November 2022

3D High Bandwidth Memory with Optical Connectivity Stacking

n/a

3D High Bandwidth Memory with Optical Connectivity Stacking

ABSTRACT

This disclosure describes techniques for thermal management of 3D stacked high bandwidth memory (HBM) with optical interfaces. Per techniques of this disclosure, the HBM thermal (heat) load is configured to be in a separate package from the ASIC package, thereby enabling optimized cooling of both components. The HBM package includes multiple HBM DRAM dies and a DRAM base die. A DRAM interface light bundle provides optical connectivity to the HBM package, and is coupled to an optical connector.A cold plate or other thermal management solution can be deployed at an upper surface of the HBM package via direct contact to the HBM DRAM die, thereby providing superior thermal management. The configuration also enables utilization of coldplates for cooling the HBM via direct contact. Power consumption of the HB package is also reduced due the elimination of 2 high speed PHY circuits from the HBM package.

KEYWORDS

- High bandwidth memory (HBM)
- 3D memory
- 3D stacking
- Surface mount technology
- Cooling stack
- Thermal management
- Optical connectivity
- Micro-LED

BACKGROUND

High performance processors such as graphics processing units (GPU), vision processing units (VPU), application-specific integrated circuits (ASIC), etc., commonly include three-dimensional (3D) stacked high bandwidth memories (HBMs), e.g., as memory core dies within the same package. The processors typically include a base die that hosts a physical layer (PHY) interface that provides optical connectivity and connects the high bandwidth memory to the main application-specific integrated circuit (ASIC). The PHY consumes power and dissipates heat. Combined with the memory and ASIC, this can lead to significant heat dissipation that needs management. For example, a typical 3D stacked HBM is rated at over 50W and has a complex thermal profile. In this scenario, the position of the optical connectivity prevents effective cooling. This can lead to either reduced performance if the power consumption is controlled (mitigated) or poor reliability due to thermal effects.

DESCRIPTION

This disclosure describes techniques for thermal management of 3D stacked high bandwidth memory (HBM) with optical interfaces that can be utilized for high performance computing (HPC) applications. Per techniques of this disclosure, the HBM thermal (heat) load is configured to be in a separate package from the ASIC package, thereby enabling optimized cooling of both components. This design can support larger ASIC designs (e.g., monolithic or as a chiplet) since the design no longer needs to account for the HBM. The configuration also enables utilization of cold plates for cooling the HBM via direct contact. Power consumption of the HBM package is also reduced due the elimination of two high speed PHY circuits from the HBM package.

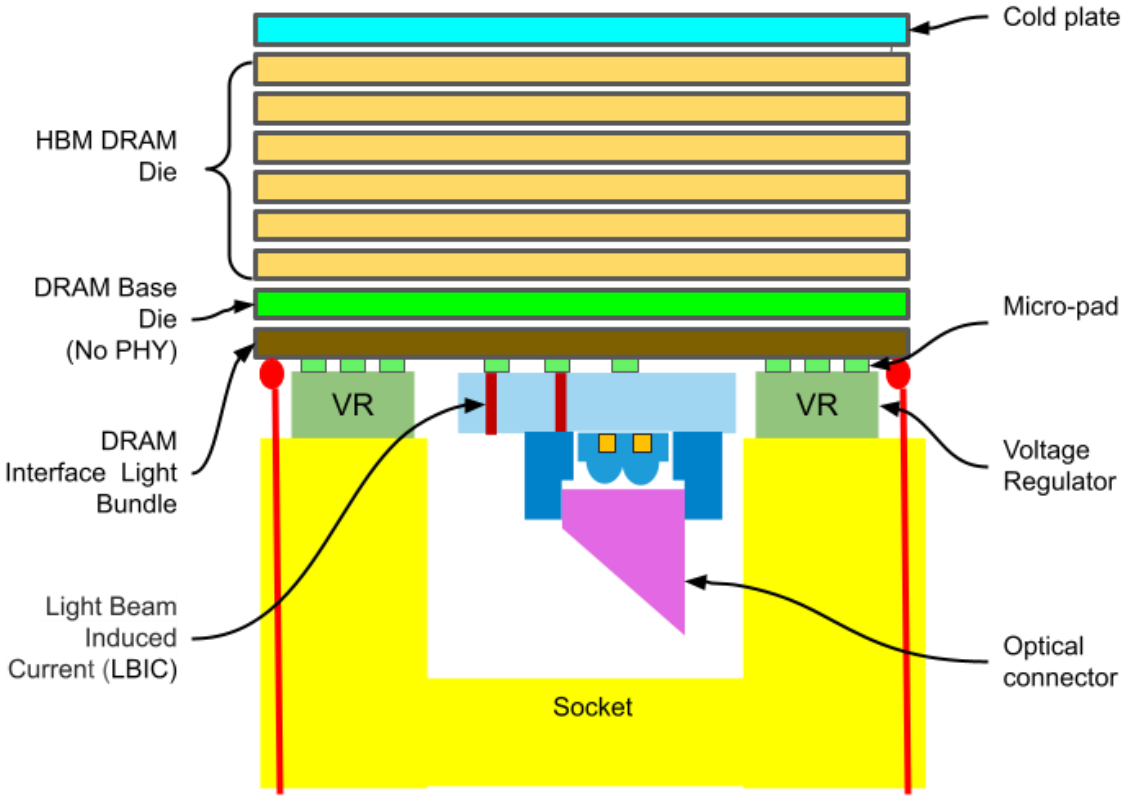

Fig. 1: 3D stacked high bandwidth memory (HBM) package

Fig. 1 depicts an example 3D stacked high bandwidth memory (HBM) package. As depicted in Fig. 1, the HBM package includes multiple HBM DRAM dies and a DRAM base die that does not include a PHY. A DRAM interface light bundle provides optical connectivity to the HBM package and is coupled to an optical connector. A cold plate or other thermal management solution can be deployed at an upper surface of the HBM package via direct contact to the HBM DRAM die, thereby providing superior thermal management.

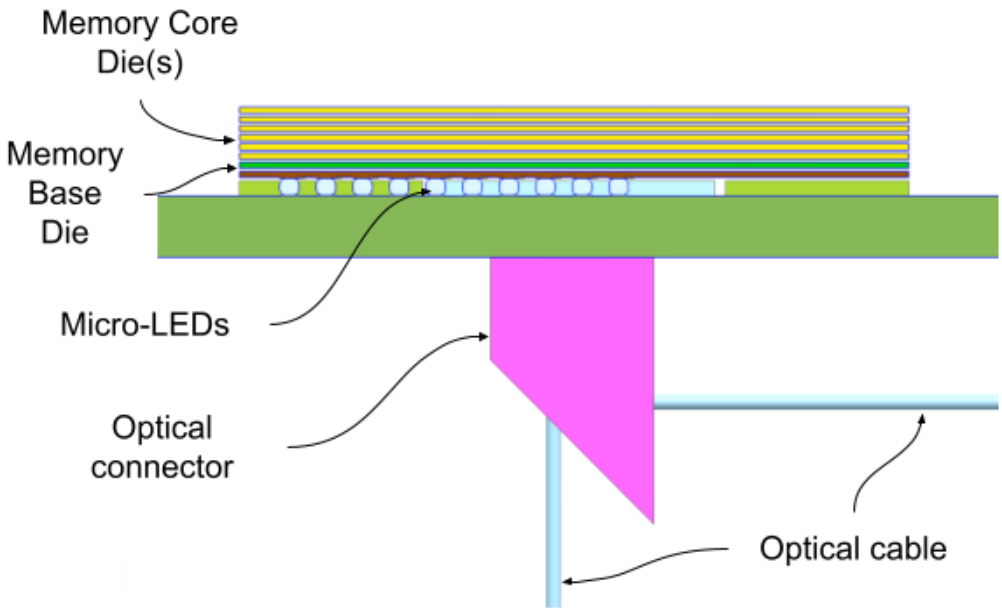

Fig. 2: View of 3D stacked HBM package

Fig. 2 depicts another view of a HBM package. As depicted in Fig. 2, memory core dies are provided at an upper portion of the package to enable utilization of thermal management solutions via direct contact. A memory base die is also depicted that can provide support for Memory Built-In Self Tests (MBIST) and other tests.

The base die provides connectivity to the optical interface. Microbumps can be utilized for providing connectivity between the optical system and the base die. Micro-LEDS are provided for electrical to optical signaling. In some implementations, the base die can include a gear box. An optical connector that is connected to one or more optical cables is depicted.

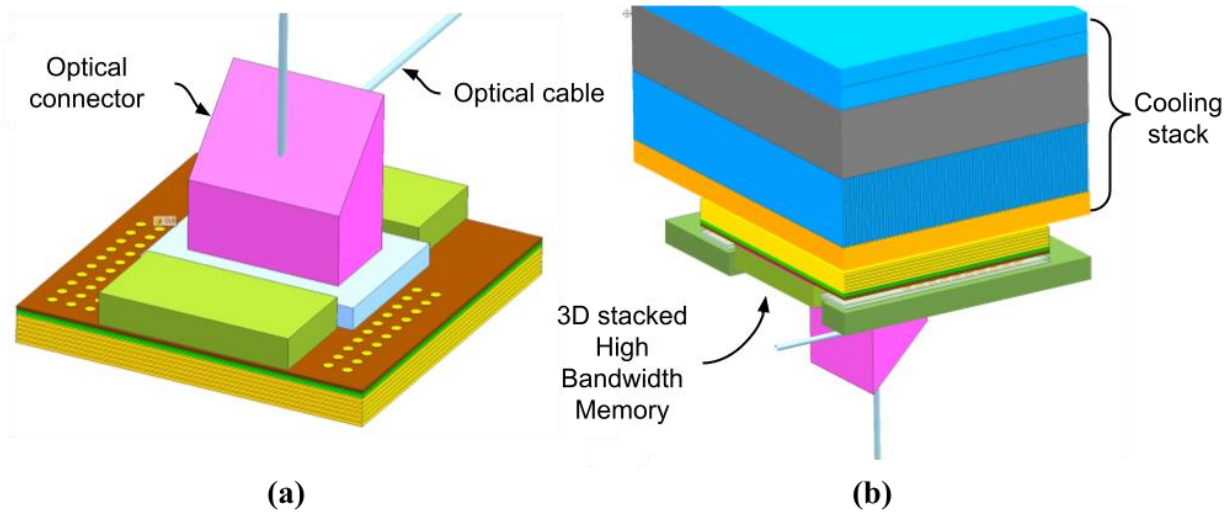

Fig. 3: (a) Isometric view of HBM package; (b) deployment with cooling stack

Fig. 3 depicts additional views of the 3D stacked HBM package. Fig. 3(a) depicts an isometric view of the package while Fig. 3(b) depicts an example deployment of a cooling stack for thermal management of the HBM package. As depicted in Fig. 3(b), locating the memory dies at an upper portion of the HBM package enables deployment of cooling solutions, e.g., cold plates, via direct contact.

The high bandwidth memory components can be manufactured at low cost and tested separately from the main ASIC package and installed as memory banks. In this design, unlike traditional HBM packages which include the ASIC and HBM within the same package, the HBM can be swapped as needed to improve productivity, besides providing superior thermal management.

Techniques of this disclosure can enable increased available capacity for high performance processors by increasing (e.g., doubling) available memory. When compared to traditional HBM packages, techniques of this disclosure provide increased flexibility of

communication solutions, greater bandwidth, higher capacity availability, increased system reliability, and superior thermal management.

CONCLUSION

This disclosure describes techniques for thermal management of 3D stacked high bandwidth memory (HBM) with optical interfaces. Per techniques of this disclosure, the HBM thermal (heat) load is configured to be in a separate package from the ASIC package, thereby enabling optimized cooling of both components. The HBM package includes multiple HBM DRAM dies and a DRAM base die. A DRAM interface light bundle provides optical connectivity to the HBM package, and is coupled to an optical connector.A cold plate or other thermal management solution can be deployed at an upper surface of the HBM package via direct contact to the HBM DRAM die, thereby providing superior thermal management. The configuration also enables utilization of coldplates for cooling the HBM via direct contact. Power consumption of the HB package is also reduced due the elimination of 2 high speed PHY circuits from the HBM package.