S1 EP30 - Defining Heterogeneous Integration

Friday, December 23, 2022 · 11:03

"Listen to Radha Nagarajan, Senior Vice President and Chief Technology Officer of Marvell's Optical and Copper Connectivity Business Group and Chris Banuelos on this week's episode, discussing a recent white paper "2.5D Heterogeneous Integration for Silicon Photonics Engines in Optical Transceivers." Learn more about the key takeaways of heterogeneous integration, what it means for cloud data center infrastructure and use cases".

Speakers
Radha Nagarajan
SVP & CTO, Optical and Copper Connectivity

Host
Christopher Banuelos
Senior Manager of
Global Social Media Marketing

C **Christopher Banuelos 00:04**
Welcome to the Marvell Essential Technology Podcast. I'm your host, Chris Banuelos on today's episode jump into a conversation with Raja Nagarajan, Senior Vice President and CTO for the Optical and Copper Connectivity Business Group discussing the recent white paper 2.5D Heterogeneous Integration for Silicon Photonics Engines in Optical Transceivers. Learn more about the key takeaways of heterogeneous integration, what it means for cloud data center infrastructure, and use cases, be sure to click the link in this episode's description, to read more. To stay up to date on future episodes, please be sure to subscribe to the Marvell Essential Technology Podcast. Hey Radha, thanks so much for joining today's episode. You and your team recently published a white paper "2.5D Heterogeneous Integration for Silicon Photonics Engines in Optical Transceivers." Can you describe what is heterogeneous integration?

R **Radha Nagarajan 01:02**
So, heterogeneous integration is a process by which you combine materials or different types on a common substrate, or generally in the silicon ASIC industry side, you try to combine ASICs from different sources. So let's take this one at a time. The silicon ASIC size is a little bit easier to understand, you have a silicon main chip, perhaps from one foundry, perhaps a CPU or a GPU. And you want to integrate it with a memory module from a completely different supplier. On the same substrate, a single substrate this forms a highly integrated system. This is heterogeneous from the foundry perspective, the paper we published takes it a little further, we use the silicon photonics chip as the integration platform on which we integrated amplifiers, not only from a different foundry, but a different technology itself, silicon germanium being integrated directly onto a silicon CMOS like Silicon Photonics chip. And on the silicon CMOS like Silicon Photonics chip, we also integrated lasers, which are from a completely different material system, in this case in the phosphide. So heterogeneity, if we can use that word. It's one of varied materials or varied foundry, or both. So it is the ability to take the best of multiple walls and put them on a common substrate in close proximity to enhance the value and the performance of the final system.

C **Christopher Banuelos 03:06**
From a technology standpoint, what are some of the benefits of heterogeneous integration?

R **Radha Nagarajan 03:12**
From the technology point of view, there's also a commercial point of view, from a technology point of view, no one material system, or no one foundry can get you everything you want. So sometimes the desire is to make the integration monolithic, meaning a single chip, and have all functions including the single chip, there are some downsides, a couple of downsides. One is the chip size grows, the power consumption, the power density grows. And at some point, you reach what's called a reticle limit. The dye that you can build as one coherent unit, you reach the limit and you can no longer put any more functions on a single dye. So that forces you actually even in the same material system to break the functionality among multiple different dyes on a single substrate. So when you go to that point, now you can pick different technology notes for the other dyes. Compared to the main dye, the main dye could be the advanced or while the typically it's the IO dye, could be in a different technology or we use the memory on a completely different technology. Now you've managed to integrate, you've beaten the size limits and you have managed to integrate technologies and devices from multiple different concepts. And if you take it a step further, so commercially, the advantages you can as well see you have diversified risk, better supply chain and also a somewhat of a shortened cycle time when you can have these chips built at different foundries. And also, you just can't build them monolithically. That's one of the advantages having two chips. In our case, when we deal with a lot of analog components, silicon germanium probably has a better performance than silicon CMOS, when it comes to high bandwidth, low noise amplifiers. So this allows us to actually get the best of both worlds in technology. And of course, lasers are only built out of the three five systems. So one is out of need, you just can't do it on a single system. Secondly, a diversified supply chain. And third, it just allows you to integrate systems into larger and larger units without having to rely on a single dye.

C **Christopher Banuelos 05:59**
What about this question? What about, you know, at a macro level? What does this mean for cloud datacenter Infrastructure?

R **Radha Nagarajan 06:07**
A couple of things. One is, as we go to higher and higher speeds, close integration is critical, just purely from loss and bandwidth perspective. Heterogeneous integration allows you to do that, or just closer integration allows you to do that. And when things get integrated together on a common substrate, you save a lot on power that's needed to drive across large printed circuit boards between components. When you save on the power, you save overall power consumption and energy efficiency of the system. And third, it allows you to realize the types of functionalities that you have not been able to realize in the past, and fourth, actually, it's turning out, as you integrate these components closer and closer together, they have a latency benefit. The latency refers to the time propagation of signals between processing units, and that allows for a better or rather a more efficient computational model. So yes, there are several advantages to integration from not just from just a technology perspective, or the ability to integrate or the sort of the diversity in supply chain, but from a number of other application perspectives

C **Christopher Banuelos 07:42**
And what are some of the use cases that come to mind?

R **Radha Nagarajan 07:45**
The one use case that we shared, which is becoming more important, more and more important, in the paper is called co-packaged optics. The premise or the co-packaged optics is to bring optical interconnects closer to the electronic computational elements or the electronic switching elements. More often than not, traditionally CPU has been has been touted as integration of optics, close to a switch ASIC, that does not have to be the case all the time, more and more and AI ML applications, we are talking about integrating optical components closer and closer to computational elements like a GPU or a tensor PU type, processing unit type element. So in these applications, we eventually get to a point where these high bandwidth, high computation intensive elements have an optical IO. So you've sort of eliminated the communication bottleneck between the various elements. So

heterogeneous integration allows you to place the higher bandwidth lower loss, optical interconnect elements, lower power consuming optical interconnect elements, as close as you can. So all these electronic components, thereby eliminating the need for larger and larger systems that end up consuming more power and dissipating more heat. So that would be a an example.

C **Christopher Banuelos 09:41**
And just before we finish, are there any other use cases that you could share?

R **Radha Nagarajan 09:46**
The other one that's commonly used and actually is in full sign of deployment right now is integration of memory to processors. It's a very traditional application. So the access speeds to minimize the access speed bottleneck, which in the computation on the memory element, increasingly people are resorting to heterogeneous integration. Although the components are of the same silicon node, putting them together via a very short, very low power consuming interconnects achieve both an inquiry, both a advantage of lower latency and low power consumption, that will be the other one Chris.

C **Christopher Banuelos 10:27**
Radha I wanted to say great work for you and your team on the white paper. Always a pleasure to have you on the podcast and thanks for joining today's episode.

R **Radha Nagarajan 10:34**
Thank you very much, Chris. I very much appreciate the opportunity to share the excitement. 2.5D are generally in heterogeneous integration with a larger audience.

C **Christopher Banuelos 10:49**
Thank you for listening to the Marvell Essential Technology Podcast. As always, please feel free to visit our website to learn more, and we'll see you on the next episode.