METHOD AND APPARATUS FOR DECREASING THREAD SWITCH LATENCY IN A MULTITHREAD PROCESSOR

ABSTRACT

The method and apparatus for decreasing thread switch latency in a multithread processor stores instructions for an active thread in a primary instruction queue (10), and stores instructions for a dormant thread in a thread switch instruction queue (14). The active thread is the thread currently being processed by the multithread processor, while the dormant thread is a thread not currently being executed by the multithread processor. During execution of the active thread, instructions are dispatched from the primary instruction queue (10) for processing. When a thread switch occurs, instructions are dispatched from the thread switch instruction queue for execution. Simultaneously, instructions stored in the thread switch instruction queue (14) are transferred to the primary instruction queue (10). In this manner, the thread switch latency resulting from the amount of time to refill the primary instruction queue with instructions of the dormant thread is eliminated.

Fig. 2