Prasad Bachiraju

Monita Pau

How AI is helping optimize AI chip production

PRASAD BACHIRAJU and **MONITA PAU**, Onto Innovation

IN EARLY FEBRUARY, OPENAI CEO SAM ALTMAN issued a rallying cry to transform the semiconductor industry by calling for the rapid expansion of AI chip production. The magnitude of this undertaking is reflected in its $5 to $7 trillion price tag.

To create the kind of integrated packages needed for AI – ones with graphic processing units (GPUs), high-bandwidth memory (HBM) and advanced integrated circuit substates (AICS) – manufacturers are increasingly turning to metrology, inspection and thin film solutions that lean on data-hungry, machine learning-based software technology.

As a company, Onto Innovation increasingly leverages machine learning tools in our process control portfolio. By amassing large amounts of data for each component in an AI package during various process steps, and then using machine learning to identify potential issues, this combination of hardware and software enables customers to make process control adjustments that will optimize their AI package-making abilities.

But this new era in manufacturing brings with it new challenges, whether we are dealing with microbump inspection in GPUs, metal film measurement in HBM or critical dimension (CD) in through-silicon vias (TSVs) in 2.5D packages using AICS. This brings us to the elephant in the clean room: for machine learning software to operate on its own without human input, it needs large amounts of data to learn from and a place to store the terabytes and terabytes of data generated weekly, if not daily. And this deluge of data is only increasing, exponentially so in the case of AI packages.

Take GPUs for instance. Advanced logic is now approaching 3nm and 2nm, and with smaller sizes comes increased complexity. These advanced nodes have increasingly higher numbers of input/output (I/O) in the form of microbumps. These microbumps connect the GPU to a silicon interposer which then allows the GPU to communicate with the HBM. (More on memory in a minute.)

Today, bump counts in GPUs can reach 100 and 200 million bumps on a wafer. As bumps increase, the risk of defectivity and chip failures increases as well. Integrated process control solutions combining cutting edge hardware and machine learning software with access to large amounts of data will help isolate potential failures and enable manufacturers to make process adjustments before future errors arise. To accomplish this an AI solution will need to churn through large amounts of bump data, and that data may include information about bump height, bump diameter, bump dimension in the X and Y directions, etc. In other words: more vectors of data than in years past.

This is also occurring at a time when customers want a better understanding of what is going on inside the product, and, equally as important, how those individual components interact with each other. While we can test these components individually, that is not a guarantee they will work in concert with the other components in an AI package. Consider this scenario: a faster than normal GPU is paired with low-end memory. They may both pass individual requirements, but the combination may be mismatched and result in a nonfunctional package. As such, manufacturers would be better served with integrated software tools that can examine the diverse kinds of chips that go into the package and make sure they work together optimally.

With HBM, the key memory block for the AI package, manufacturers are stacking more DRAM layers on top of each other, going from eight layers to 12 and, even, 16 layers. Microbumps play a role here since these bumps connect layer to layer and finally the HBM to the silicon interposer. As with GPUs, bump density is increasing while bump size is shrinking, bringing with it an increased need for more stringent bump inspection, 2D/3D metrology and thin film measurement tools. And, as with most everything AI and machine learning related, that means a considerable increase in data being generated and churned through.

TSVs also play a vital role in HBM. They enable increased numbers of I/O between layers and the DRAM stack and silicon interposer. And like microbumps, TSVs are getting smaller in terms of dimension and density. As with GPUs, the amount of data being collected and

Continued on page 38

Continued from page 40

consumed is considerable. Further complicating matters: the TSV process also requires etch, deposition, fill and chemical mechanical planarization, making this piece in the AI production puzzle one that requires lithography, inspection and metrology. In the future, people are looking into doing hybrid bonding instead of using microbumps to stack DRAM, bringing with it additional challenges. But that era may still be a ways off.

Designed for panel-based products like AI chips, AICS are key to AI chip production. Why? AICS provide the connective tissue, if you will, between the silicon interposer and the printed circuit board (PCB). But as is often the case, solving one problem creates another. In this case, an AICS is made up of individual RDLs, and as each layer is built upon the previous one, overlay drift becomes a mounting concern. And overlay isn't the only issue. AICS is also subject to errors caused by dry film resist and laminate non-uniformity, RDL line defects, and under-laminate bubbles and particles. Once again, machine learning software, paired with a metrology or inspection tool, will require large amounts of data in order to be useful in spotting potential defects and errors.

Further complicating matters: as a panel-based technology using extremely large package sizes (120mm x 120mm), AICS offers far fewer packages per 510 x 515mm panel than fan-out panel-level packing. The difference in numbers is dramatic, with FOPLP offering 2,300 packages per panel and AICS offering a scant 16 packages. As a result, one bum package represents a yield loss of 6.25% compared to 0.04%, and this issue is only going to become more significant as AICS packages transition to 150mm x 150mm. To address these issue, AICS manufacturers need to employ stringent metrology and inspection process control measures, along with AI software solutions, at each important process step.

We are on the cusp of an exciting era in technology, both on the manufacturing floor and as end-product users. To meet the challenges of this coming age, each process control sector will play an important role, from metrology to inspection, lithography to software. The HVM of AI chips will need new solutions, whether that is in the manufacturing of GPUs, HBM or AICS-based advanced packages. More importantly, these solutions will need to be more integrated than ever with machine learning software. sd

Prasad Bachiraju is Director, Sales and Customer Solutions, and Monita Pau is Strategic Marketing Director, Advanced Packaging, at Onto Innovation.