Latency on the Time Processor Unit

By Sharon Darley
Austin, Texas

Introduction

Each time function on the time processor unit (TPU) is divided into several individual states that are executed by the microengine.

These states consist of micro instructions that are stored inside the TPU or stored in RAM if the TPU is running in emulation mode. Each state requires a certain amount of time to execute. Furthermore, states for one channel are not necessarily executed back-to-back. Other channels can be serviced between states.

Thus, there is a certain amount of latency involved in anything the TPU does. Worst-case latency is defined as the longest amount of time that can elapse between the end of execution of one function state on a particular channel and the beginning of the next function state. The Time Processor Unit Reference Manual, Motorola document order number TPURM/AD, deals with worst-case latency calculations in detail.

This engineering bulletin discusses the factors involved in latency calculations as well as some of the effects of TPU latency.
Factors Involved in Latency Calculations

Several factors are involved in making latency calculations.

- Probably the most important is the number of channels that are active.

Since all 16 TPU channels share the same execution unit, only one channel can be serviced at a time and performance varies depending on the number of channels active and the frequency at which they request service.

- A second factor involved is the time required to service each state that is executed.

Each micro instruction takes two CPU clocks to execute. State execution times are listed in each individual function note.

- The TPU also requires a certain amount of setup time before it can service a channel.

There is always a 10-clock time slot transition time between the servicing of any two channels. In addition, once all channels on a particular priority level requesting service have been granted service, a 4-clock NOP is added to the time slot transition time. This means that if only one channel is enabled, 14 CPU clocks will elapse between the servicing of each state.

- Another factor in determining TPU latency is the RAM collision rate (RCR).

RCR is the percentage of time during which both the TPU and the CPU try to access the parameter RAM. Sometimes, the CPU may wait for the TPU to finish accessing the parameter RAM. At other times, the TPU may stall during a parameter RAM access while waiting for the CPU to finish accessing the RAM. A stall can take up to two CPU clocks.

Usually, the RCR is quite low, sometimes even 0. An RCR of about 0.1 is a common estimation. The estimation of the TPU stall time is equal to the number of RAM accesses in the longest state * RCR * 2 CPU clocks.
The number of RAM accesses in each state is provided in each individual TPU function note in the TPU manual.

Although the time to enter an interrupt service routine is not included in worst-case latency calculations, it is often of interest. The time to enter an interrupt routine is approximately 30 CPU clocks plus the time of the longest instruction that might be executing when the interrupt is requested.

**Effects of TPU Latency**

TPU latency affects the performance that can be obtained on each channel. In particular, it affects anything that requires calculations or the updating of parameters between two states. For example, latency determines the minimum amount of time that can elapse between two consecutive edges on an output channel. However, it does not affect hardware-controlled events, such as matches.

Each channel has its own capture register, match enable register, and greater-than-or-equal comparator. Thus, once a channel is initialized to capture an event, the hardware will record the event when it occurs, regardless of whether or not other channels are currently being serviced. The TCR time of the event will be recorded accurately. Likewise, once the microcode has set up a match (for example, for an output transition), that match will occur at the proper time regardless of whatever else the TPU is doing.

Anything that requires the execution of microcode will be affected by latency. This mainly includes calculations (including those to set up matches) and updating of the parameter RAM. Each system designer must determine how latency will affect the particular system in question and what the minimum system requirements are.

There is no single formula that determines the exact latency time that will be seen by each channel in every system. General worst case times are best determined by probability and statistics.

The *Time Processor Unit Reference Manual* presents an effective method for determining worst-case latencies.