# D2S
Optimizing Semiconductor Manufacturing with GPU Acceleration

It is now widely accepted that GPU-accelerated computing is revolutionizing high-performance computing (HPC). Driven by an aim to deliver the highest performance for the semiconductor manufacturing industry is D2S, a developer of innovative GPU-acceleration solutions at a rapid-fire sequence for a number of software applications as well as different manufacturing equipment capabilities. D2S plays an enabling role by leveraging the strength of GPUs for simulation of nature, image processing, and deep learning, particularly in model-based processing.

"The key is the difference that GPU acceleration provides in the rate of improvement for computing single-instruction, multiple-data (SIMD) tasks," remarks Aki Fujimura, Chairman and CEO, D2S. Programs written for SIMD can continue to scale, whereas conventional programs scale by coarse-level parallelism and are limited by Amdahl's Law. The Law says that coarse-level parallelism hits an asymptotic limit in performance because of overheads associated with dividing and assembling the problems. "Any application that is suited for SIMD will be GPU-accelerated sooner or later, and D2S wants to be one of the sooner ones," he adds.
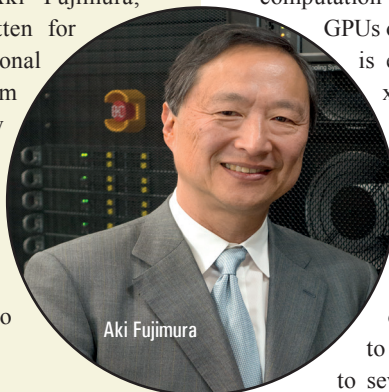
> " Deep Learning completes the third leg of the triangle that makes CDP the ideal platform for semiconductor manufacturing applications "

With more than a decade of experience, D2S has developed GPU-accelerated engines, models and the Computational Design Platform (CDP) to serve multiple applications within the semiconductor photomask manufacturing segment. The company launched its fifth-generation CDP in April last year, featuring NVIDIA Pascal-based Tesla P40 GPUs which allow these CDPs to achieve 888 Teraflops of processing speed in Single Precision (that's more than double the speed of previous-generation CDPs from D2S). Keeping in mind that the semiconductor-manufacturing environment is sensitive to

Aki Fujimura

any downtime, the new CDPs are architected to ensure the high speed, precision, and reliability required for round the clock clean room production environments.

"Deep Learning completes the third leg of the triangle that makes CDP the ideal platform for semiconductor manufacturing applications," mentions Fujimura. Since semiconductor manufacturing requires massive simulation of nature and image processing, GPU acceleration is already a necessity. In addition, GPU is an ideal platform for Deep Learning. "For semiconductor manufacturing applications that we focus on, there's a need to 'see' the data in a special way. They require filters for the data to transform the data. This often takes single or double precision computation that is subject to 10 to 100 times speedup on GPUs over CPUs." As the amount of data on one chip is enormous—0.1nm precision spanning 2.5cm x 3cm space on a wafer and 0.1nm precision spanning 10cm x 13cm on a typical mask—540TB is what it takes to represent pixel data of a multi-beam mask today. PLDC, a capability embedded inside NuFlare's MBM-1000 multi-beam mask writer, runs on a CDP to perform real-time in-line correction of mask shapes and enhancement of resiliency to manufacturing variation. That's equivalent to seven years worth of compressed 4K video to manufacture one mask. Fujimura says, "We predict a very broad-based adoption of GPU-acceleration everywhere, but specifically in electronic design automation (EDA), and D2S is well-positioned in EDA for GPU acceleration."

Fujimura also stresses on the fact that successful GPU-accelerated systems don't equate to complete replacement of CPU-based computing, but rather a balance of GPUs and CPUs is necessary. As an illustration, D2S ran Gaussian convolution, a specific type of computation that is useful in both simulation of nature and in image processing, on an arbitrarily sized piece of mask data (~80µm by 80µm, 10nm pixels), using a node of CDP to demonstrate runtimes for a CPU-only implementation, and for a CPU+GPU implementation. Algorithms and implementations were optimized for each platform separately. The CPU+GPU version was found to be ten times faster.

Reliability and serviceability are critical factors in the semiconductor manufacturing equipment space, and currently, D2S is nailing it with their new-generation balanced computing solution for better performance and optimal acceleration. **CR**