High Bandwidth Memory (HBM) is a new type of memory device bringing higher bandwidth, smaller form factor and lower power consumption to keep up with processor performance growth. This is achieved by stacking multiple DRAM dies onto a base controller die, which are interconnected by through-silicon vias (TSV) and micro bumps. The advanced wafer level packaging to fabricate HBM chips also introduce increasing challenges in testing and handling of these delicate devices.

Current testing of these HBM devices is done at pre-singulation stage using traditional ATE and wafer prober followed by rigorous vision inspection steps after singulation. This approach is very costly because of the extremely long test times (tens of minutes) and high pin count (>200) of the devices limiting test parallelism. Over rejection from the vision inspection is also very costly as the cost of manufacture for these devices is very high. It also does not provide the level of quality needed as there are many more handling steps of these very fragile and delicate devices before shipment to the end customers.

A better approach to testing HBM devices is to have it done after singulation and compress the remaining back end finishing steps into one to reduce the number of times that these delicate devices are moved and contacted. This approach will require an integrated system with leading edge technologies in vacuum chuck with vision alignment to get reliable fine pitch contact and batch handling of the devices, thermal capability to achieve multiple temperature tests with single insertion, inspection capability to detect defects and contamination and automation to achieve high throughput and automated factory processes. This approach also gives an option to have the devices tested using traditional ATE or using a system level test (SLT) method to achieve massive parallelism (>1000) to significantly reduce the cost and improve quality of test.