［様式－学 5］

主　論　文　要　旨

2010 年 12 月 24 日

論文題名　MICRO-ARCHITECTURAL MECHANISMS FOR LOW COST HIGH PERFORMANCE ON-CHIP NETWORKS

ふりがな　ぐぇん　ちゅおん　そん
学位申請者　NGUYEN TRUONG SON

主論文要旨

In recent years, as the technology scales to integrate a larger number of cores in a single chip, Network-on-Chips are becoming more important for communication on System-on-Chips. Designing high throughput and low latency on-chip networks with reasonable area overhead is becoming a major technical challenge. This thesis proposes a number of micro-architectural mechanisms for developing low cost high performance on-chip networks.

First, the thesis introduces a low cost low latency router architecture based on the single virtual output queuing (VOQ) scheme for on-chip networks. By maintaining a dedicated virtual channel (VC) for each output channel, the critical path of the router can be minimized without any additional control. Our experiments showed that our proposed single VOQ router can significantly reduced the area overhead for control logic and communication latency by up to 64.5% and 45.5% respectively as compared to the conventional design.

Next, this thesis proposes the multiple VOQ router architecture for alleviating the impact of traffic congestion and avoiding deadlocks under heavy workloads. Instead of a single VC, multiple smaller VCs are dedicated for each output channel. This helps to reduce the traffic congestion without an increase of the critical path delay. The experimental results showed that the multiple VOQ architecture offers an alternative, which enables to reduce the hardware amount for control logic by up to 15.6% and communication latency by up to 45.5%.

In addition, a mechanism called on-the-fly VC allocation is developed. By performing the VC allocation during the time a packet is traversing the crossbar switch, the router pipeline can be shortened in a non-speculative fashion. In this manner, the router throughput can be significantly improved. The proposed router with on-the-fly VC allocation can reduce the communication latency by up to 40.9%, and improve the throughput by up to 47.6% without the penalty of area.