

# A Cache Consistency Protocol with Improved Architecture

Qiao Tian, Jingmei Li<sup>(✉)</sup>, Fangyuan Zheng, and Shuo Zhao

College of Computer Science and Technology,  
Harbin Engineering University, Harbin, China  
lijingmei@hrbeu.edu.cn

**Abstract.** The effective cache consistency protocol plays an important role in improving the processor performance. This paper designed an improved architecture of consistency protocol for multi-core environment, adding the D-Cache virtual bus to achieve the point-to-point consistency transaction transmission which avoided the bus idle phenomenon caused by the polling query method that the broadcast consistency transaction must be observed. The experimental results show that the architecture can improve the bus utilization.

**Keywords:** Consistency protocol · Multi-core environment · Virtual bus

## 1 Introduction

Cache consistency as one of the hot issues in the processor research, it is a technical problem to be solved that determines whether the multi-core technology can be further developed [1]. Therefore, the design of an effective cache consistency protocol to improve the processor performance is of great significance.

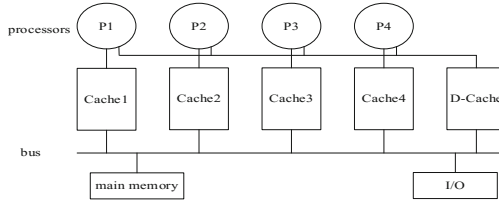
The traditional consistency protocols such as bus listening protocol and directory consistent protocol have their own advantages and disadvantages. Based on deep research and analysis, this thesis realized the point-to-point consistent transaction transmission by adding D-Cache virtual bus in the architecture, which improved the effective utilization of bus.

## 2 Consistency Protocol Optimization

### 2.1 The Analysis and Optimization of Bus Listening Protocol

The bus listening protocol [2] uses bus to connect processor private cache with the main memory, propagating consistent transaction messages on the bus in broadcast, so the bus is the ordering point, all nodes connected to the bus can observe the messages in same order.

In bus listening protocol, the polling query method [3] that the broadcast consistent transaction must be observed produces bus idle occupancy. Based on the above shortcomings, the D-Cache virtual bus architecture model is added to improve the equipment utilization. The D-Cache virtual bus structure is shown in Fig. 1.



**Fig. 1.** The system structure with D-Cache virtual bus

The D-Cache virtual bus is used to store directory entry that records data information. By designing and modifying the directory entry, constructing request transaction collection unit, directory entry lookup and update unit and listening response transaction unit, the improved directory entry structure is shown in Fig. 2.

Valid_Bit	Status_Bit	Ident_Bit	Share_Bit	Busy_Bit	Count_Bit
-----------	------------	-----------	-----------	----------	-----------

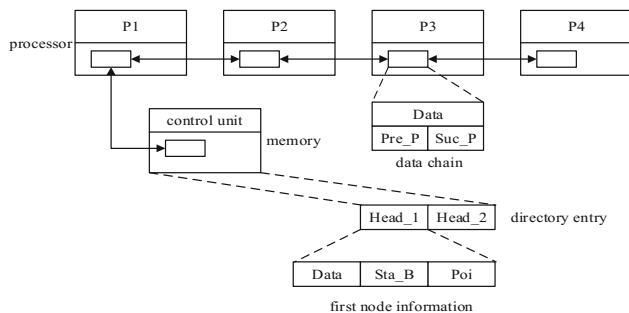
**Fig. 2.** The directory entry structure

In Fig. 1, raising the location of D-Cache to the private cache of each processor core speeds up the search and reduces the access delay. In Fig. 2, the read and write requests from processor cores are first cached in the request transaction collection unit. The lookup and update unit matches the directory entry in D-Cache with every request, which is identified by Ident\_Bit, 1 for hit 0 for miss, after the hit, checking whether Valid\_Bit is 1, 1 represents that the entry is valid, otherwise it is invalid; then checking Busy\_Bit, 1 represents that the data block is being used, the data block can only be read and wrote until Busy\_Bit is 0; when Busy\_Bit is 0, the Status\_Bit and Share\_Bit can get the state of target data and which processor core contains the data in their private cache. When the corresponding read and write requests are met, the entry information will be updated, Count\_Bit will be incremented by one, which is used as a reference bit when the data block is replaced, the data block with the smaller number is preferentially replaced.

## 2.2 Directory Consistency Protocol

The directory protocol uses the directory to store information about the cache data copy, it serves as the ordering point. The requested data is obtained in point-to-point communication after finding the directory. All consistent messages are forwarded through a directory structure. The directory protocol is represented by fully associative directory, limited directory and chained directory [4, 5].

By combining the fully associative directory and chained directory, a new cache consistency protocol of two-level directory structure is proposed. The system architecture is shown in Fig. 3.



**Fig. 3.** The two-level directory structure

Each directory entry in main memory consists of Head\_1 and Head\_2, which consists of Data, Sta\_B and Poi. The head node points to first address of shared data. A data chain contains Pre\_P and Suc\_P is added to each data block in the private cache.

When a processor core sends read requests, the request first reaches the main memory directory. After matching the data block, the head node sends the data to the processor core, and the private cache of processor is added to the chain of head node.

When a processor core sends write requests, the request also reaches the main memory directory. If Sta\_B of head node is the state except “M”, firstly, all the data blocks connected to head node should be discarded, then doing write and modifying t Sta\_B, finally, the private cache with latest data is connected to the head node. If Sta\_B of the head node is “M”, it can be wrote directly after transferring the data, and it is not necessary to modify Sta\_B, after the completion of write invalidate, the private cache of processor core is connected to the chain which the head node is in.

To sum up, the optimization of cache consistency protocol should from the protocol itself and consider the importance of architecture.

### 3 Experimental Verification

In order to test the performance of architecture, it is compared with MESI protocol by selecting GEMS system multi-core simulator platform. The thesis uses SPLASH-2 centralized test program LU, Ocean, Radix, FFT and Water-SP to test the performance, as is shown in Table 1.

**Table 1.** Test procedures

Name of test procedure	Characteristic parameters
LU	512 * 512 matrix
Ocean	258 * 258 ocean
Radix	1M keys, 1024 radix
FFT	256K points
Water	512 molecules

As is shown in Fig. 4, based on the running time of the five test procedures in MESI protocol environment, the unit of running time is CPU cycle. It can be concluded that the average running time of the test procedures in the architecture is 3.84% less than that in MESI. As a result, the architecture improves the efficient utilization of bus and system performance to a certain extent.

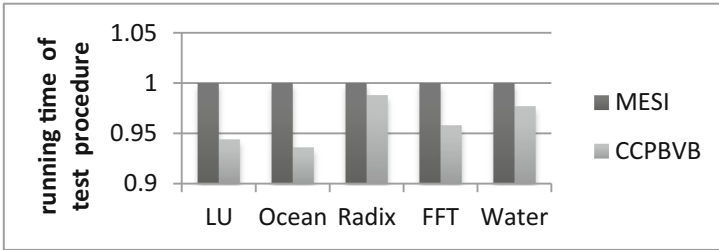


Fig. 4. Comparison of the running time of test procedures

## 4 Conclusion

Cache consistency problem has become one of the hot issues in multi-core processor research. The paper summarizes the current problem of consistency protocol, the effective use of shared bus resources of bus listening consistency protocol is lack, its broadcast consistent transaction mechanism leads to inadequate use of resources. The directory-based consistency protocol has long access delay. Aimed to the shortcomings of these two consistency protocols, D-Cache virtual bus architecture model in the paper effectively solves the shortcomings of the bus effective utilization. The paper has some shortcomings, which will be further studied and resolved in the following scientific research work.

**Acknowledgments.** This work is supported by Research on Compiling Technology Based on FPGA Reconfigurable Hybrid System (No. 61003036). The authors would like to thank all of the co-authors of this work.

## References

1. Hsia, A., Chen, C.W., Liu, T.J.: Energy-efficient synonym data detection and consistency for virtual cache. *Microprocess. Microsyst.* **40**(C), 27–44 (2016)
2. Selvin, L.S., Palanichamy, Y.: Push-pull cache consistency mechanism for cooper caching in mobile ad hoc environments. **24**(5), 3459–3470 (2016)
3. Guo, S., Wang, H., et al.: Hierarchical cache directory for CMP. *J. Comput. Sci. Technol.* **25**(2), 246–256 (2010)
4. Li, G.: Research on Cache Consistency Model in On-chip Multiprocessor Architecture. University of Science and Technology of China, pp. 57–65 (2013)
5. Shu, J., Lu, Y., Zhang, J., et al.: Research study of storage system technology based on nonvolatile memory. *Sci. Technol. Rev.* (14) (2016)