By BRAD GRIFFIN, Product Marketing Group Director,
Systems Analysis Group, Cadence Design Systems

Electronic Design

Conquer Signal and Power Challenges When Implementing GDDR6 Interfaces

The GDDR interface has evolved beyond the original applications for gaming GPUs. However, designing GDDR6 systems is fraught with challenges, thus requiring advanced tools for analysis/workflow to handle the exhaustive signoff process.

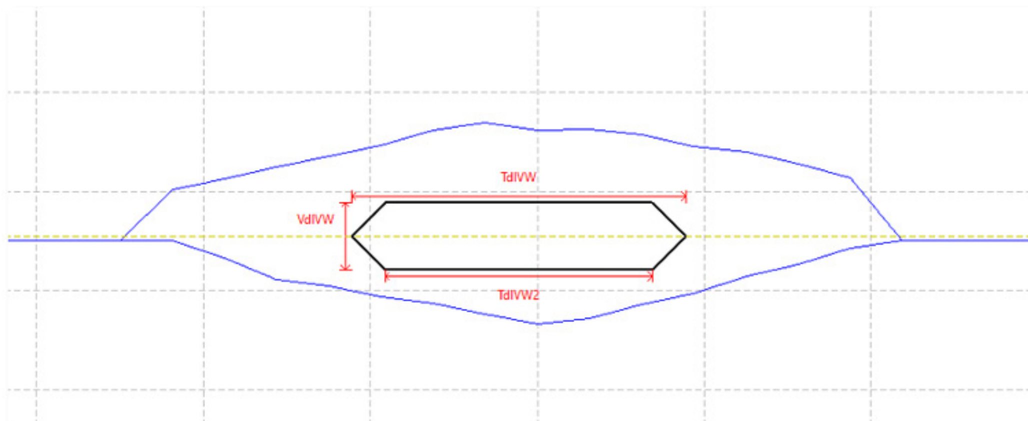
Graphics processing units (GPUs) and graphics double-data-rate (GDDR) memory interfaces are essential to graphics cards, game consoles, high-performance computing (HPC), and machine-learning applications. Signal integrity (SI) and power integrity (PI) are becoming intertwined with the thermal issues caused by ultra-fast data-transfer rates, ultra-low-voltage swings, and high-density GDDR6 designs that are often implemented on silicon interposers.

This article describes how power-aware SI analysis and thermal-aware PI analysis are used as part of a system design and signoff methodology for GDDR6 designs. Such designs enable data-transfer speeds of over 665 GB/s today and will continue to support such speeds of well over a terabyte per second (TB/s) in next-generation GDDR interfaces.

What is GDDR6?

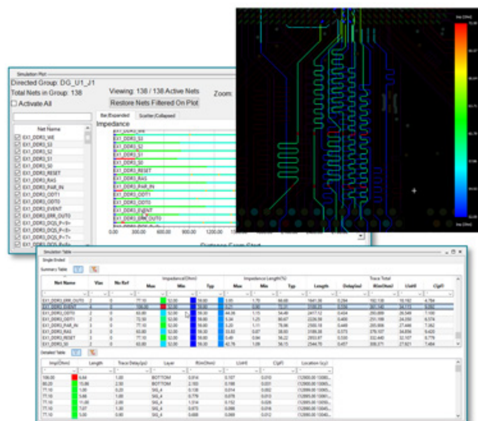
GDDR is an evolving interface used for many electronic applications. It's typically used in graphics-intense applications such as gaming consoles and graphic cards. However, usage of GPUs has progressed to other applications today because of the high bandwidth, enabling high-performance applications such as machine learning (ML), artificial intelligence (AI), graphics, advanced driver-assistance systems (ADAS) for automobiles, and high-performance computing (HPC).

As more companies invest in heterogeneous integration of semiconductors to address these next-generation applications featuring GDDR6 interfaces, they look to partner with an intellectual-property (IP) provider for the GDDR6 memory controller.

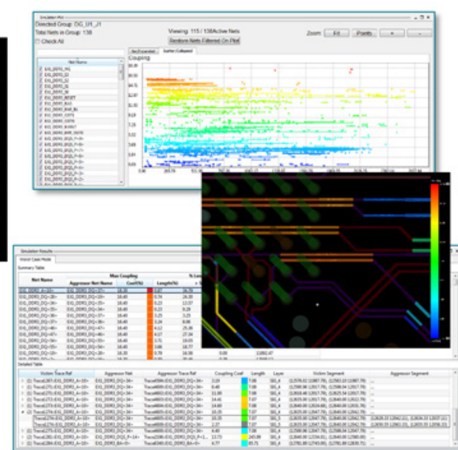


1. GDDR6 compliance is measured with mask requirements.

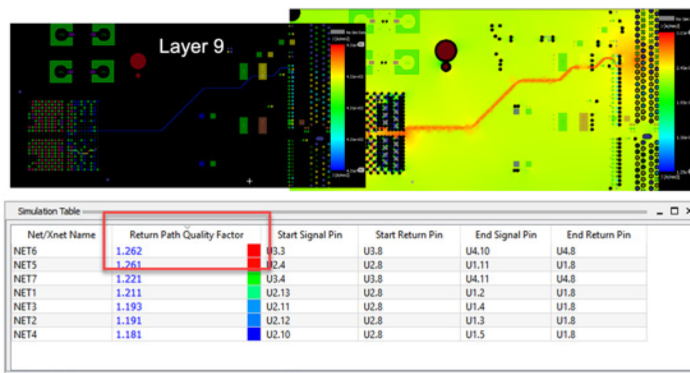
Impedance Analysis Screening



Coupling Analysis Screening



Return Path Screening



2. Signal-quality in-design analysis.

Key Issues to Consider in GDDR6 Design

As DDR interfaces have evolved, setup-and-hold checks are no longer the sole approach used in checking the interface timing compliance. While still part of the process, timing is now managed through various mask requirements. *Figure 1* shows a mask used for the data bus. There are multiple mask checks to examine the relationship between address, control, and data, and the various strobes/clocks used to latch the signals both at the rising and falling edge of the clock/strobe.

One of the changes with GDDR6 compared to previous versions of GDDR is that a data inversion now exists for both the data bus (DBI) and command/address bit inversion (CABI), which reduce the number of signals that need to switch simultaneously. This results in a reduction in simultaneous switching noise (SSN), which in turn cuts down on bit errors, ensuring that GDDR6 interfaces achieve the required bit error rate (BER).

Accurate simulation of the latest GDDR data-transfer speeds requires that both the controller and memory devices support modeling of feed-forward equalization (FFE), continuous-time linear equalization (CTLE), and decision feedback equalization (DFE) with input/output buffer information specification (IBIS) algorithmic-modeling-interface (AMI) models.

In-Design Analysis

GDDR6 design requires an exhaustive signoff process, but design teams can advance through the signoff stage faster by using in-design workflows early in the design process. In-design analysis complements a constraint-driven flow.

Constraints provide rules that designers must follow, but in-design analysis delivers simulation engines to the designer while they're in the process of layout out their design. This enables design teams to find problems earlier in the design cycle, which in turn means less time spent in the exhaustive signoff analysis stage.

Designers of memory interfaces like GDDR6 must pay close attention to both impedance discontinuities and return-path quality. *Figure 2* illustrates the electrical-rule-check (ERC) screening technologies that can be performed during the PCB design.

Impedance analysis helps designers identify outliers such as traces routed over splits in power/ground planes *(Fig. 2, top left)*. Coupling also can be analyzed, providing early insight into potential crosstalk problems without having to uti-lize IBIS models *(Fig. 2, top right*. Finally, the return path can be analyzed to uncover nets with possible return-path problems using the return-path quality factor as a figure of merit. Selected signals highlight the return current flow on the planes so that the return path quality can be visualized and changed directly within the PCB design environment *(Fig. 2, bottom)*.

Reflection analysis output

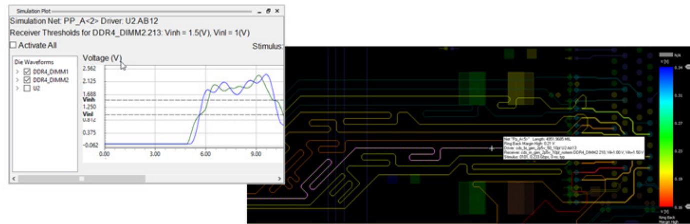
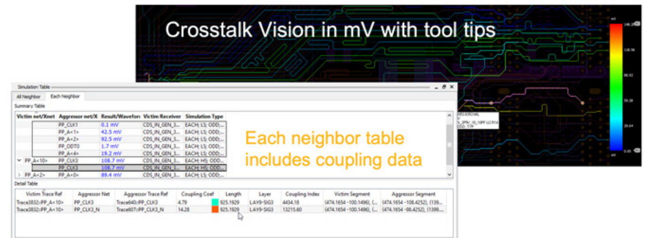
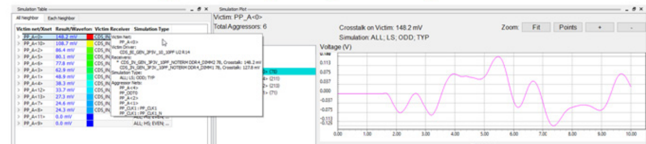


Table supports cross-probing, ability to show details, and .CSV output

Crosstalk analysis output



Both tables support cross-probing and ability to show details



3. Signal-integrity and crosstalk in-design analysis.

For more sophisticated analysis. designers can call on reflection and crosstalk workflows that enable viewing of time-domain waveforms. Simulation is enabled by industry-standard IBIS models that are assigned to the components in the implementation environment. With those IBIS models in place to model the I/O, reflection and crosstalk analyses reveal the overall signal quality *(Fig. 3)*.

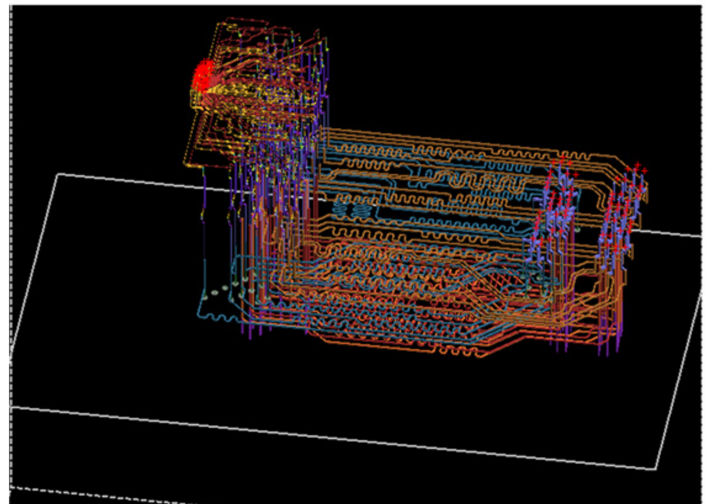
SI Testbench for GDDR6 Memory Interface

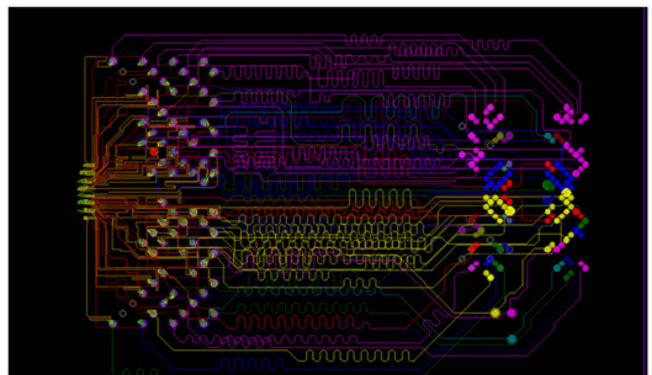
One of the key steps in GDDR6 design is extracting the interconnect from the PCB and IC package. There are different approaches to modeling system interconnect using Cadence tools: the Sigrity X PowerSI hybrid solver, the full-wave 3D finite-element-method (FEM) Clarity 3D Solver, or a combination called the “cut-and-stitch flow.” Companies with access to both PCB and IC package design data can merge the PCB and the package to perform an extraction that will include the coupling between the IC package and PCB *(Fig. 4)*.

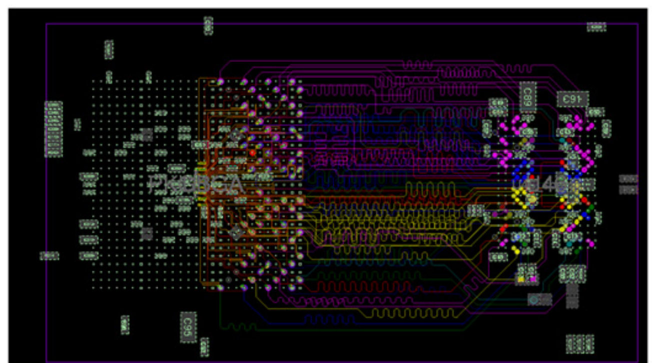
PCB design teams without access to the package design should ask their semiconductor vendors for package models of the memory controller and the memory. These can be cascaded with the extracted PCB model to perform system-analysis design.

If the design is implemented using the Cadence PCB or package design tools, the design data, including the stackup and material properties, will all be automatically read into the extraction tool without needing any manual inputs. Designs from non-Cadence design tools are read in through IPC-2581 or ODB++ manufacturing file formats.

Clarity 3D Solver is a robust FEM solver offering parallelization and distributed computing technologies that allow for full-wave 3D extraction of large,



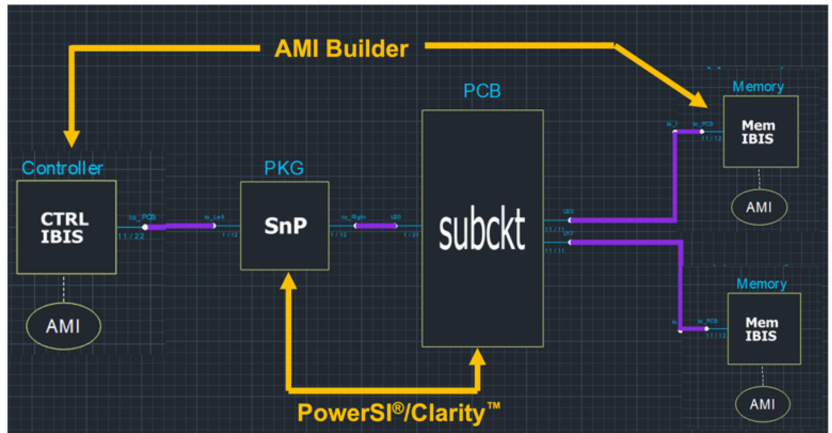




4. The IC package and PCB can combine into a single environment for extraction.

complex designs done in a fraction of the time as legacy FEM tools. Distributed computing technologies also have been incorporated into the Sigrity X PowerSI tool to quickly extract S-parameters using a hybrid solver.

In this GDDR6 example, PowerSI provided a full model in 15 minutes while the Clarity 3D Solver solved the same problem with a full 3D FEM approach in 1.5 days using 64 cores. The cut-and-stitch methodology blended the two approaches for improved accuracy over the hybrid solver alone, providing results in hours.



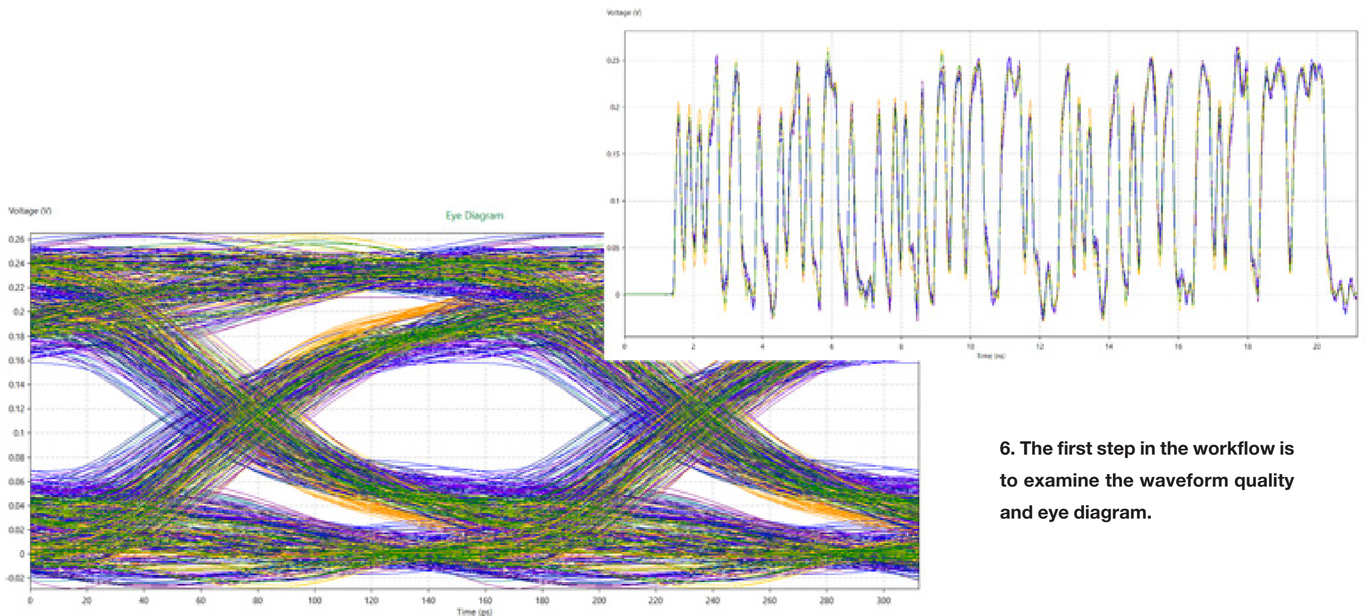
5. The topology workbench utilizing the SystemSI parallel-bus-analysis workflow.

Which extraction technology should be used and when? As a first step, PowerSI hybrid extraction can be used to obtain a quick extraction to take a first-pass look at the S-parameters. Designers can observe the insertion loss, return loss, and fundamental frequencies and determine if modifications need to be made to the layout.

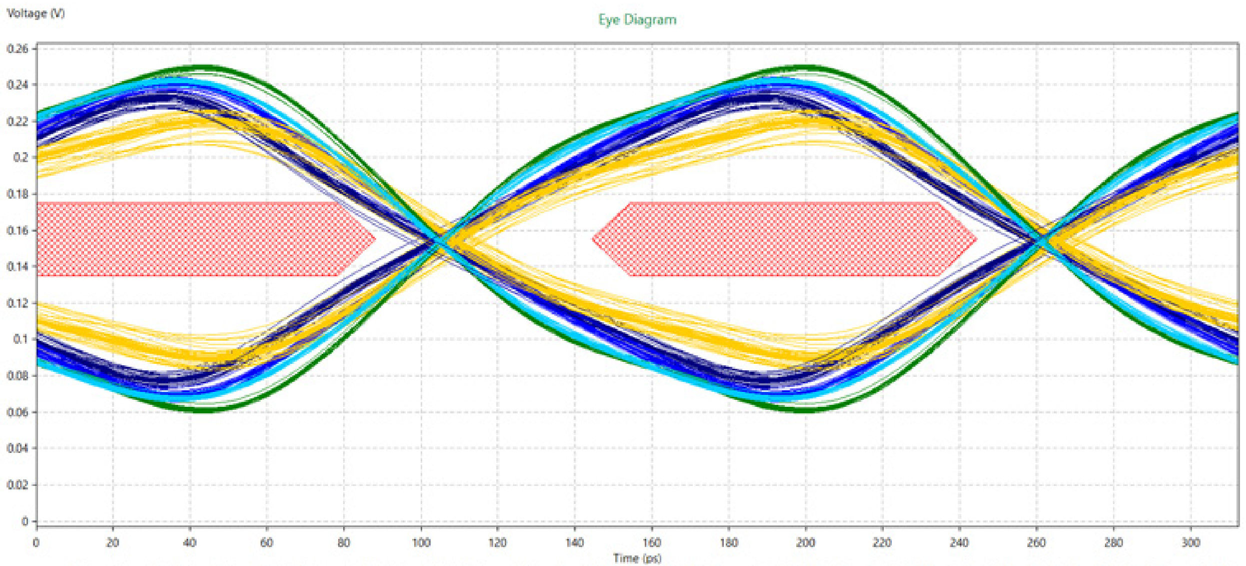
For signoff, Clarity 3D FEM extraction should be used. Between initial modeling with PowerSI and final signoff, the cut-and-stitch flow can be employed to focus on areas of interest around the interface and assign extraction techniques to each area. Areas with vias, via-stubs, non-uniform return paths, etc. may be assigned to the Clarity 3D Solver, whereas areas with long transmission lines that are well behaved and have a uniform return path would be assigned to the hybrid solver.

As the design matures, these two solvers can be combined to provide fast and reasonably accurate results in a fraction of the time of a full 3D FEM extraction. Moreover, they serve as a balance between accuracy and extraction time.

The scalability of the Clarity 3D Solver provides an alternative way to accelerate the signoff process. With additional CPU cores made available, the GDDR6 example that was completed in 1.5 days could have been accelerated. Doubling the number of compute cores will typically cut the simulation time nearly in half. In addition, the process is memory efficient—32-core machines with 256 GB of memory are more than sufficient to complete the task.



6. The first step in the workflow is to examine the waveform quality and eye diagram.



7. Check to make sure the eye mask fits.

If there's a lack of compute power on the premises, designers could turn to the Clarity 3D Solver Cloud solution. With Clarity 3D Solver Cloud, all of the setup is done on the local machine and the simulation is sent to a secure AWS cloud. Results are returned to the on-premises machine just as if the simulation had been run locally or on an on-premises server farm.

Parallel bus topology creation

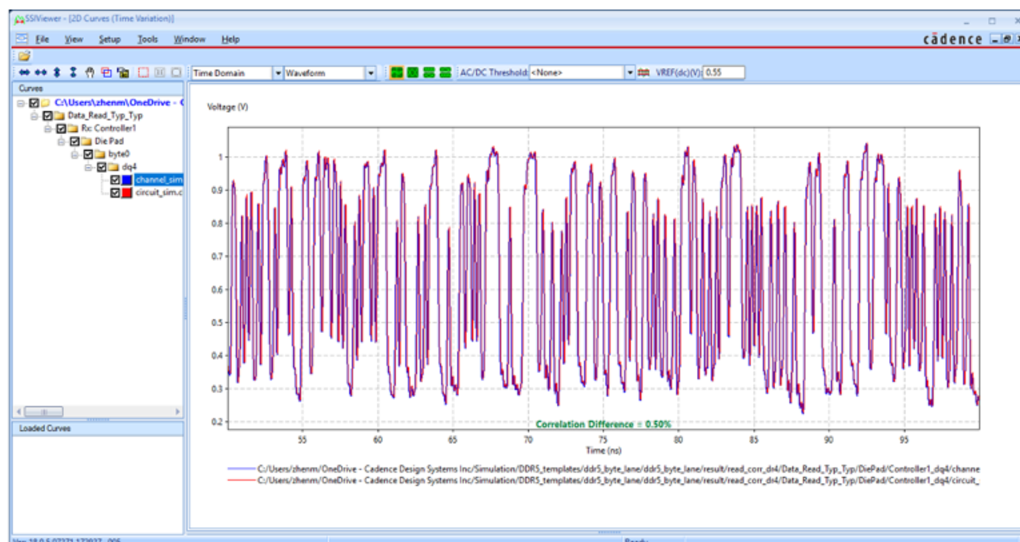
After the extraction is performed, the topology is created using the topology workbench within the Sigrity X SystemSI parallel-bus-analysis (PBA) tool *(Fig. 5)*. This GDDR6 shows the block-based topology where the designer places blocks in the different portions of the interface.

Figure 5 shows an on-board memory example with IBIS models for the controller and two memory devices. Between the IBIS models are separate interconnect models for the IC package and the PCB. As mentioned above, an alternative approach would be to combine the package and PCB into a single interconnect model so that the electromagnetic impact between IC package and PCB will be captured.

Also note that IBIS AMI bubbles have been attached to the IBIS models. All of the blocks are connected—the purple connectors represent multiple connections between the blocks (only one line is visible but there are multiple connections). Once the connections are made, the analysis can be started.

Traditionally, channel simulations have been done on serial links and more specifically with differential pairs. Cadence has developed and patented channel simulation on single-ended signals that enables designers to simulate the millions of bits required for interfaces like GDDR6 that include both SSN and crosstalk effects.

A correlation of the circuit simulation with the channel simulation can easily be performed within the workflow, either using the default options or rising ramp responses, falling ramp responses, or both rising and falling can be consid-

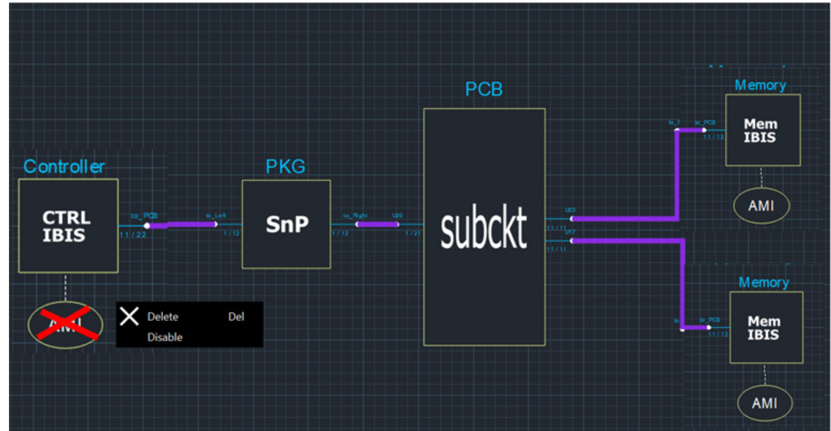


8. The circuit/channel correlation simulation in the workflow.

ered. The workflow steps for this flow are unique for signoff of high-speed memory interfaces.

The IBIS AMI portion of the channel simulation requires models that ideally will come from the component vendor. When the vendor is unable to supply an IBIS AMI model, a wizard-based workflow allows you to create an IBIS AMI model based off the component specification.

All of the steps in the parallel bus SystemSI PBA environment are workflow-based. There are workflows for setup, simulation, and evaluating the results. The workflows support different types of analysis, including circuit simulation and channel simulation, and the ability to incorporate those with a bus simulation. Lastly, there's an intuitive simulation parameter setup for stimulus definition and model selection.



9. Channel simulation and AMI models.

Workflow Steps

The first step of the workflow is to run a circuit simulation that examines the waveform quality to ensure there are no issues with the eye diagram *(Fig. 6)*.

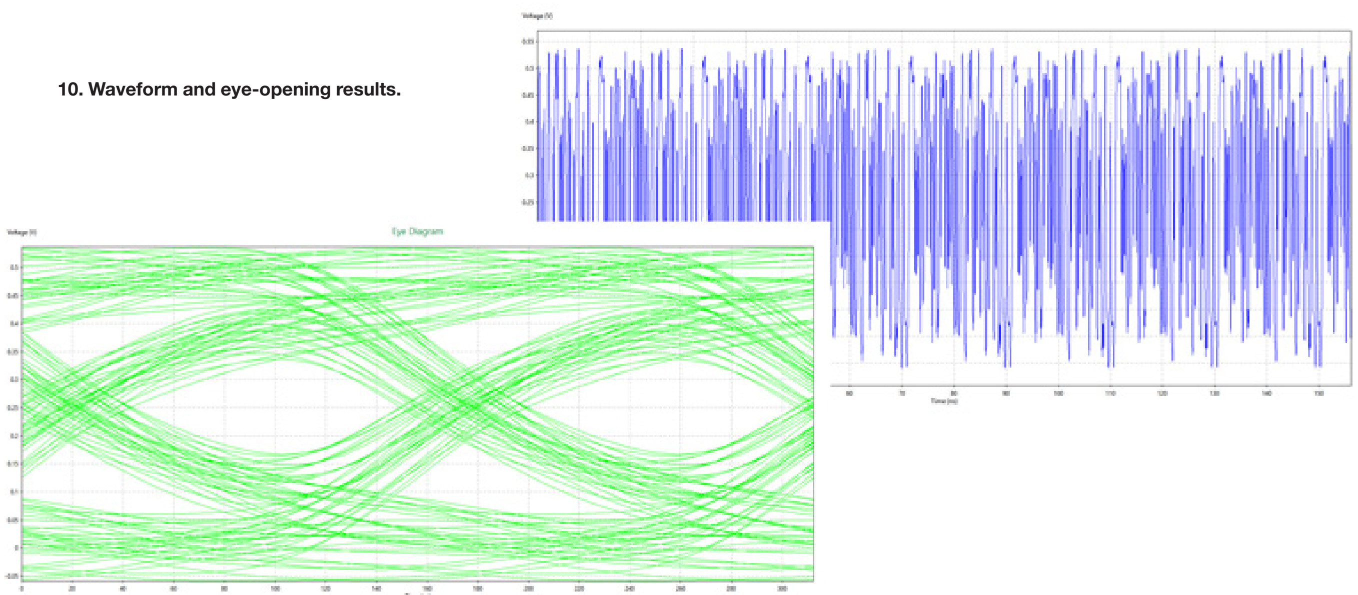
The bus responses from the circuit simulation can be examined and the eye mask checked to make sure it fits *(Fig. 7)*.

After running the initial simulation, the next step is to run a circuit/channel correlation. This also is easily done in the workflow, with the tool doing all of the setup and simulations automatically and providing a view of the overlay for comparison *(Fig. 8)*.

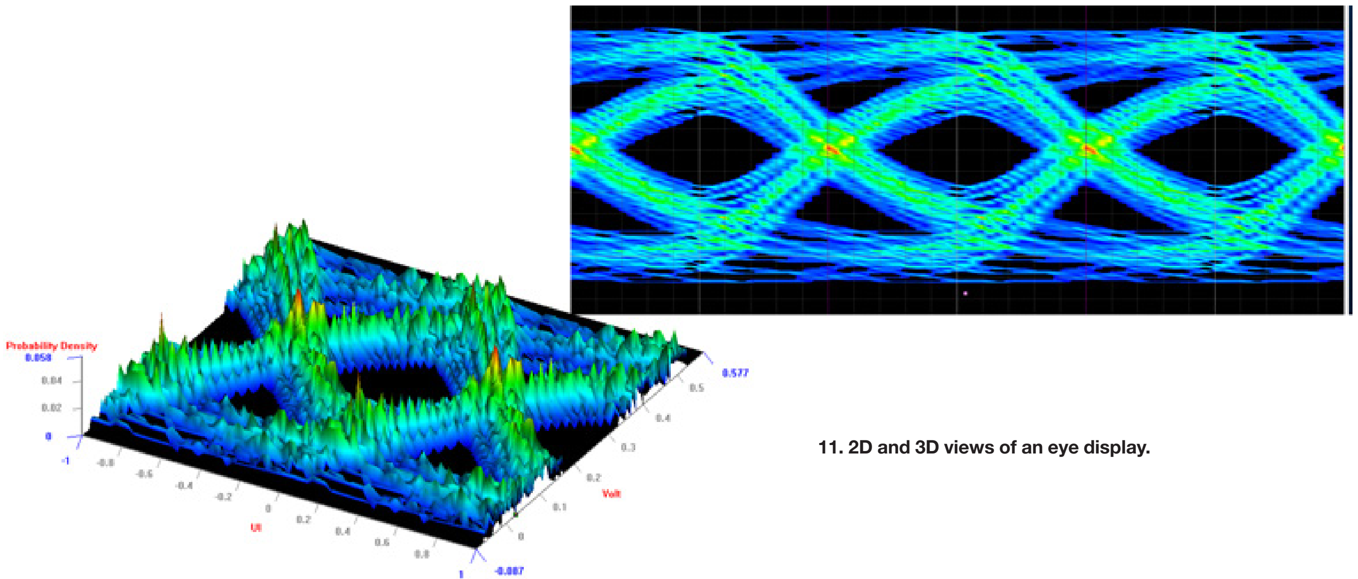
A match of 3% or better is needed to qualify the usage of the channel simulation and go on to the next step. That's because, ultimately, a channel simulation of millions of bits will be run to see if the GDDR6 specification for BER eye mask and eye opening is met. Selecting the different options for ramp response will allow the SI engineer to find the best option for achieving correlation within tolerance.

Next, a channel simulation is run, looking at various AMI model configurations to determine the need to enable equalization settings. The tool offers the ability to disable any of the AMI models and run a baseline simulation to determine if equalization is needed for the controller, memory, or both *(Fig. 9)*.

Once the equalization needs are determined, a what-if analysis can be run to change the AMI parameters and



10. Waveform and eye-opening results.



11. 2D and 3D views of an eye display.

see how that affects the eye opening or sweep certain AMI parameters. Multiple parameters are able to be swept, including AMI parameters. In this example, the transmit and receive AMI parameters were swept with three step counts each, which produced a total of nine iterations. The results are overlaid, enabling the designer to determine which simulation gave the best result and then set those values in the AMI model.

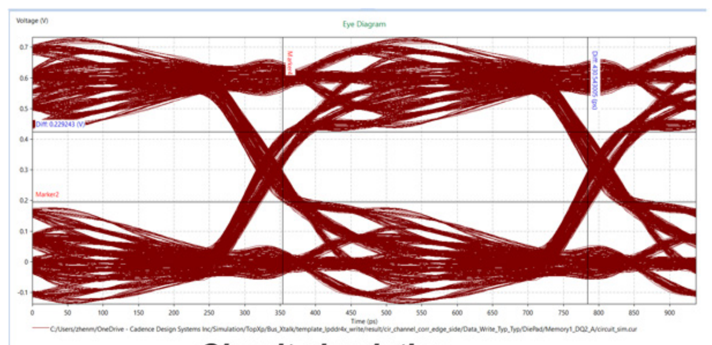
After the proper AMI parameters are determined, a channel simulation is run to look at the waveform and the eye opening *(Fig. 10)*.

The number of bits used for the eye diagram can be specified with the default being last 1,000 bits saved. Other results returned by the tool include the bathtub curve and a report of the eye-height opening for the different signals to see how much the eye height is being reduced with higher BERs.

Also provided is an eye display showing the distribution of the noise and jitter in both 2D and 3D views *(Fig. 11)*.

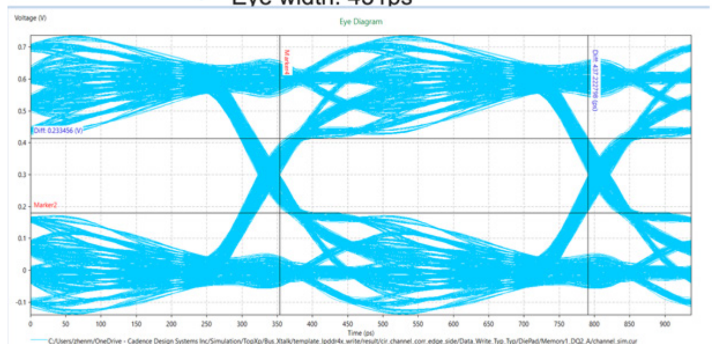
After running the channel simulation, additional characterization of the bus is performed as part of the workflow to capture the crosstalk on the individual signals. Because this is a channel simulation, the crosstalk needs to be characterized a little differently than the circuit simulation, which is embedded.

Figure 12 compares the circuit and channel eye measurement simulations, which correlate well with the eye height and width. This gives the designer confidence that million+ bit simulations can be run to capture the impact of crosstalk on the BER. This is essential to understanding how the design is going to



Circuit simulation

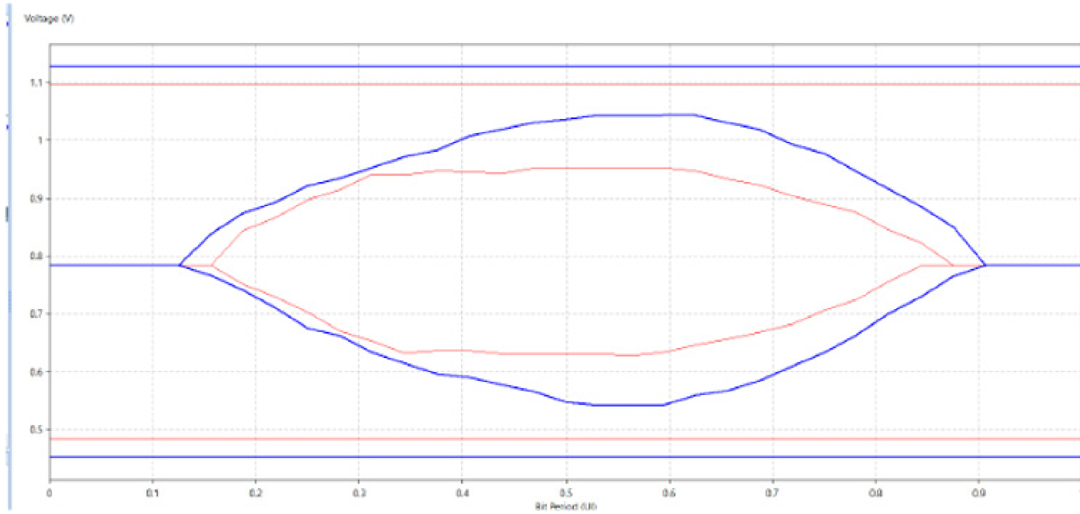
- Eye height: 229mv
- Eye width: 431ps



Channel simulation

- Eye height: 233mv
- Eye width: 436ps

12. Eye measurement comparison with circuit simulation on the top and channel simulation on the bottom.



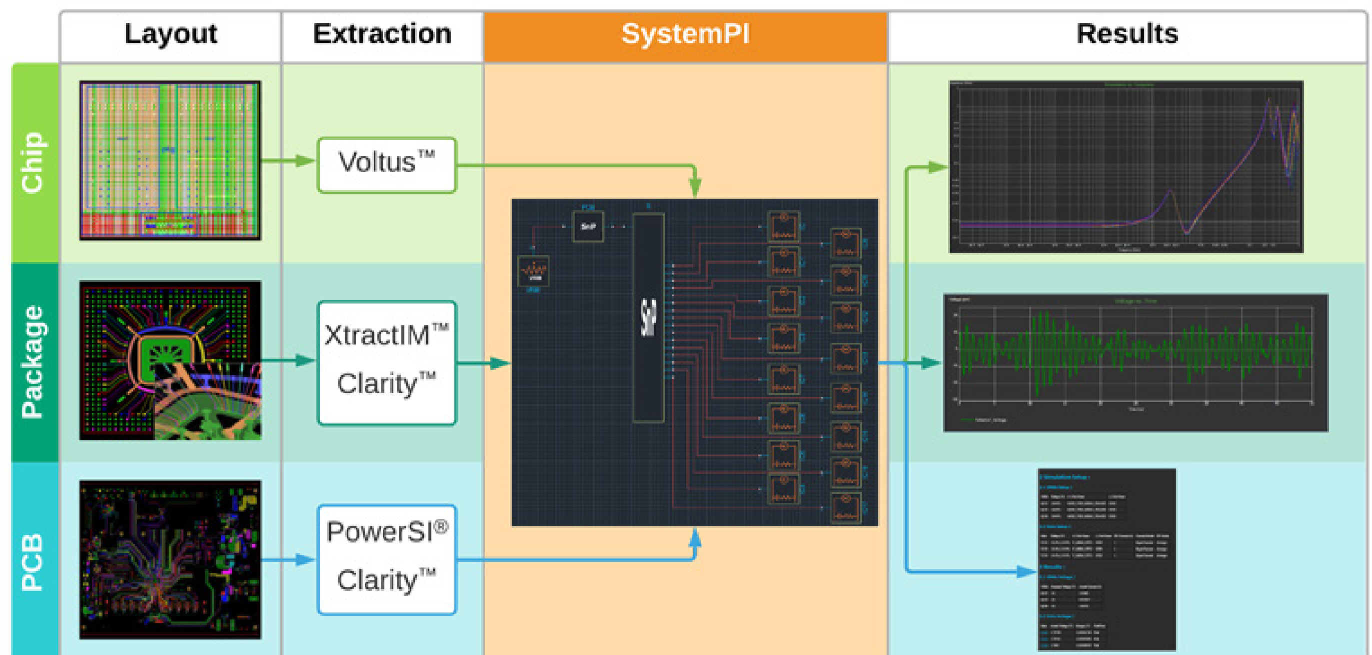
13. Characterization response comparison with blue representing ideal power supply and red representing non-ideal supply and crosstalk.

run in the field and ensuring that the GDDR6 compliance requirements are met.

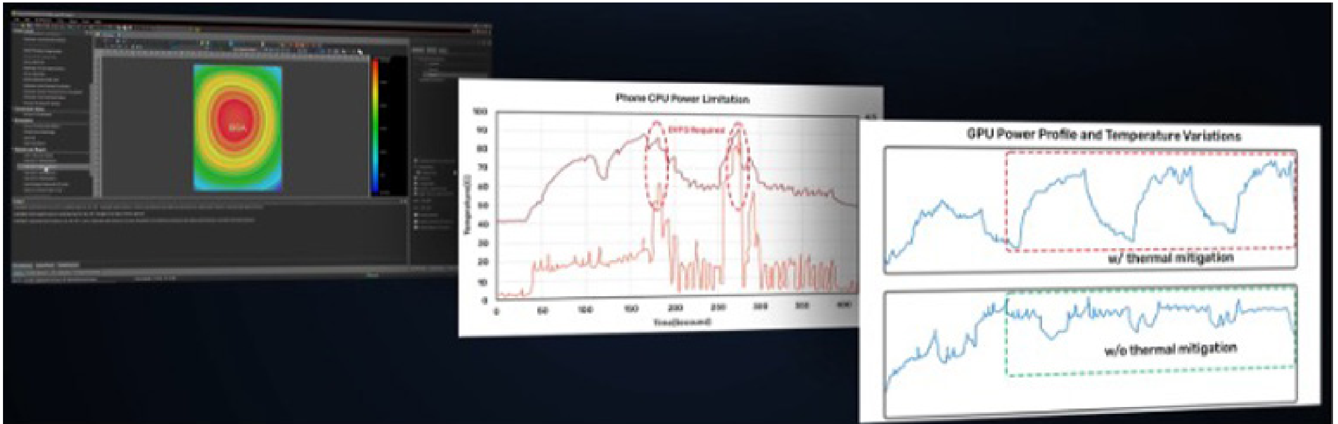
The next step in the workflow is to include power noise effects in the channel simulation. Up until now, an ideal power (VED and VCC) for both the controller and memory design is assumed. The power noise is affected and needs to be included in both the circuit and channel simulations. The channel-simulation power effects are captured through characterization, whereas the circuit power effects are captured through the interconnect model itself.

To verify the power noise in the channel, two channel simulations are run—one with ideal power supply and one with non-ideal power supply. The results enable the designer to compare the characterization responses and other channel-simulation results to see the effects of non-ideal power on the eye opening *(Fig. 13)*.

Once this step is complete, a report can be generated for compliance check and signoff. The GDDR6 JEDEC specification is built into the report generation and an enhanced data-rate-dependent measurement is included. The report highlights any violations. Within the report table, any of the signals can be selected in order to view the eye opening with the mask automatically calculated. If there's a violation, what-if analysis can be performed to come up with a solution.



14. SystemPI data flow for power-integrity analysis.



15. Thermal mitigation of the GPU in a cellphone application.

PI Testbench for GDDR6 Memory Interface

Many signal-integrity designers are responsible for power integrity, too. The GDDR6 testbench example also can be used to verify sufficient, efficient, stable, and reliable power delivery through the PCB and IC package for power-integrity signoff *(Fig. 14)*.

From a SI perspective, simulation is performed from the transmitter to the receiver. For PI, simulation is done from the voltage source or voltage regulator module (VRM) to the voltage sink (component pins). The new topology-based user interface makes it easy to connect multiple fabrics across cables, connectors, and other fabrics such as PCN, package, interposer, etc.

Proper PI analysis needs to be thermal aware. IR drop analysis should be performed integrated with thermal analysis. Celsius PowerDC and Celsius Thermal Solver provide electrical-thermal co-simulation as well as 3D thermal insight from a common user interface used by other Cadence analysis tools *(Fig. 15)*.

Conclusion

The memory interface roadmap is constantly advancing. Looking into the future, the GDDR7 specification now in development that will likely support 1.5 TB/s and the interim GDDR6X samples at 1 TB/s are already in production by Micron and NVIDIA.

The move toward pulse-amplitude modulation 4 (PAM4) used in GDDR6X and GDDR7 will come with new challenges. And parallel-bus-analysis technology from Cadence is advancing with the roadmap utilizing many years of experience of PAM4 simulation for serial-link analysis.

The GDDR interface has evolved beyond the original applications that addressed the need for fast data access for data-hungry GPUs in gaming applications. Today, design software is enabling GDDR6 to be leveraged for many high-performance applications, including automotive, high-performance computing, AI/ML, and 5G. Design teams can utilize a holistic solution from Cadence for areas of the memory interface design and analysis challenges, including IP, systems on chip (SoCs), interposers, IC packages, and PCBs.

Brad Griffin is a product marketing group director, System Analysis Group in the Cadence Design Systems System Custom IC and PCB (CPG) Group. He has over 25 years of experience in EDA technologies that enable the design and analysis of integrated-circuit packaging and PCB systems. Griffin is a graduate of Arizona State University.

References

Sigrity X Redfines SI/PI Analysis
Sigrity X White Paper