Prototyping Imagination’s PowerVR Series6XT Dual-Cluster 64-Core GPU

Imagination’s approach to graphics processor unit (GPU) development is to provide scalable IP that can support the demands of a broad range of applications, from wearables and mobile devices to high-end gaming and computing. Satisfying the requirements of all of these markets is only possible by creating architectures that can deliver the highest levels of performance while minimizing both area and power.

Power to Perform

The PowerVR® Rogue architecture has helped Imagination take a market-leading position — measured in terms of GFLOPS/mm² and GFLOPS/mW — over several generations of GPUs. In order to achieve the highest levels of performance, GPU architectures harness significant parallel processing power to perform the most demanding graphics and GPU compute tasks. Consequently, configuring GPU IP to deliver advanced processing features and high performance can result in complex, high gate-count devices — often substantially larger than an SoC’s application processor IP. Across the PowerVR Series6 GPU cores, the largest GPU configuration is eight times that of the smallest GPU (Figure 1).

A key challenge faced by Imagination’s design team is how to test GPUs as they become larger and more complex. Historically, use of FPGA-based prototyping has been limited as top-end GPUs exceed the capacity of the largest available FPGA devices and the manual partitioning of the data-path intensive GPU structures to multiple FPGAs has proven to be a time-consuming and difficult process. As a result, the only feasible approach has been to fabricate test chips — an increasingly expensive and time-consuming process, which delays final product lead times.

*PowerVR® is a registered trademark of Imagination Technologies

Figure 1: Imagination’s PowerVR architecture (image provided at the courtesy of Imagination Technologies)

About the Author

Andy Jolley is Senior Staff Application Consultant – Worldwide Product Line Lead, FPGA-Based Prototyping at Synopsys. Andy has been working with FPGA technologies for over 25 years, originally in a design capacity in the telecommunications, radar and video industries before supporting FPGA synthesis and prototyping technologies at Synplicity and then Synopsys. Most recently, Andy has been supporting UK customers with their complex CPU SoC and GPU IP prototyping needs on the Synopsys HAPS platforms while also providing support for worldwide engagements to deploy the same SoC and GPU IPs embedded into user applications. Andy holds a Bachelor’s Degree in Electronic Engineering from the University of Brighton, England.
The size and complexity of GPU IP is not just an issue for Imagination’s GPU development teams; its customers face similar testing challenges when it comes to integrating the IP into a system on chip and partitioning it into multiple FPGAs for prototyping purposes.

**GPU Prototyping Requirements**

Synopsys has been working with Imagination’s design teams to explore ways of using multiple FPGAs to model their largest GPUs. An FPGA-based prototyping methodology must be able to support all configurations of the entire GPU family for standalone implementation and testing, as well as rapid GPU integration for SoC development.

The Synopsys team first worked on a proof-of-concept project to demonstrate an FPGA-based prototype for Imagination’s PowerVR Series6 devices. The prototyping environment required a top-level test infrastructure (Figure 2) to enable standalone regression tests to be run. The test infrastructure had to support connection to a PC host via PCIe, and a DDR3 memory interface to support the storage of test stimuli and results. The test interface required extensive use of rate conversion logic. The test infrastructure would enable the test team to control and analyze the GPU, including the ability to configure the system through a Universal Multi-Resource BUS (UMRBus) and access the test and results data from the PC host.

The team manually partitioned the design for implementation on a Synopsys HAPS-70 S48 prototyping system comprising four FPGAs. To successfully partition the largest GPU configuration took two weeks of manual design effort, iterating the partitioning to trade off the I/O multiplexing.

**New Developments**

While the proof-of-concept project was underway to prototype the PowerVR Series6 GPUs with Synopsys HAPS prototyping boards and Certify software, Imagination’s design team was busy working on the next generation: PowerVR Series6XT GPU family.

FPGA utilization varied between 55% and 90%, and the prototype achieved a clock speed of 8 MHz, which enabled successful completion of 7,000 regression tests – all without the need to implement a test chip (Figure 3).
The Series6XT GPUs offer processing performance that scales to the TFLOPS range, thanks to use of significant parallelism within the GPU implementation. The Imagination design team wanted to see if prototyping could handle the largest Series6XT device.

The challenge was not only to partition the derivative design, which was an even larger device than the Series6 GPU, but also to create additional test logic and sufficient performance to enable output of live video as part of the test.

Initial estimates of the test logic and GPU top-level design suggested that these blocks alone would exceed 100% utilization of a single Virtex 7 2000T FPGA, hence the need to repartition the original Series6 prototype design.

To compound the engineering challenge, the Imagination team needed some kind of prototype as soon as possible; reducing time to first prototype was absolutely key.

The Synopsys prototyping team considered two possible development routes. The first option was to repartition the existing HAPS-70 48 system. While this was perfectly possible, it required a muxing ratio of 32:1, which would reduce the system performance to 2 MHz – a significant compromise in terms of run times and too slow to support analysis of live video output.

While Imagination had been developing its next generation GPU, Synopsys had also been working on a second generation of automated FPGA partitioning tools. ProtoCompiler is designed to minimize the effort and time required to bring-up and then deploy a Synopsys HAPS series system for IP validation and software development. It incorporates automation features for design planning, logic synthesis, debug, and connectivity to other verification environments like Synopsys VCS and ZeBu. The prototyping software is tightly integrated with the HAPS series to deliver system performance.

The Synopsys team captured the additional test infrastructure logic, which included a frame buffer and logic for the DriverLive Active Video Out, and initially allocated six FPGAs to the prototyping environment. The team also ported the basic partitioning constraints from the previous PowerVR Series6 configuration to seed the repartitioning process. The team used other constraints to ensure there was sufficient FPGA capacity, to limit FPGA utilization to 80% and select a simple pin-muxing strategy.

The prototyping team made use of abstraction flows (Figure 4) to investigate FPGA-to-FPGA interconnect — a feature within ProtoCompiler that enables the effects of various partitions to be explored very quickly. Typically scenarios can be investigated in runtimes of a minute or so.

ProtoCompiler’s abstraction flows feature enabled the prototyping team to identify communication bottlenecks between FPGAs and quickly analyze the effect of increasing the interconnect capacity on the hardware. One of the key features of the HAPS environment is that the prototype configuration is not constrained by fixed I/O, a feature that ProtoCompiler is able to exploit. The team concluded that constraining the multiplexing ratio to 12 would enable the prototype to run at a respectable 7.3 MHz.

The additional logic to support the DriverLive Active Video output was merged with the existing test logic, which had no impact on the overall FPGA partitioning. The prototype design can easily be modified as ProtoCompiler supports incremental updates without having to re-compile the entire design. Adding the live video output feature gives the Imagination design team a very powerful debug capability — the ability to review moving images in real time at the GPU prototyping stage.

Figure 4: Abstraction flows used to investigate FPGA-to-FPGA interconnect
With the prototyping design mapped to just five of the six available FPGAs, the team took the opportunity to optimize for performance. High-speed time-domain multiplexing (HSTDM) I/O sharing increases FPGA interconnection bandwidth. All HAPS-70 connectors support HSTDM, while ProtoCompiler automates the complete HSTDM implementation, including assigning source synchronous clocks, splitting multi-terminal nets for point-to-point HSTDM communication and handling direct and HSTDM net assignment to the HAPS platform. By using an HSTDM ratio of 24x2 the team improved the performance of the prototype to 12.0 MHz. A single engineer spent around half a day to explore the prototype, and a further half-day to implement it.

**Compelling Results**

The collaboration between Imagination and Synopsys showed how, with the right environment, design teams can use FPGAs to support early prototyping of the largest and most complex GPU devices.

The success of both the PowerVR Series 6 and Series 6XT prototypes will mean that the Imagination design team is less dependent on the use of test chips to bring new GPUs to market. Early access to physical prototypes enables system validation, earlier software development and eases hardware-software integration.

The 12 MHz performance achieved with the prototyping platform helped Imagination to execute thousands of tests in a matter of hours and provided a platform for early software development. In addition, supporting the video output from the HAPS system allowed the use of real-time, real-world I/O to enable inspection of the correctness and quality of the image processing.

The ability to accelerate the time to first prototype assists Imagination in bringing new products to market, and also helps Synopsys’ and Imagination’s mutual customers when they integrate Imagination GPUs into their SoCs. Synopsys is working with Imagination to investigate the feasibility of adding extensive configurability to the test infrastructure. This will enable mutual customers to more easily integrate Imagination GPUs within Synopsys design and verification flows.

**More Information**

- Synopsys FPGA-Based Prototyping
- Imagination Technologies