of this is that many existing video IP cores originally designed to run at 300 MHz in an Intel Arria 10 FPGA can now comfortably run at 600 MHz in an Intel Agilex device with limited modification.

IP Core Resource Usage

The table shows the resources required for three video IP cores configured with either 2 or 1 pixel-in-parallel support sufficient to handle “4K” resolution video with a processing clock of 300 MHz or 600 MHz respectively.

	ALMs		M20Ks		DSPs	
3D LUT 2 PIP	2,464		222		12	
3D LUT 1 PIP	1,326	54%	111	50%	6	50%
Color space 2 PIP	711.4		0		12	
Color space 1 PIP	484.6	69%	0	-	6	50%
Scaler 2 PIP	3,713.8		54		48	
Scaler 1 PIP	2,134.8	58%	50	93%	24	50%
Tone Mapper 2 PIP	10,758		71		107	
Tone Mapper 1 PIP	7,504	70%	49	70%	58	55%
Warp 2 PIP	9,550.0		477		72	
Warp 1 PIP	5,767.1	61%	347	73%	36	50%
Average size of 1 PIP IP core compared against 2 PIP IP core		64%		68%		52%

Figures should be considered approximate and have been taken from the Intel® Quartus® Prime Pro Edition Software v21.2.

It can be clearly seen that doubling the processing clock frequency results in significant resource savings, particularly in ALM and DSP usage. When extrapolated over an entire design, such savings could easily mean a design fits in a smaller part or allows space for additional functionality.

Power Savings

The dynamic power required by a single register switching at 600 MHz is similar to two registers switching at 300 MHz as can clearly be seen by creating entries in the Intel FPGA Power and Thermal Calculator (PTC) tool.

Module	# Half ALMs	# FFs	Clock Freq. (MHz)	Toggle %	Routing Factor	Power (W)			User Comment
						Routing	Block	Total	
56	10000	0	600	25%	3	0.004	0.050	0.055	
57	20000	0	300	25%	3	0.004	0.050	0.055	
58	0	10000	600	25%	3	0.086	0.072	0.158	
59	0	20000	300	25%	3	0.086	0.072	0.158	

The reduction in static power, however, often outweighs this increase in dynamic power leading to a power advantage when switching to 600 MHz. While it is likely that the largest reduction in static power would be achieved by switching to a smaller FPGA, it is still possible to achieve meaningful static power reductions while remaining in the same part. For example, Intel Agilex devices support DSP and M20K power gating, so any resources saved in these areas will directly lead to static power reduction.

To demonstrate this, two test designs were constructed, each with four instances of the 3D LUT, Tone Mapper, and Warp IP cores. The two designs were configured for either 600 MHz 1 pixel-in-parallel (PIP) video data, or 300 MHz 2PIP video data. The reduction in resource usage is shown in the following table.

Resource	300 MHz Variant	600 MHz Variant	600 MHz as % of 300 MHz
ALMs	94,289 (19% of device)	61,907 (13% of device)	65.7%
M20Ks	2,294 (32%)	1,642 (23%)	71.6%
DSPs	724 (16%)	376 (8%)	51.9%

The power usage figures estimated by the PTC tool, assuming a 25% toggle rate and a constant 85 degree junction temperature are shown below. These figures clearly show an overall power reduction, with the 600 MHz design on the right, of around 0.5W.

Resource Type	Power (W)
Logic	3.466
RAM	1.760
DSP	0.866
Clock	1.215
PLL	0
IO	1.243
Transceiver	0.246
HPS	0
HBM	0
Miscellaneous	0.883
Static Power (Before Savings)	13.181
Static Power Savings	-5.183
SmartVID Power Savings	-1.751
Total Power	15.927

Resource Type	Power (W)
Logic	3.357
RAM	1.535
DSP	0.869
Clock	1.206
PLL	0
IO	1.244
Transceiver	0.246
HPS	0
HBM	0
Miscellaneous	0.883
Static Power (Before Savings)	12.917
Static Power Savings	-5.172
SmartVID Power Savings	-1.682
Total Power	15.403

These figures are taken from the PTC tool included with the Intel Quartus Prime Pro Edition Software v21.2 and should be considered approximate. Also, note that preliminary Intel Agilex FPGA power models have been used.

Wider System Improvements

Complete video systems typically include a wide range of functionality from other areas of technology, for example they often rely on embedded processors or interconnect such as PCIe for control, and external DDR memory for storage. Improvements in FPGA clock frequencies also enable progress to be made in these areas.

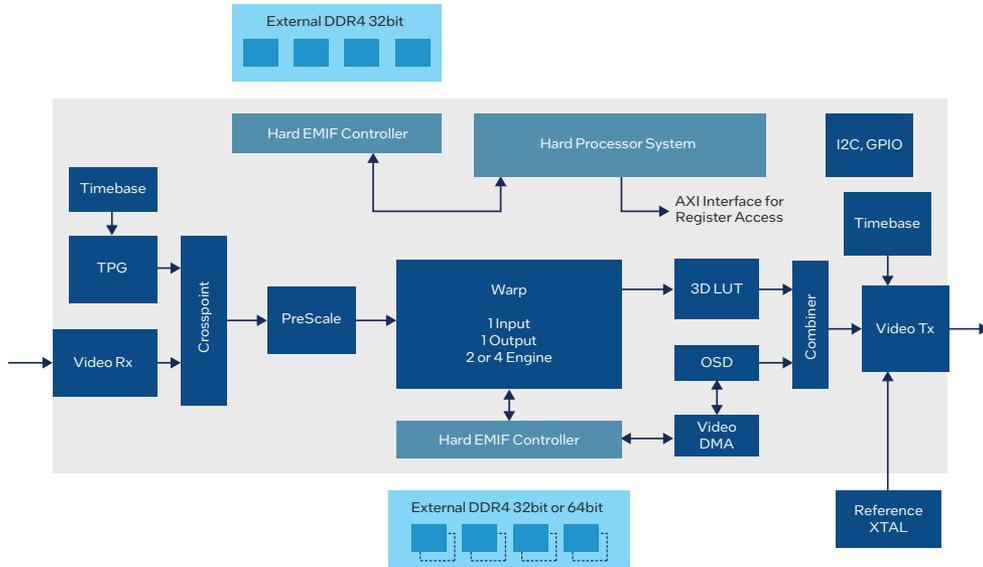
The ability to support the latest external memory standards and speeds are increasingly linked to the FPGA fabric speed. The internal interface is usually clocked at a quarter of the memory clock frequency (or an eighth of the headline DDR figure). For example, a 64 bit DIMM of DDR4 3,200 MHz will result in an internal interface that is 512 bits wide and clocked at 400 MHz. Future support for DDR5 4,400 MHz memory in Intel Agilex M-Series devices will require the FPGA fabric to support 550 MHz.

The same is true of the latest PCIe interfaces. For a number of years, designers have chosen to increase interface width rather than increase the clock frequency beyond 250 MHz. The ability of Intel Agilex FPGA fabric to comfortably meet 500 MHz effectively allows an upgrade from Gen4 to Gen5, or from 8 lanes to 16 lanes without increasing the interface width and consuming more routing resources.

Full System Case Study

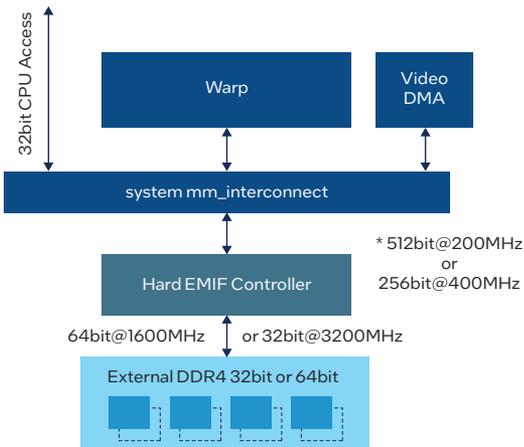
We have recently considered the design of a warp solution for a 4K120 projector, and specifically compared the options of running in an Intel Arria 10 FPGA at 300 MHz, or an Intel Agilex FPGA running at 600 MHz.

Control was to be handled by an embedded processor, which would also compute the warp mesh required, a mathematically complex task. For this reason, an SoC was chosen and external communication would be via Ethernet.



The external video interfaces were specified to be HDMI at the input, and V-by-One at the output. To handle 4K120 V-by-One requires 16 transceiver lanes, each running at a fixed line rate of 3 Gbps. Genlock and alternative resolution video were to be handled by adjusting horizontal and vertical blanking periods as required, simplifying the clocking of the output video interface.

Internally, the warp processor is required to process 1,200*106 pixels per second. This requires two processing engines running at 600 MHz, or four processing engines running at 300 MHz.



The memory bandwidth required for a 4K120 bounce through memory equates to 66.4 Gbps (3,840*2,250*120*32*2). The design also required a 1080p overlay to be read from external memory, adding 8.3Gbps (1,920x1,125x32x120), for a total memory bandwidth of 74.7 Gbps. This was considered too high to be comfortably accommodated by a 32 bit DDR 2,400 MHz interface (76.8 Gbps). As a result, the Intel Arria 10 FPGA variant required a 64 bit memory interface at 1,600 MHz (102.4 Gbps), whereas the Intel Agilex FPGA could stick with a 32 bit interface and instead use faster 3,200 MHz memory (also 102.4 Gbps). The internal memory interface is therefore 512 bits at 200 MHz for Intel Arria 10 FPGA versus 256 bits at 400 MHz for Intel Agilex FPGA.

Resource Usage

Module	300 MHz			600 MHz		
	ALMs	M20Ks	DSPs	ALMs	M20Ks	DSPs
PreScaler	7,735	58	96	3,733	54	48
Warp	16,590	750	144	9,502	477	72
3D LUT	2,464	222	12	1,326	111	6
Video DMA	2,699	36	0	2,405	21	0
VideoIO	5,000	20	0	5,000	20	0
Miscellaneous	6,000	10	6	5,000	6	3
Total	40,488		258	26,966	689	129

Figures for video I/O and miscellaneous are estimates for illustration

Pin Count

The faster Intel Agilex FPGA I/O allows the use of a narrower DDR4 3,200 MHz memory interface. This saves 44 FPGA pins that has several advantages including, reduced I/O power consumption, potential for smaller device package, fewer external memory components, and simpler PCB layout and routing.

Function	I/O	Transceiver (XCVR)
Clock / reset / HPS (flash, ethernet, uart)	~25	
Video I/O	~10	4 RX + 16 TX
CPU memory – DDR4 1,600 MHz * 32 bit	75	
Video memory – DDR4 1,600 MHz * 64 bit or 3,200 MHz * 32 bit	75 or 119	

Device Selection

For the 300 MHz Intel Arria 10 FPGA variant, the number of M20Ks required means that the SX480 is the smallest possible device, even though this provides many more ALMs and pins than required. If the Intel Arria 10 FPGA family was capable running these video IP cores at 600 MHz, then it would have been possible to move two devices smaller to the SX270 device, which is also available in a smaller package. This solution would obviously have offered significant cost advantages and potentially power savings.

Either design would easily fit in what is the smallest AGF006 Intel Agilex device but a smaller design, even though running faster, is still likely to offer a power reduction as the reduction in static power usually outweighs the slight increase in dynamic power. Future “8K” video designs will likely come closer to filling Intel Agilex devices, in which case the resource saving offered by 600 MHz operation is again likely to mean selecting a smaller device.

Conclusions

- Intel Agilex FPGA is the first FPGA to comfortably reach 600 MHz, enabling integration of complex video systems at this important clock frequency
- This performance is “push-button”, delivered with only limited changes to RTL
- 600 MHz reduces resource count and increases value by fitting into a smaller device, yielding significant static power savings
- 8K video will drive future adoption of 600 MHz processing as larger designs mean larger savings



Intel technologies may require enabled hardware, software or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates.

Some results have been estimated or simulated.

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.