THE KEYS TO SUCCESS

Designing a product from the ground-up is challenging. **Bethan Grylls** talks to Imagination about how its team successfully delivered on the PowerVR Series2NX

Above: A handful of the Imagination team behind the PowerVR Series2NX

Delivering a successful product from the ground-up in 14 short months requires a hardworking team, especially when that team is made up of 90 engineers located around the globe.

Imagination, who took home the Design Team of the Year award at last year's British Engineering Excellence Awards (BEEAs), overcame these challenges with the development of its PowerVR Series2NX.

According to Imagination, its team of engineers were looking for an industry solution capable of executing the compute-intensive operations required for neural network acceleration (NNA) on edge devices.

The PowerVR project was designed to deliver high performance computation of neural networks (NN), but with very low power consumption across a smaller silicon area.

Begun in 2016, two years were allocated to the project from initial concept to evaluation, and by October 2017, despite a geographical spread team, the first evaluation systems were released. While the first production release of the 2NX IP, the AX2180, was achieved soon thereafter.

Given its experience in designing GPUs for mobile and embedded systems, Imagination started the design by modifying the PowerVR GPU architecture to achieve the performance levels needed for NN acceleration. However, this resulted in a solution that was too large, so the team started to work on a new system from the ground-up – the 2NX, which was intended as a specific, optimised hardware acceleration unit for convolutional NN.

Once they made the decision to create a new design, the target was to achieve ~10-100x higher performance per mW and performance per mm^2 compared to other embedded NNA solutions then on the market.

According to Russell James, VP vision and AI, Imagination, who led the business unit, success could be attributed to a combination of: strong leadership, clear communication and a well-organised timeline of work.

"At each stage we had to check that our product was still valuable," James said. "We had to keep abreast of new developments because the market for NNA on edge devices emerged so quickly – that was a challenge. As we moved forwards, we kept adding more detail to the concept, checking it, adding more detail and so on."

Every feature of the 2NX was designed specifically for it. With no pre-existing flows, tools or system architecture to help ease or speed up the process.

"Because everything needed to be custom designed, it presented the team with a unique opportunity to create the most efficient and competitive product they could without any constraints," James said.

More with less

Initially, the project started off as a small, dedicated group of engineers with limited resources. "As a British IP developer in semiconductors we don't have the resources or the budget. We have to do a lot more with a lot less."

During the first stages of the development process, James said it became evident that the market opportunities for the solution were vast, and it was a clear-cut decision to grow the team.

"I did expect that it would become a big project," James said. "The technology is absolutely crucial. Providing you have enough compute and performance, you can apply NN to some very interesting tasks. Once we had a proof-of-concept and knew the direction we were going in, we had to grow the team."

But "you can't hire a good team overnight," James admitted – particularly in a smaller company.

"Quite often we have bought small companies just to get a team together quickly. Organic growth in this industry takes a while – especially when you want to build a good team. That's why you start small – you don't really have much choice."

For the 2NX, James said they were able to allocate people internally, re-tasking in-house engineers as the project progressed. But he confessed that it did take several months to integrate the groups as one focused team because many of them were accustomed to executing projects in smaller subgroups.

"When you have a large, segregated team you have to divest efficient responsibilities and accountability to the team leads in each area, so they are empowered to take ownership of their tasks," he said. "This means they can press on with work independently and take on the responsibility that the work gets done."

He continued: "With teams in different time zones, you also don't want to be waiting 8 hours for them to resolve a problem."

James pointed to "clear and regular communication at the right level of granularity," as crucial. "You don't need to recite an essay on every single thing that happens during your meetings, you just need a concise report on what's been done, what is going to be done next, and what the blocking issues are. With those three questions answered, provided you keep up that regular communication, most issues can be resolved."

In this particular project, James said that video conferencing was a key form of communication. "Of course, email is the most common way of connecting, but I am very much in favour of video. You gain something from being able to see someone's face."

He emphasised the importance of relevant communication too. "If you have a follow-up meeting about an issue that was flagged, only involve those who need to be involved.

"You need compromise too," he advised. "When you have a geographical spilt someone somewhere is going to have to get up early or stay late. In those situations,

"It's very easy to give attention to the most interesting parts of the system, but they aren't always the most used by the end-customer."
Russell James

it's not just about accepting that, but also rotating that responsibility fairly."

James also pointed to limiting the need for awkward combinations of time zones. "If an issue arose, we made it so there was something else a team could work on in the meantime, so that the time difference didn't hold up overall progress."

Despite the challenges it presented, James said the geographical divide also had its advantages in the form of "diversity of thought", as he described it, and with teams able to handle work in multiple time zones, he said that the speed of development, research and customer support increased.

Focus was also a key element of James' management process. "It's very easy to give attention to the most interesting parts of a system. But, they aren't always the most used by the end-customer. Get the key functionality in place first and then add the lower priority parts later."

He continued, "It's also about knowing when to stop. You can always add new features, but if you keep doing that, you'll never actually deliver a project. Pick the right balance between features and deadlines. Identify key customers to work with and remove as many obstacles for your team as you can."

Despite being a seasoned manager and successfully coordinating the 2NX team, James admitted that he did learn a thing or two during the process. "There is always opportunity to bring new ideas whether they're yours or someone else's."

He added, thoughtfully: "A day without learning something new is a very sad day indeed."

Since taking home Design Team of the Year, Imagination is set to release the 3NX, an update to the IP that incorporates a flexible processing element. "We are, in fact, already working on the next generation beyond that," James said, "it's moving extremely fast."
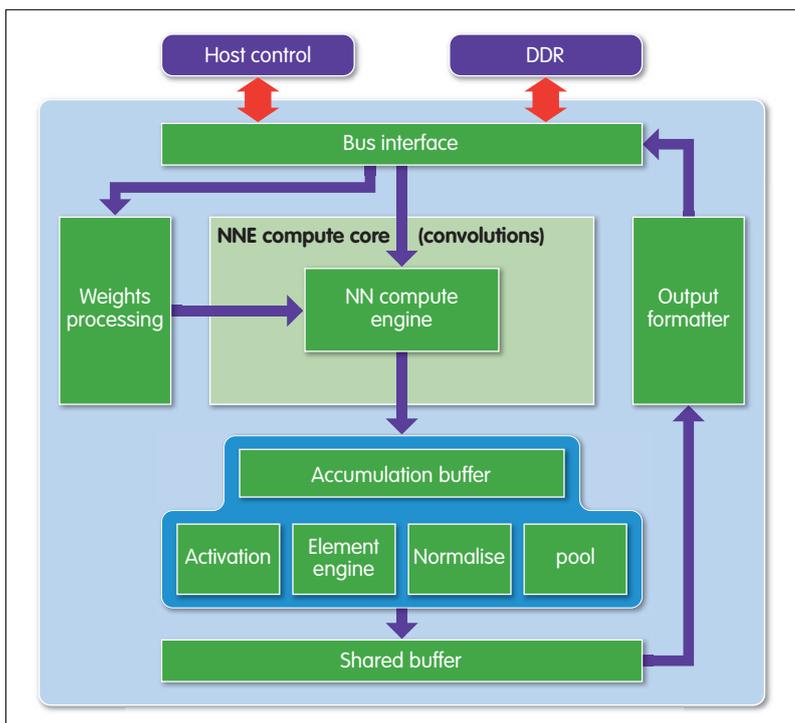
Left: The 2NX is a new architecture from Imagination, designed from the ground-up as a dedicated inference engine across a range of markets including mobile and surveillance