

Build a Super Simple Tasker

By Miro Samek and Robert Ward

[Embedded Systems Design](#)

(07/12/06, 11:26:00 AM EDT)

Almost all embedded systems are event-driven; most of the time they wait for some event such as a time tick, a button press, a mouse click, or the arrival of a data packet. After recognizing the event, the systems react by performing the appropriate computation. This reaction might include manipulating the hardware or generating secondary, "soft" events that trigger other internal software components. Once the event-handling action is complete, such *reactive systems* enter a dormant state in anticipation of the next event.¹

Ironically, most real-time kernels or RTOSes for embedded systems force programmers to model these simple, discrete event reactions using tasks structured as continuous endless loops. To us this seems a serious mismatch--a disparity that's responsible for much of the familiar complexity of the traditional real-time kernels.

In this article we'll show how matching the discrete event nature typical of most embedded systems with a simple run-to-completion (RTC) kernel or "tasker" can produce a cleaner, smaller, faster, and more natural execution environment. In fact, we'll show you how (if you model a task as a discrete, run-to-completion action) you can create a prioritized, fully preemptive, deterministic real-time kernel, which we call *Super Simple Tasker (SST)*, with only a few dozen lines of portable C code.²

Such a real-time kernel is not new; similar kernels are widely used in the industry. Even so, simple RTC schedulers are seldom described in the trade press. We hope that this article provides a convenient reference for those interested in such a lightweight scheduler. But more importantly, we hope to explain why a simple RTC kernel like SST is a perfect match for execution systems built around state machines including those based on advanced UML statecharts. Because state machines universally assume RTC execution semantics, it seems only natural that they should be coupled with a scheduler that expects and exploits the RTC execution model.

We begin with a description of how SST works and explain why it needs only a single stack for all tasks and interrupts. We then contrast this approach with the traditional real-time kernels, which gives us an opportunity to re-examine some basic real-time concepts. Next, we describe a minimal SST implementation in portable ANSI C and back it up with an executable example that you can run on any x86-based PC. We conclude with references to an industrial-strength single-stack kernel combined with an open-source state machine-based framework, which together provide a deterministic execution environment for UML state machines. We'll assume that you're familiar with basic real-time concepts, such as interrupt processing, context switches, mutual exclusion and blocking, event queues, and finite state machines.

Preemptive multitasking with a single stack

Conventional real-time kernels maintain relatively complex execution contexts (including separate stack spaces) for each running thread or task, as explained in the excellent book *MicroC/OS-II: The Real-Time Kernel* by Jean Labrosse.³ Keeping track of the details of these contexts and switching among them requires lots of bookkeeping and sophisticated mechanisms to implement the context switch magic. The kernel we'll describe can be ultra simple because it doesn't need to manage multiple stacks and all of their associated bookkeeping.

By requiring that all tasks run to completion and enforcing fixed-priority scheduling, we can instead manage all context information using the machine's natural stack protocol. Whenever a task is preempted by a higher-priority task, the SST scheduler uses a regular C-function call to build the higher-priority task context on top of the preempted-task context. Whenever an interrupt preempts a task, SST uses the already established interrupt stack frame on top of which to build the higher-priority task context, again using a regular C-function call. This simple form of context management is adequate because we're insisting that every task, just like every interrupt, runs to completion. Because the preempting task must also run to completion, the lower-priority context will never be needed until the preempting task (and any higher-priority tasks that might preempt it) has completed--at which time the preempted task will naturally be at the top of the stack, ready to be resumed.

The first consequence of this execution profile is that an SST task cannot be an endless loop, unlike most tasks in traditional kernels.³ Instead, an SST task is a regular C-function that runs to completion and returns, thus removing itself from the execution stack as it completes. An SST task can be activated only by an ordinary C-function call from the SST scheduler and always returns to the scheduler upon completion. The SST scheduler itself is also an ordinary C-function that's called each time a preemptive event occurs and that runs to completion and returns once all tasks with priorities higher than the preempted task have completed.

The second consequence of the SST execution profile is that events associated with lower-priority tasks must be queued until the higher-priority tasks complete. The scheduling algorithm will work with a wide range of prioritization mechanisms; here we assume the simplest. For the sake of this discussion we'll assume a priority numbering scheme in which SST tasks have priorities numbered from 1 to **SST_MAX_PRIO**, inclusive, and a higher number represents a higher urgency. We reserve the priority level 0 for the SST idle loop, which is the only part of SST structured as an endless loop.

Simply calling a task function for every preemption level may seem too naive to work, but the analysis in the next section will show that it works perfectly for exactly the same reason that a prioritized hardware-interrupt system works.

Prioritization of tasks and interrupts in SST

It's interesting to observe that most prioritized interrupt controllers (for example, the 8259A inside the x86-based PC, the AIC in AT91-based ARM MCUs from Atmel, the VIC in LPC2xxx MCUs from Philips, the interrupt controller inside M16C from Renesas, and many others) implement in hardware the exact same asynchronous scheduling policy for interrupts as SST implements in software for tasks. In particular, any prioritized interrupt controller allows only higher-priority interrupts to preempt the currently active interrupt. All interrupts must run to completion and cannot block. All interrupts nest on the same stack.

In the SST execution model, tasks and interrupts are nearly symmetrical: both tasks and interrupt service routines are one-shot, run-to-completion functions. In fact, SST views interrupts very much like tasks of "super high" priority, as shown in Figure 1, except that interrupts are prioritized in hardware (by the interrupt controller), while SST tasks are prioritized in software.

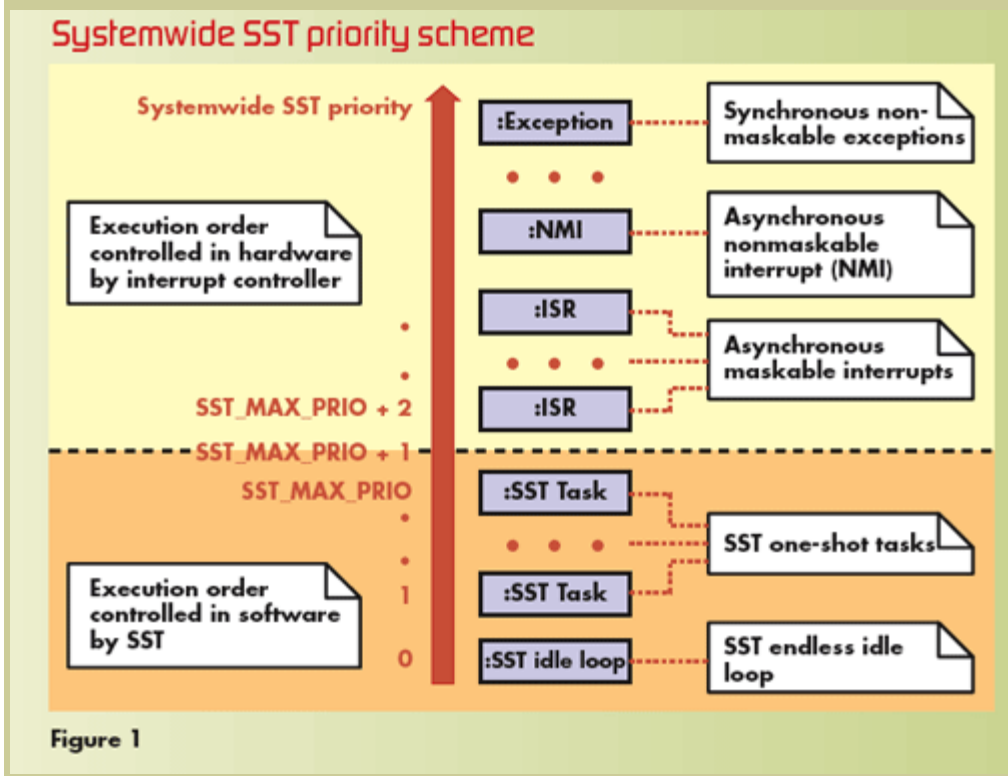


Figure 1

Synchronous and asynchronous preemptions

As a fully preemptive kernel, SST must ensure that at *all times* the CPU executes the highest-priority task that's ready to run. Fortunately, only two scenarios can lead to readying a higher-priority task:

A lower-priority task posts an event to a higher-priority task:

SST must immediately suspend the execution of the lower-priority task and start the higher-priority task. We call this type of preemption *synchronous preemption* because it happens synchronously with posting an event to the task's event queue.

An interrupt posts an event to a higher-priority task than the interrupted task:

Upon completion of the interrupt service routine (ISR), the SST must start execution of the higher-priority task instead of resuming the lower-priority task. We call this type of preemption **asynchronous preemption** because it can happen any time interrupts are not explicitly locked.

Figure 2 illustrates the synchronous preemption scenario caused by posting an event from a low-priority task to a high-priority task. Initially a low-priority task is executing (1). At some point during normal execution, the low-priority task posts an event to a high-priority task, thus making the high-priority task ready for execution. Posting the event engages the SST

scheduler (2). The scheduler detects that a high-priority task has become ready to run, so the scheduler calls (literally a simple C-function call) the high-priority task (3). Note that because the SST scheduler launches the task on its own thread, the scheduler doesn't return until after the higher-priority task completes. The high-priority task runs (4), but at some time it too may post an event to other tasks. If it posts to a lower-priority task than itself (5), the SST scheduler will again be invoked (again a simple C call), but will return immediately when it doesn't find any ready tasks with a higher priority than the current task. When this second call to the scheduler returns, the high-priority task runs to completion (6) and naturally returns to the SST scheduler (7). The SST scheduler checks once more for a higher-priority task to start (8), but it finds none. The SST scheduler returns to the low-priority task, which continues (9).

Synchronous preemption by a high-priority task

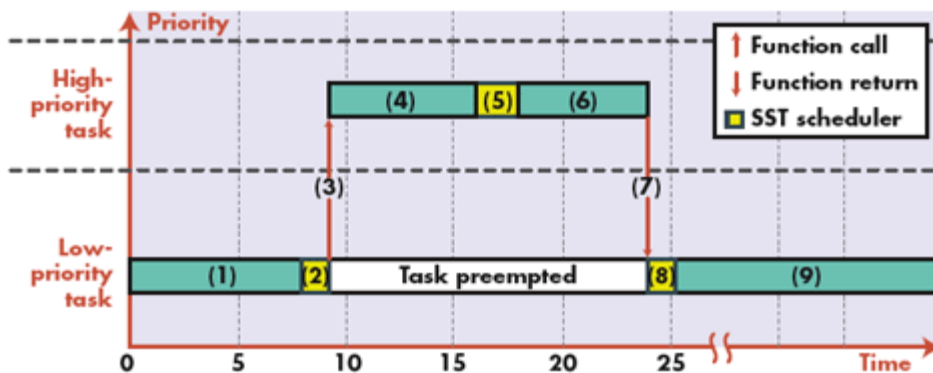


Figure 2

Obviously, synchronous preemptions are not limited to only one level. If the high-priority task were to post an event to a still higher-priority task at point (5) of Figure 2, the high-priority task would be synchronously preempted and the scenario would recursively repeat itself at a higher level of priority.

Figure 3 illustrates the asynchronous preemption scenario caused by an interrupt. Initially a low-priority task is executing and interrupts are unlocked (1). An asynchronous event interrupts the processor (2). The processor immediately preempts any executing task and starts executing the ISR (interrupt service routine). The ISR executes the SST-specific entry (3), which saves the priority of the interrupted task on the stack and raises the current priority to the ISR level. The ISR performs its work (4) which includes posting an event to the high-priority task (5). Posting an event engages the SST scheduler, but the current priority (that of the ISR) is so high that the scheduler returns immediately, not finding any task with priority higher than an interrupt. The ISR continues (6) and finally executes the SST-specific exit (7). The exit code sends the end-of-interrupt (EOI) instruction to the interrupt controller, restores the saved priority of the interrupted task, and invokes the SST scheduler (8). Now, the scheduler detects that the high-priority task is ready to run, so it enables interrupts and calls the newly ready, high-priority task (9). Note that the SST scheduler doesn't return. The high-priority task runs to completion (10) unless it also gets interrupted. After completion, the high-priority task naturally returns to the scheduler (11). The scheduler checks for more higher-priority tasks to execute (12) but doesn't find any and returns. The original interrupt returns to the low-priority task (13), which has been asynchronously preempted all that time. Note that the interrupt return (13) matches the interrupt call (2). Finally, the low-priority task runs to completion (14).

Asynchronous preemption by an interrupt and a high-priority task

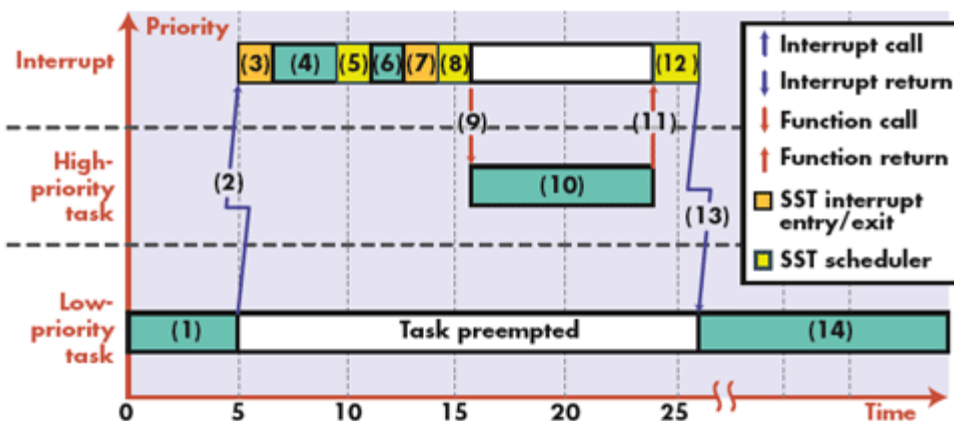


Figure 3

It's important to point out that conceptually the interrupt handling ends in the SST-specific interrupt exit (8), even though the interrupt stack frame still remains on the stack and the *IRET* instruction has not yet executed. (The *IRET* instruction is

understood here generically to mean a specific CPU instruction that causes hardware interrupt return.) The interrupt ends because the EOI instruction is issued to the interrupt controller and the interrupts get re-enabled inside the SST scheduler. Before the EOI instruction, the interrupt controller allows only interrupts of higher priority than the currently serviced interrupt. After the EOI instruction and subsequent call to the SST scheduler, interrupts get unlocked and the interrupt controller allows all interrupt levels, which is exactly the behavior expected at the task level.

Consequently, asynchronous preemption is not limited to only one level. The high-priority task runs with interrupts unlocked, shown as segment (10) in Figure 3, so it too can be asynchronously preempted by an interrupt, including the same-level interrupt that launched the task at point (9). If the interrupt posts an event to a still higher-priority task, the high-priority task will also be asynchronously preempted and the scenario will recursively repeat itself at a higher-level of priority.

SST versus traditional kernels

By managing all contexts in a single stack, SST can run with significantly less RAM than a typical blocking kernel. Because tasks don't have separate stacks, no unused private stack space is associated with suspended tasks. SST task switches also tend to incur less execution overhead; a traditional kernel doesn't distinguish between the synchronous and asynchronous preemptions and charges you a uniformly high price for all context switches. Finally, SST uses much simpler and smaller task control blocks (TCBs) for each task.

Because of this simplicity, context switches in SST (especially the synchronous preemptions) can involve much less stack space and CPU overhead than under any traditional kernel. But even the asynchronous preemptions in SST end up typically using significantly less stack space and fewer CPU cycles. Here's why.

Upon an interrupt, a traditional kernel must establish a strictly defined stack frame on each private stack into which it saves all CPU registers. Unlike SST, which can exploit the compiler's natural register-management strategy, a traditional kernel must be prepared to restore all registers when it resumes the preempted task. Often the construction and destruction of these stack frames must be programmed in assembly, because a traditional kernel cannot leave the context switch magic to the C compiler. In contrast, SST doesn't really care about any particular stack frame or whether the registers are stored immediately upon the ISR entry or stepwise, as needed. The only relevant aspect is that the CPU state be restored exactly to the status before the interrupt, but it's irrelevant *how* this happens. This means that the compiler-generated interrupts that most embedded C compilers support are typically adequate for SST, but are often inadequate for traditional kernels. Not only does SST not require any assembly programming, but in this case the compiler-generated interrupt entry and exit code is often superior to custom assembly, because the C compiler is in a much better position to globally optimize interrupt stack frames for specific ISRs. In this respect, SST allows you to take advantage of the C compiler's capabilities, something a traditional kernel can't do.

The last point is perhaps best illustrated by an example. All C compilers for ARM processors, for instance, adhere to the ARM Procedure Call Standard (APCS) that prescribes which registers must be preserved across a C-function call and which can be clobbered. The C-compiler-generated ISR entry saves initially only the registers that might be clobbered in a C function, which is only about half of all ARM registers. The rest of the registers get saved later, inside C functions invoked from the ISR, if and only if such registers are actually used. This example of a context save occurring in several steps is perfectly suited to the SST. In contrast, a traditional kernel must save *all* ARM registers in one swoop upon ISR entry, and if the assembly ISR "wrapper" calls C functions (which it typically does) many registers are saved again. Needless to say, such policy requires more RAM for each private stack and more CPU cycles (perhaps by factor of two) than SST.

On a side note, we would like to emphasize that all traditional blocking kernels in the industry share the problems we've described. We refer here to Labrosse's book *MicroC/OS-II* not because *MicroC/OS-II* has any special deficiency in this regard but because his book is an excellent treatment of traditional kernels and is familiar to many embedded systems programmers.³

Sidebar 2: Interrupt duration in traditional kernels and SST

If you have some experience with traditional preemptive kernels, you'll notice that SST seems to encourage programmers to violate time-honored advice about what to put in an ISR.

Most of us have been taught that an ISR should be as short as possible and that the main work should always be done at the task level. In SST, however, *everything* appears to be done at the ISR context and nothing at the task context. It seems to be backwards.

But really, the problem centers on the distinction between an interrupt and a task. SST forces us to revise the naive understanding of interrupt duration as beginning with saving interrupt context on a stack and ending with restoring the interrupt context followed by **IRET** because, as it turns out, this definition is problematic even for traditional kernels.

Figure 4 shows the per-task data structures maintained by a traditional preemptive kernel where each task has its own stack and a task control

block (TCB)³ for the execution context of each task. In general, an interrupt handler stores the interrupt context on one task's stack and restores the context, from *another* task's stack. After restoring a task's context into the CPU registers, the traditional scheduler always issues the IRET instruction. The key point is that the interrupt context remains saved on the preempted task's stack, so the saved interrupt context *outlives* the duration of the interrupt handler. Therefore defining the duration of an interrupt from saving the interrupt context to restoring the context is problematic.

Per-task data structures maintained by a traditional preemptive kernel [3]

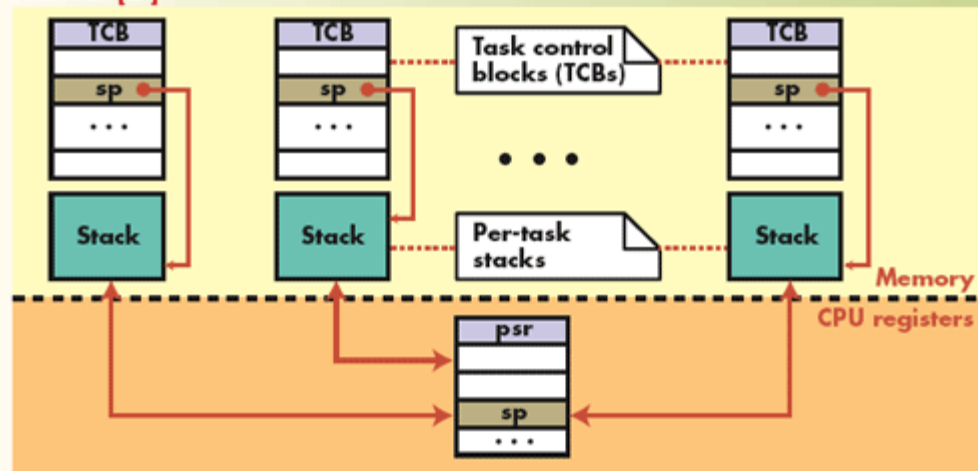


Figure 4

The situation is not really that much different under a single-stack kernel, such as SST. An ISR stores the interrupt context on the stack, which happens to be common for all tasks and interrupts. After some processing, the ISR calls the scheduler, which internally unlocks interrupts. If no higher-priority tasks are ready to run, the scheduler exits immediately, in which case the ISR restores the context from the stack and returns to the original task exactly at the point of preemption. Otherwise, the SST scheduler calls a higher-priority task and the interrupt context remains saved on the stack, just as in the traditional kernel.

The point here is that the ISR is defined from the time of storing interrupt context to the time of sending the EOI (End Of Interrupt) command to the interrupt controller followed by invoking the scheduler that internally enables interrupts, not necessarily to the point of restoring the interrupt context. This definition is more precise and universal, because under any kernel the interrupt context remains stored on one stack or another and typically outlives the duration of an interrupt's processing. Of course, you should strive to keep the ISR code to a minimum, but in SST, you should remember that the final call to the SST scheduler is conceptually not a part of the ISR, but rather it constitutes the task level.

The definition of ISR duration is not purely academic, but has important practical implications. In particular, debugging at the ISR level can be much more challenging than debugging at the task level. Indeed, in SST debugging the ISR to the point of sending the EOI command to the interrupt controller and invoking the SST scheduler is as hard as in any traditional kernel, especially when debugging is accomplished through a ROM monitor. However, debugging an SST task is as easy as debugging

the **main()** task, even though a task is called indirectly from the ISR. This is because the SST scheduler launches each task with interrupts unlocked at the CPU level and with the interrupt controller set to enable all interrupts.

SST implementation

We first present a minimal standalone implementation of SST. The presented code is portable ANSI C, with all CPU and compiler-specific parts clearly separated out. However, the implementation omits some features that might be needed in real-life applications. The main goal at this point is to clearly demonstrate the key concepts while letting you execute the code on any Windows-based PC. We've compiled the example with the legacy Turbo C++ 1.01, available as a free download from the Borland Museum.⁴

The example The SST example demonstrates the multitasking and preemption capabilities of SST. This example, shown in Figure 5, consists of three tasks and two ISRs. The clock tick ISR produces a "tick event" every 5ms, while the keyboard ISR produces a "key event" every time a key is depressed or released, and at the auto-repeat rate of the keyboard when a key is depressed and held. The two "tick tasks": **tickTaskA()** and **tickTaskB()** receive the "tick events" and place a letter A or B, respectively, at a random location in the right-hand panel of the screen. The keyboard task **kbdTask()**, with the priority between **tickTaskA()** and **tickTaskB()**, receives the scan codes from the keyboard and sends "color events" to the tick tasks, which change the color of the displayed letters in response. Also, the **kbdTask()** terminates the application when you depress the Esc key.

The SST example running in the Turbo C++ IDE. Note that the SST code size is just 421 bytes

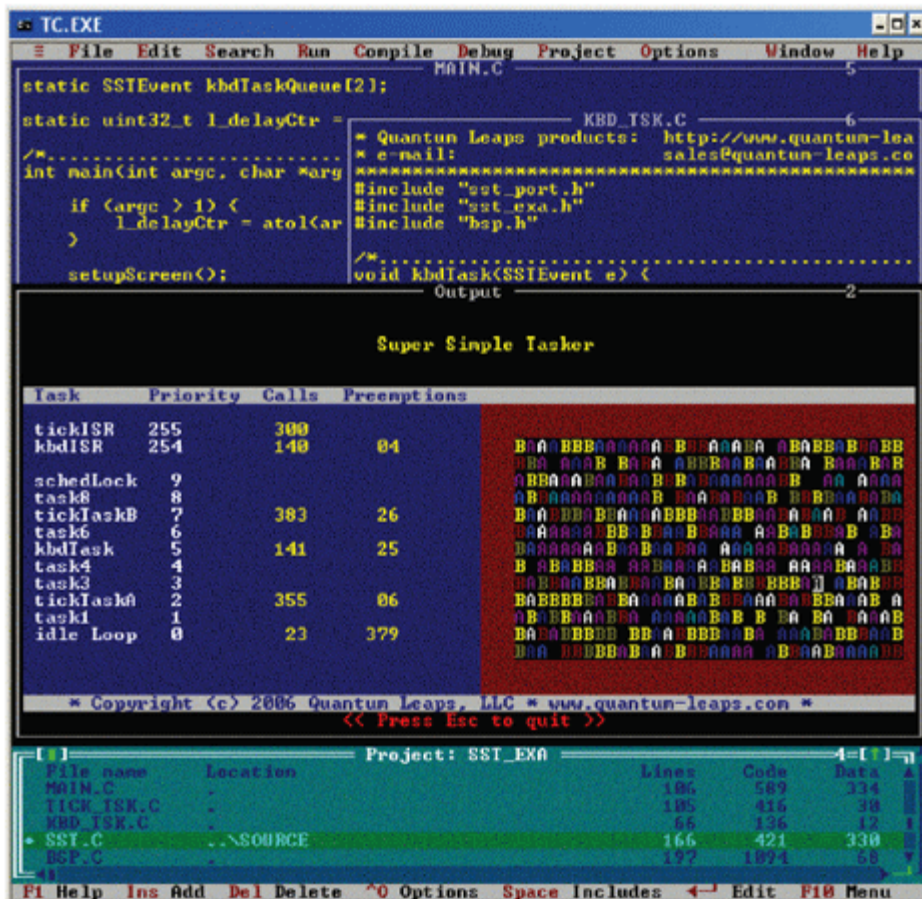


Figure 5

The left-hand side of the screen in Figure 5 shows the basic statistics of the running application. The first two columns display the task names and priorities. Note that there are many unused priority levels. The "Calls" column shows the number of calls with 3-digit precision. The last "Preemptions" column shows the number of *asynchronous* preemptions of a given task or ISR.

The SST example application intentionally uses two independent interrupts (the clock tick and keyboard) to create asynchronous preemptions. To further increase the probability of an interrupt preempting a task, and of an interrupt preempting an interrupt, the code is peppered with calls to a **busyDelay()** function, which extends the run-to-completion

time in a simple counted loop. You can specify the number of iterations through this loop by a command line parameter to the application. You should be careful not to go overboard with this parameter, though, because larger values will produce a task set that is not schedulable and the system will (properly) start losing events.

The following subsections explain the SST code and the application structure. The complete SST source code consists of the header file `sst.h` located in the `include\` directory and the implementation file `sst.c` located in the `source\` directory. The example files are located in the `example\` directory, which also contains the Turbo C++ project file to build and debug the application. (NOTE: because the standard keyboard ISR is replaced by the custom one, debugging this application with the Turbo C++ IDE might be difficult.)

Critical sections in SST

SST, just like any other kernel, needs to perform certain operations indivisibly. The simplest and most efficient way to protect a section of code from disruptions is to lock interrupts on entry to the section and unlock the interrupts again on exit. Such a section of code is called a *critical section*.

Processors generally provide instructions to lock/unlock interrupts, and your C compiler must have a mechanism to perform these operations from C. Some compilers allow you to include inline assembly instructions in your C source. Other compilers provide language extensions or at least C-callable functions to lock and unlock interrupts from C.

To hide the actual implementation method chosen, SST provides two macros to lock and unlock interrupts. Here are the macros defined for the Turbo C++ compiler:

```
reformat this as:
#define SST_INT_LOCK() \
    disable()
#define SST_INT_UNLOCK() \
    enable()
```

In the minimal SST version, we assume the simplest possible critical section: one that unconditionally locks interrupts upon entry and unconditionally unlocks interrupts upon exit. Such simple critical sections should never nest, because interrupts will always be unlocked upon exit from the critical section, regardless of whether they were locked or unlocked before the entry. The SST scheduler is designed to never nest critical sections, but you should be careful when using macros `SST_INT_LOCK()` and `SST_INT_UNLOCK()` to protect your own critical sections in the applications. You can avoid this limitation by using smarter (though somewhat more expensive) code to lock and unlock interrupts. Please note, however, that the inability to nest critical sections does not necessarily mean that you can't nest interrupts. On processors equipped with an internal or external interrupt controller, such as the 8259A PIC in the x86-based PC, or the AIC in the AT91 ARM microcontroller, you can unlock the interrupts inside ISRs at the processor level, thus avoiding nesting of the critical section inside ISRs, and let the interrupt controller handle the interrupt prioritization and nesting before they even reach the CPU core.

Interrupt processing in SST

One of the biggest advantages of SST is the simple interrupt processing, which is actually not much more complicated with SST than it is in a simple "super-loop" (a.k.a., main+ISRs). Because SST doesn't rely in any way on the interrupt stack frame layout, with most embedded C compilers, the ISRs can be written entirely in C. One notable difference between a simple "super-loop" and SST ISRs is that SST requires the programmer to insert some simple actions at each ISR entry and exit. These actions are implemented in SST macros `SST_ISR_ENTRY()` and `SST_ISR_EXIT()`. The code snippet in Listing 1 shows how these macros are used in the clock tick and keyboard ISRs from the example application defined in the file `example\bsp.c`.

Listing 1: SST ISRs from the example application

```
void interrupt tickISR() { /* every ISR is entered with interrupts locked */
    uint8_t pin; /* temporary variable to store the initial priority */
    SST_ISR_ENTRY(pin, TICK_ISR_PRIO);

    SST_post(TICK_TASK_A_PRIO, TICK_SIG, 0); /* post the Tick to Task A */
    SST_post(TICK_TASK_B_PRIO, TICK_SIG, 0); /* post the Tick to Task B */

    SST_ISR_EXIT(pin, outportb(0x20, 0x20));
}
/*.....*/
void interrupt kbdISR() { /* every ISR is entered with interrupts locked */
    uint8_t pin; /* temporary variable to store the initial priority */
    uint8_t key = inport(0x60); /* get scan code from the 8042 kbd controller */
    SST_ISR_ENTRY(pin, KBD_ISR_PRIO);

    SST_post(KBD_TASK_PRIO, KBD_SIG, key); /* post the Key to the KbdTask */

    SST_ISR_EXIT(pin, outportb(0x20, 0x20));
}
```

Note in Listing 1, the compiler specific keyword "**interrupt**", which directs the Turbo-C compiler to synthesize appropriate

context saving, restoring, and interrupt return prologues and epilogues. Please also note the SST interrupt-entry and interrupt-exit macros at the beginning and end of each ISR. (If the interrupt source requires clearing, this should be done before calling `SST_ISR_ENTRY()`). The macros `SST_ISR_ENTRY()` and `SST_ISR_EXIT()` are defined in the `includes\sst.h` header file as shown in Listing 2. (The `do..while(0)` loop around the macros is only for syntactically correct grouping of the instructions.)

Listing 2: Definition of the SST interrupt entry/exit macros

```
#define SST_ISR_ENTRY(pin_, isrPrio_) do { \
    (pin_) = SST_currPrio_; \
    SST_currPrio_ = (isrPrio_); \
    SST_INT_UNLOCK(); \
} while (0)

#define SST_ISR_EXIT(pin_, EOI_command_) do { \
    SST_INT_LOCK(); \
    (EOI_command_); \
    SST_currPrio_ = (pin_); \
    SST_schedule(); \
} while (0)
```

The `SST_ISR_ENTRY()` macro is invoked with interrupts locked and performs the following three steps: (1) Saves the initial SST priority into the stack variable 'pin_' (2) Sets the current SST priority to the ISR level (3) Unlocks the interrupts to allow interrupt nesting. The `SST_ISR_EXIT()` macro is invoked with interrupts unlocked and performs the following four steps: (1) Locks the interrupts (2) Writes the EOI command to the interrupt controller (for example, `outportb(0x20, 0x20)` writes the EOI to the master 8259A PIC) (3) Restores the initial SST priority (4) Calls the SST scheduler, which performs the "asynchronous preemption," if necessary.

The task control blocks
 Like other real-time kernels, SST keeps track of tasks in an array of data structures called task control blocks (TCBs). Each TCB contains such information as the pointer to the task function, the task mask (calculated as $(1 \ll (\text{priority} - 1))$), and the event queue associated with the task. The TCB takes only 8 to 10 bytes, depending on the size of the function pointer. Additionally, you need to provide an event queue buffer of the correct size when you create a task. The TCB used here is optimized for a small, fixed number of priority levels and simple events--restrictions that make sense in a classic embedded environment. Neither of these limitations, however, is required by the SST scheduling algorithm.

Posting events to tasks
 In this minimal version of SST, events are represented as structures containing two byte-size elements: an event type identifier (for example, the key-press occurrence), and the parameter associated with the occurrence (for example, the scan code of the key). The events are stored in event queues organized as standard ring buffers. SST maintains the status of all event queues in the variable called the SST ready-set. As shown in Figure 6, the SST ready-set `SST_readySet_` is just a byte, meaning that this implementation is limited to eight priority levels. Each bit in the `SST_readySet_` represents one SST task priority. The bit number 'n' in the `SST_readySet_` is 1 if the event queue of the task of priority 'n+1' is not empty (bits are numbered 0..7). Conversely, bit number 'm' in `SST_readySet_` is 0 if the event queue of the task of priority 'm+1' is empty, or the priority level 'm+1' is not used.

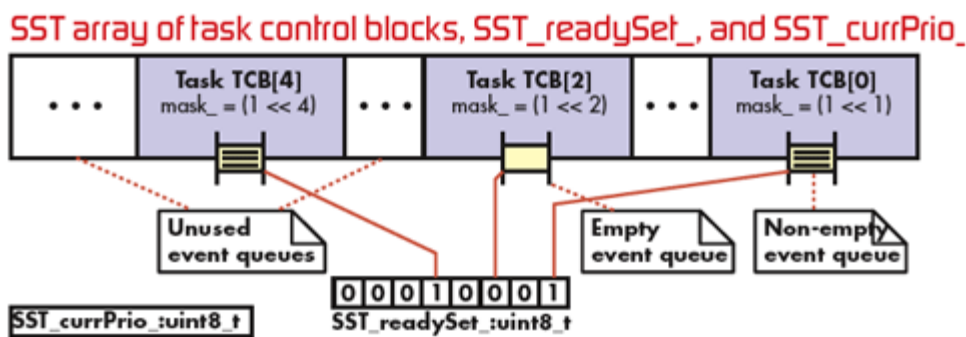


Figure 6

Listing 3 shows the event posting function `SST_post()`, which uses a standard ring buffer algorithm (FIFO). If the event is inserted to an empty queue (1), the corresponding bit in `SST_readySet_` is set (2), and the SST scheduler is invoked to check for the "synchronous preemption" (3).

Listing 3: Event posting in SST

```
uint8_t SST_post(uint8_t prio, SSTSignal sig, SSTParam par) {
    TaskCB *tcb = &l_taskCB[prio - 1];
    SST_INT_LOCK();
    if (tcb->nUsed__ < tcb->end__) {
        tcb->queue__[tcb->head__].sig = sig; /* insert the event at the head */
        tcb->queue__[tcb->head__].par = par;
        if ((++tcb->head__) == tcb->end__) {
            tcb->head__ = (uint8_t)0; /* wrap the head */
        }
    }
}
```



```

(1)         if ((++tcb->nUsed_) == (uint8_t)1) { /* the first event? */
(2)             SST_readySet_ |= tcb->mask_; /* insert task to the ready set */
(3)             SST_schedule_(); /* check for synchronous preemption */
        }
        SST_INT_UNLOCK();
        return (uint8_t)1; /* event successfully posted */
    }
    else {
        SST_INT_UNLOCK();
        return (uint8_t)0; /* queue full, event posting failed */
    }
}

```

The SST scheduler

The SST scheduler is a simple C-function `SST_schedule_()` whose job is to efficiently find the highest-priority task that is ready to run and run it, if its priority is higher than the currently serviced SST priority. To perform this job, the SST scheduler uses the already described `SST_readySet_` and the current priority level `SST_currPrio_` shown in Figure 6. Both variables `SST_currPrio_` and `SST_readySet_` are always accessed in a critical section to prevent data corruption. (See also the `SST_ISR_ENTRY/SST_ISR_EXIT` macros.) Listing 4 shows the complete SST scheduler implementation. (Yes, it really is that small.) The function `SST_schedule_()` must be called with interrupts locked and returns also with interrupts locked.

Listing 4: The SST scheduler

```

void SST_schedule_(void) {
    static uint8_t const log2Lkup[] = { /* log-base-2 lookup table */
        0, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4,
        5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
        6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
        6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
        7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,
        7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,
        . . .
        8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8,
        . . .
    };
(1)     uint8_t pin = SST_currPrio_; /* save the initial priority */
        uint8_t p; /* the new priority */
        /* is the new priority higher than the initial? */
(2)     while ((p = log2Lkup[SST_readySet_]) > pin) {
            TaskCB *tcb = &l_taskCB[p - 1]; /* get the event out of the queue */
(3)         SSTEvent e = tcb->queue__[tcb->tail_];
            if ((++tcb->tail_) == tcb->end_) {
                tcb->tail_ = (uint8_t)0;
            }
            if ((--tcb->nUsed_) == (uint8_t)0) { /* is the queue becoming empty?*/
(4)         SST_readySet_ &= ~tcb->mask_; /* remove from the ready set */
            }
(5)         SST_currPrio_ = p; /* this becomes the current task priority */
(6)         SST_INT_UNLOCK(); /* unlock the interrupts */
(7)         (*tcb->task_)(e); /* call the SST task */
(8)         SST_INT_LOCK(); /* lock the interrupts for the next pass */
        }
(9)     SST_currPrio_ = pin; /* restore the initial priority */
}

```

Just as the hardware-interrupt controller does when servicing an interrupt, the SST scheduler starts by saving the initial priority into a stack variable '`pin`' (1). Next, this initial priority is compared to the highest priority of all tasks ready to run (computed from the SST ready-set `SST_readySet_`). This latter computation is efficiently implemented as a binary-logarithm lookup (more precisely $\log_2(x) + 1$), which delivers the bit number of the most-significant 1-bit in the `SST_readySet_` byte (2). If the new priority '`p`' is higher than the initial value, the scheduler needs to start the task of priority '`p`'. First, the scheduler removes the event from the tail of the queue (3), and if the queue becomes empty clears the corresponding bit in `SST_readySet_` (4). Subsequently, the current priority `SST_currPrio_` is raised to the new level '`p`' (5), the interrupts are unlocked (6) and the high-priority task is called (7). When the task completes and returns, the interrupts are locked (8), and the `while` loop again computes the highest-priority task ready to run based on the potentially changed `SST_readySet_` (2). The loop continues until no more tasks above the initial priority level '`pin`' are ready to run. Before the exit, the scheduler restores the current priority `SST_currPrio_` from the stack variable '`pin`' (9).

Mutual exclusion in SST

SST is a preemptive kernel, and as with all such kernels, you must be very careful with *any* resource sharing among SST

tasks. Ideally, SST tasks should not share any resources, limiting all intertask communication to events. This ideal situation allows you to program all SST tasks with purely sequential techniques, allowing SST to handle all the details of thread-safe event exchange and queuing. However, at the cost of increased coupling among tasks, you might choose to share selected resources. If you go this path, you must accept responsibility for synchronizing access to such resources (shared variables or devices). One option is to guard access to the shared resource with a critical section. This requires the programmer to consistently lock and unlock interrupts around each access. For very short accesses this might well be the most effective synchronization mechanism. However, SST also provides a more selective mechanism: a priority-ceiling mutex. Priority-ceiling mutexes are immune to priority inversions,⁵ but still allow hardware interrupts and higher-priority tasks to run as usual. The SST implementation of a priority-ceiling mutex is remarkably simple. Recall that the SST scheduler can only launch tasks with priorities higher than the initial priority with which the scheduler was entered. This means that temporarily increasing the current SST priority **SST_currPrio_** blocks any tasks with priorities lower than this priority "ceiling." This is exactly what a priority-ceiling mutex is supposed to do. Listing 5 shows the SST implementation.

Listing 5: Priority-ceiling mutex locking and unlocking

```
uint8_t SST_mutexLock(uint8_t prioCeiling) {
    uint8_t p;
    SST_INT_LOCK();
    p = SST_currPrio_;          /* save the original SST priority to return */
    if (prioCeiling > SST_currPrio_) {
        SST_currPrio_ = prioCeiling; /* set the SST priority to the ceiling */
    }
    SST_INT_UNLOCK();
    return p;
}
/*.....*/
void SST_mutexUnlock(uint8_t orgPrio) {
    SST_INT_LOCK();
    if (orgPrio < SST_currPrio_) {
        SST_currPrio_ = orgPrio; /* restore the saved priority to unlock */
        SST_schedule_(); /* the scheduler unlocks the interrupts internally */
    }
    SST_INT_UNLOCK();
}
}
```

Unlike the simple **INT_LOCK** macros, the SST mutex interface allows locks to be nested, because the original SST priority is preserved on the stack. Listing 6 shows how the mutex is used in the SST example to protect the non-reentrant DOS random number generator calls inside the clock-tick tasks **tickTaskA()** and **tickTaskB()**.

Listing 6: Priority-ceiling mutex used in the SST example to protect the non-reentrant random-number generator

```
void tickTaskA(SSTEvent e) {
    . . .
    uint8_t x, y;
    uint8_t mutex;          /* mutex object preserved on the stack */

    mutex = SST_schedLock(TICK_TASK_B_PRIO);          /* mutex lock */
    x = random(34); /* call to non-reentrant random number generator */
    y = random(13); /* call to non-reentrant random number generator */
    SST_schedUnlock(mutex);          /* mutex unlock */
    . . .
}
```

Starting multitasking and the SST idle loop

Listing 7 shows the **SST_run()** function, which is invoked from **main()** when the application transfers control to SST to start multitasking.

Listing 7: Starting SST multitasking

```
void SST_run(void) {
(1)    SST_start();          /* start ISRs */
(2)    SST_INT_LOCK();
(3)    SST_currPrio_ = (uint8_t)0; /* set the priority for the SST idle loop */
(4)    SST_schedule_();      /* process all events produced so far */
(5)    SST_INT_UNLOCK();

(6)    for (;;) {          /* the SST idle loop */
(7)        SST_onIdle();    /* invoke the on-idle callback */
    }
}
```

The **SST_run()** function calls the **SST_start()** callback (1), in which the application can configure and start interrupts (see file **examples\bsp.c**). After locking interrupts (2), the current SST priority **SST_currPrio_** is set to zero, which corresponds to the priority of the idle loop. (The SST priority is statically initialized to 0xFF.) After lowering the priority the SST scheduler is invoked (4) to process any events that might have accumulated during the task initialization phase. Finally, interrupts are unlocked (5), and **SST_run()** enters the SST *idle loop* (6), which runs when the CPU is not processing any of the one-shot SST tasks or interrupts. The idle loop continuously calls the **SST_onIdle()** callback (7), which the application can use to put the CPU into a power-saving mode.

Where to go from here

The minimal SST implementation presented in this article is remarkably powerful, given that it takes only 421 bytes of code (see Figure 5), 256 bytes of lookup tables, and several bytes of RAM per task (for the TCB and the event queue buffer) plus, of course, the stack. This might be all you need for smaller projects. When you tackle bigger systems, however, you will inevitably discover shortcomings in this implementation. From our experience, these limitations have little to do with the multitasking model and much to do with the infrastructure surrounding the kernel, such as the limited event parameter size, lack of error handling, and the rather primitive event-passing mechanism.

One thing that you'll surely discover when you go beyond toy applications suitable for publication in an article is that your task functions will grow to contain more and more conditional code to handle various modes of operation. For example, a user-interface task for a digital camera must react differently to button-press events when the camera is in the playback mode than to the same button-press events in the picture-taking mode. The best-known method of structuring such code is through a *state machine* that explicitly captures the different modes of operations as different states of the system.

And here is the most important significance of SST. The inherent run-to-completion (RTC) execution model responsible for the simplicity of this kernel perfectly matches the RTC execution semantics universally assumed in all state machine formalisms, including advanced UML statecharts. Simply put, SST and state machines are born for each other. We would even go so far as to suggest that if you're using any other type of real-time kernel for executing concurrent state machines, you are probably paying too much in ROM, RAM, and CPU usage.

While the full integration of SST into a generalized, state machine-based system is beyond the scope of this article, you can explore the fit yourself in an open-source system produced by one of the authors. You can download complete code, examples, and documentation from www.quantum-leaps.com. For the code and other supporting documentation about the implementation in this article, go to [Embedded.com's code archive](#).

Miro Samek is the founder and CTO of Quantum Leaps, LLC, a provider of real-time, state machine-based application frameworks for embedded systems. He is the author of *Practical Statecharts in C/C++* (CMP Books, 2002), has written numerous articles for magazines, and is a regular speaker at the Embedded Systems Conference. He welcomes contact at miro@quantum-leaps.com.

Robert Ward, in a former life, founded The C Users Journal and its sister publications. He is now a principal software engineer at Netopia, Inc, where he designs and builds Java-based servers to support web-based collaboration. You can reach him at ward@codecraftsman.com.

Endnotes:

1. Selic, Bran. "The Challenges of Real-Time Software Design," *Embedded Systems Programming*, October 1996. (Article not available online.)
2. Ward, Robert. "Practical Real-Time Techniques," Proceedings of the Embedded Systems Conference, San Francisco, 2003. [Article pdf](#)
3. Labrosse, Jean J. [MicroC/OS-II: The Real Time Kernel, 2nd Edition](#). CMP Books 2002.
4. Borland Developer Network, "Antique Software: Turbo C++ version 1.01" <http://bdn.borland.com/article/0,1410,21751,00.html>
5. Kalinsky, David. "Mutexes Prevent Priority Inversions," *Embedded Systems Programming*, August 1998. (Article not available online.)

Reader Response

In my experience building operating systems, "Run-to-Completion" and "fully-preemptive" are complete polar opposites. Code cannot "run to completion" if its execution is preempted!

Maybe you are using these terms differently somehow?

- Mike O'Dell
President
Compass Rose Labs LLC

Oakton, VA

Robert Ward responds: *Here we mean that tasks don't block waiting for any resource other than the CPU—and then only because some higher priority task owns it.*

The simple tasker described does reduce the overhead of the task switcher. However this reduction in overhead comes at the cost of increased state logic within each thread in order to control thread execution.

A traditional RTOS permits you to write simpler, straight line application code since thread execution pauses in place while waiting for events. The simple switcher requires additional state logic within each thread to control execution on each pass through the thread. In this respect the simple switcher requires application code more like what would be used in a round-robin scheduler

- Phil Ouellette
Senior Engineer
Mettler-Toledo, Inc.
Worthington, OH

Robert Ward responds: *This is exactly why we like to couple the switcher with applications written as statecharts. The statechart is the expression of the statefulness of the relationship between the application and its environment. The tasker is all that's necessary to support scheduling for the statechart. The statechart solution is a different expression for the algorithm than many are familiar with, but the statechart is arguably more natural (more isomorphic) to the problem in stimulus/response applications. Whether the traditional form produces "simpler" code, I think, is a highly subjective matter.*

What I miss in this article is what happens when a task must wait for a resource other than the priority ceiling mutex. Think of a timer, a message or an event. In this case the task can not run to completion and is preempted by another task with a lower priority. As far as I understand your mechanism does no longer work then since this can not be solved with synchronous function calls. The lower priority thread must continue where it originally has been interrupted. I think that a mix of your mechanism and an original system is ideal. Concluding, I think your system looks very good but is not able to deal with all possible situations and therefore on its own only suitable for a specific class of systems.

- Leon van Snippenberg
Software Architect TA (Technical Automation)
Atos Origin TA - IPS
Den Bosch, The Netherlands

Robert Ward responds: *Ahhh, but you are still trying to frame "waiting" using the traditional "block-on-resource" model. We would instead assume there is no active thread until the timer fires (it is, after all an interrupt) and that it would post the task that services it to the SST ready queue. SST would then schedule the service as soon as possible. The basic requirement is that the entire program must be written in an asynchronous fashion; each portion that depends on an external event (for example, a resource becoming ready) must be a separate SST "task." When the external event occurs, that portion runs. The "program" never waits within a task . . . waits happen between tasks.*

Again, that's what makes statecharts so natural to this environment. When expressed as a statechart, the program is naturally partitioned into non-blocking action sequences (SST tasks == statechart transitions) that are triggered by external events. The wait points (states) are clearly defined and encoded as data, so they don't need to tie up a thread.

I have used a technique similar to this for many years, except I do not allow tasks to preempt tasks. I use numerous smaller tasks instead of a few more complex tasks. ISRs are the only entities allowed to "preempt". The fixed-priority scheduler runs each task to completion in priority order. Tasks are called as functions in C. The thing that people enamored with preemptive schedulers often forget is that inter-task communications must be accounted for at both task and ISR level. You cannot expect the RTOS to do all your thinking for you. The simple scheduler referenced here has typically higher latencies, but since critical preprocessing is handled in the ISR, the system remains deterministic, which is the most important factor.

- Douglas Schmidt
Sr. Development Engineer
Thermo Electron Inc.
Minneapolis, MN

Miro Samek responds: *I agree that the simpler non-preemptive scheduler that you concisely describe here can share many techniques used in SST, such as the ready-set to make priority decisions at the run-to-completion (RTC) step boundaries. I also agree that a non-preemptive scheduler is adequate for many systems. In fact, I do use such a scheduler extensively in the "bare metal" implementation of a state-machine framework.*

However, the main point of this article is preemptive multitasking. SST, just like any fully preemptive kernel, decouples tasks in the time domain. With SST, the timing of higher-priority tasks is completely independent on lower-priority tasks. In other words, you can keep changing lower-priority tasks without having an impact on timeliness of the higher-priority tasks. (This is true exactly if you don't introduce coupling via mutexes or critical sections in the lower-priority tasks.)

In contrast, the task-level response of the simpler, non-preemptive scheduler is determined by the longest RTC step in the whole system. So changing any task, regardless of its priority, can have an impact on the timeliness of the rest of the application. Even though the simple non-preemptive scheduler is in many ways similar to SST, it is a solution of an entirely different class.

Traditionally, the price tag for fully preemptive multitasking has been multiple stacks, TCBs, and significant number of CPU cycles for context switches. SST shows how to reduce these costs in all these respects. For example, a long-time user of SST reported an 80% reduction in stack space compared with a traditional kernel running essentially the same application.

SST_schedLock and SST_schedUnlock

These appear to be equivalent to the mutex lock and unlock; was just wondering if they were from another part of the article that was edited or another article.

- Stephen Simmons
Sr. Systems Eng
ASRC
Kennedy Space Center, FL

Miro Samek responds: *Good catch. The function names SST_schedLock()/SST_schedUnlock() used in Listing 6 should be replaced with SST_mutexLock()/SST_mutexUnlock() described in the article. This name inconsistency reflects the names used in an earlier version of the SST code (the priority-ceiling mutex is equivalent to locking the scheduler to a given priority level).*

Please note that the SST code accompanying the article is correct. I apologize for the inconsistency in the text.

Miro, Robert,

Thanks, I really enjoyed reading this article. I thought some of this looked familiar, then I remembered that I hadn't finished my copy of Statecharts in C/C++ yet.

- Greg Feneis
Consultant
Thorough Calibration & Design
Palo Alto, CA

[Discuss This Article](#)

11 message(s). Last at: Sep 14, 2009 8:41:20 AM

Guest

commented on Nov 18, 2007 8:34:34 PM

Hi,
Perhaps I haven't understood the scheduler correctly, but the 256 byte table, log2Lkup[], seems like a waste of space. Yes, I know that it's const, but const doesn't work in AVR ATmegas, which have a Harvard architecture (separately addressed buses for code and data), where even consts are in data space. Yes, we can put it into program space: special AVR functions can fetch bytes from code space: but it's not clear how efficient they are.

At the moment in my current application(1), which is essentially a KSAM database using external serial Flash & FRAM, I have about 500 bytes of program Flash left, and 200 bytes of RAM, and if we need to put SST in...

The problem I'm facing is that a database update takes about 6 ms, and the device needs to service Comms requests at a much smaller latency (about 1ms).

What does using the table solve, as a problem? Understanding this could enable us to solve this using a smaller footprint, e.g. a nibble table, or scrap the table for something like an insertion

sort?

Regards,
Alf Lacis

http://au.geocities.com/lacis_alfredo

(1) current MCU, ATmega168 has 16 k Flash & 1/2 k RAM

oh, really?

fw engineer

commented on Nov 18, 2007 8:36:39 PM

(sorry, wasn't logged in)

Hi,

Perhaps I haven't understood the scheduler correctly, but the 256 byte table, `log2Lkup[]`, seems like a waste of space. Yes, I know that it's const, but const doesn't work in AVR ATmegas, which have a Harvard architecture (separately addressed buses for code and data), where even consts are in data space. Yes, we can put it into program space: special AVR functions can fetch bytes from code space: but it's not clear how efficient they are.

At the moment in my current application(1), which is essentially a KSAM database using external serial Flash & FRAM, I have about 500 bytes of program Flash left, and 200 bytes of RAM, and if we need to put SST in...

The problem I'm facing is that a database update takes about 6 ms, and the device needs to service Comms requests at a much smaller latency (about 1ms).

What does using the table solve, as a problem? Understanding this could enable us to solve this using a smaller footprint, e.g. a nibble table, or scrap the table for something like an insertion sort?

Regards,
Alf Lacis

http://au.geocities.com/lacis_alfredo

(1) current MCU, ATmega168 has 16 k Flash & 1/2 k RAM

Guest

commented on Feb 25, 2008 11:57:43 AM

Obviously, in the AVR you want to put the big lookup table in ROM.

Some compilers allow you to achieve it quite easily. For example, the IAR compiler for AVR provides the extended keyword `__flash`. You need to place this keyword in front of the declaration and definition of the `log2lkup[]` variable. The compiler does all the rest.

The WinAVR compiler (GNU) contains quite an ugly hack (described online). You need to add an extended keyword `PROGMEM` in front of the `log2lkup[]` variable. But this is not the end of story. You also need to give the compiler a "helping hand" in actually reading the bytes from the program memory. The compiler comes with the macro `pgm_read_byte_near()`, which you need to use for every `PROGMEM` access.

If you want to see how this works, you might want to download the open source QP-nano event-driven framework and a development kit for AVR:

www.quantum-leaps.com/downloads
www.quantum-leaps.com/avr/

The last website contains ports for IAR and for WinAVR toolsets.

Miro

RomHuang

Teacher

commented on Feb 26, 2008 2:02:43 AM

I am enjoyed reading this article. The process before ending IRET is great.

1.I thought one event can only post to one particular task, can not to multitask,if tasks can register the events what the task want, It's will more great.

2.The queue for each task will take memory space just like the private stack of traditional kernel(linux,uc/os II). how can we sure not overflowing of the queue?

P.S. sorry for my poor english.

Best Regards,
Rom Huang

Guest

commented on Feb 26, 2008 11:26:22 AM

Good points.

I believe that your point (1) refers to the publish-subscribe event delivery model. In this model multiple tasks can subscribe to certain events. When the event occurs, the kernel is then supposed to multicast (some people say incorrectly broadcast) the event to all subscribers. This model is also known as the "Observer" pattern (see the seminal Gang-of-Four book) and is very popular in bigger event-driven systems/frameworks, such as the Java event model and many others.

As we note in the article, SST is intentionally simple (maybe oversimplified) to clearly demonstrate the core ideas. However, the state-machine-based QP framework supports publish-subscribe event delivery. QP also contains a small preemptive kernel, called QK, which works exactly like the SST described in the article. QP is open source and is available online at www.quantum-leaps.com/downloads and also from SourceForge at <http://sourceforge.net/projects/qpc/>.

As to your point (2), you are absolutely right. Event queues are an additional burden that offsets somewhat the memory gains from less stack usage and simplified TCB structure. Event queues are necessary in all event-driven systems (see any Windows GUI application) and a precise explanation why is a bit out of the scope of this post. In real-time applications, event queues should be viewed exactly like the execution stacks in that overflowing a queue is as big a crime as overflowing a stack. In other words, you must size your event queues adequately, just like you must adequately size the stacks in a traditional RTOS.

Miro

Guest

commented on Feb 26, 2008 11:42:40 AM

I'd like to add a general comment to the article. If you've never done event-driven programming before, the internal structure of SST must certainly represent a big paradigm shift for you. In fact, I hope that it actually blows your mind, because otherwise I'm not sure that you really appreciate the complete reversal of control of the event-driven SST paradigm compared to the traditional sequential code, such as tasks structured as endless loops in a conventional RTOS. This reversal of control, known as the "Hollywood Principle" (Don't call us, we'll call you!), baffles many newcomers, who often find it "mind-boggling", "backwards", or "weird".

Michael Barr and I will be speaking exactly about this subject matter at the upcoming Embedded Systems Conference in San Jose, CA. Here is the presentation abstract for our 3-hour class "Event Driven Programming for Embedded Systems" (classes ESC-301 and ESC-321):

Presentation Abstract:

Most embedded systems are event-driven by nature. Yet, the conventional embedded software architectures, ranging from "superloops" to real-time operating systems (RTOS), are typically not truly event driven. This class helps embedded developers to make the transition from sequential to the modern event-driven programming, which can be one of the trickiest paradigm shifts. The class covers such concepts as the inversion of control ("Hollywood principle"), blocking versus non-blocking code, run-to-completion (RTC) execution semantics, events as objects, the role of event queues, dealing with time, use of state machines to maintain the context, and the concept of a generic, reusable, event-driven framework. The class uses diagrams, code, and examples running on an actual target CPU.

Here is the link to the ESC information about this class:
<https://www.cmpevents.com/ESCw08/a.asp?option=C&V=11&SessID=6697>

Miro Samek

RomHuang

Teacher

commented on Feb 28, 2008 1:12:24 AM

1. How about to design only ONE Queue for all Tasks, just like only one Stack for all Tasks in SST? The scheduler may be more difficult?
2. One Task (or Active Object) can handle many events, some events are important, and some are not. I think the priority should be binding to the Events but not Task (or AO) as OS now in embedded system. Is it right?

Best Regards,
Rom

RomHuang

Teacher

commented on Feb 28, 2008 1:56:26 AM

Yes, you recalled me the IOC. Assuredly there is IOC in SST, "if a task (an actor) Post a event(application) to kernel (Hollywood),

then the kernel will CALL(phone) the task some time,the task will not poll the kernel".

Great! I like the idea.

Best Regards,
Rom

QL

Software Engineer

commented on Feb 28, 2008 4:04:24 PM

A single queue for all tasks would not allow any prioritization of work. An event for a high-priority task would have to wait until all events preceding it in the queue are processed. The multiple queues used by SST (and this class of preemptive kernels, in general) work as priority queues, whereas I mean here the standard priority queue algorithm.

Assigning priorities to events rather than tasks is problematic because of the run-to-completion (RTC) event processing semantics. Consider a task working on a low-priority event, which takes a long time to complete. If during this time the task receives a high-priority event, the high-priority event must still wait until the low-priority event is completely processed. During that time, the task could be preempted by tasks that receive medium-priority events. In other words, you have the classic priority inversion problem. Of course, you can try to do "priority inheritance", that is, boost the priority of a task whenever it receives a high-priority event, but this becomes complicated. (On the side note, the ROOM method assigns priorities to events rather than tasks, but then ROOM maintains multiple priority-queues per task, which is really complex and can change the sequence of events)

Assigning priorities to tasks (each task owning a priority queue) evades all these problems in a simple and elegant way.

Again, the aforementioned ESC classes (ESC-301 and ESC-321) will discuss all these issues and more.

Miro Samek

QL

Software Engineer

commented on Aug 13, 2009 11:50:07 AM

The article mentions that run-to-completion kernels like the SST are widely used in the industry, but does not give any examples. Here I'd like to provide such an example, which is the OSEK/VDX ISO17356 Standard (<http://www.osek-vdx.org/>).

Specifically, SST falls somewhere between the OSEK/VDX Basic Conformance Class 1 and 2 (BCC1 and BCC2), as described in the OSEK/VDX Operating System Specification 2.2.3 (<http://portal.osek-vdx.org/files/pdf/specs/os223.pdf>). Using the OSEK/VDX nomenclature, SST supports Basic Tasks with one priority level per task (like BCC1) and allows multiple activations per task (like BCC2).

The OSEK/VDX concept of a Basic Task corresponds exactly to the run-to-completion task described in the SST article. Basic Tasks cannot block in the middle of execution and must always run to completion. However, this does not preclude preemptive scheduling and Basic Tasks can support full preemptive

scheduling (OSEK/VDX Specification 2.2, Section 4.6.1), which is exactly how SST tasks work.

Interestingly, the OSEK/VDX also uses the Priority Ceiling Protocol (see OSEK/VDX Specification 2.2, Section 8.5), which is exactly the priority-ceiling mutex described in the SST article.

In summary, the SST kernel and article provides a case study for a specific open-source implementation of the very lightweight OSEK/VDX system. Of course, the specific API used in SST does not currently conform to the OSEK/VDX specification, but can be easily adapted by simply renaming the functions.

Miro Samek
<http://www.state-machine.com>

Ari Okkonen

CTO

commented on Sep 14, 2009 8:41:20 AM

For previous steps on this approach you may read the paper "ReaGOS - An Application Specific R/T-OS" presented 1993 in 5th Euromicro Workshop on Real-Time Systems. See <http://www.reagenix.com/eurom93k.pdf>

Please [login or register here](#) to post a comment or to get an email when other comments are made on this article

[Track This Thread](#)

Sends you email alerts when someone comments on this article

[Discuss This Article](#)

11 message(s). Last at: Sep 14, 2009 8:41:20 AM

Guest

commented on Nov 18, 2007 8:34:34 PM

Hi,
Perhaps I haven't understood the scheduler correctly, but the 256 byte table, `log2Lkup[]`, seems like a waste of space. Yes, I know that it's const, but const doesn't work in AVR ATmegs, which have a Harvard architecture (separately addressed buses for code and data), where even consts are in data space. Yes, we can put it into program space: special AVR functions can fetch bytes from code space: but it's not clear how efficient they are.

At the moment in my current application(1), which is essentially a KSAM database using external serial Flash & FRAM, I have about 500 bytes of program Flash left, and 200 bytes of RAM, and if we need to put SST in...

The problem I'm facing is that a database update takes about 6 ms, and the device needs to service Comms requests at a much smaller latency (about 1ms).

What does using the table solve, as a problem? Understanding this could enable us to solve this using a smaller footprint, e.g. a nibble table, or scrap the table for something like an insertion sort?

Regards,

Alf Lacis

http://au.geocities.com/lacis_alfredo

(1) current MCU, ATmega168 has 16 k Flash & 1/2 k RAM

oh, really?

fw engineer

commented on Nov 18, 2007 8:36:39 PM

(sorry, wasn't logged in)

Hi,

Perhaps I haven't understood the scheduler correctly, but the 256 byte table, `log2lkup[]`, seems like a waste of space. Yes, I know that it's const, but const doesn't work in AVR ATmegs, which have a Harvard architecture (separately addressed buses for code and data), where even consts are in data space. Yes, we can put it into program space: special AVR functions can fetch bytes from code space: but it's not clear how efficient they are.

At the moment in my current application(1), which is essentially a KSAM database using external serial Flash & FRAM, I have about 500 bytes of program Flash left, and 200 bytes of RAM, and if we need to put SST in...

The problem I'm facing is that a database update takes about 6 ms, and the device needs to service Comms requests at a much smaller latency (about 1ms).

What does using the table solve, as a problem? Understanding this could enable us to solve this using a smaller footprint, e.g. a nibble table, or scrap the table for something like an insertion sort?

Regards,

Alf Lacis

http://au.geocities.com/lacis_alfredo

(1) current MCU, ATmega168 has 16 k Flash & 1/2 k RAM

Guest

commented on Feb 25, 2008 11:57:43 AM

Obviously, in the AVR you want to put the big lookup table in ROM.

Some compilers allow you to achieve it quite easily. For example, the IAR compiler for AVR provides the extended keyword `__flash`. You need to place this keyword in front of the declaration and definition of the `log2lkup[]` variable. The compiler does all the rest.

The WinAVR compiler (GNU) contains quite an ugly hack (described online). You need to add an extended keyword `PROGMEM` in front of the `log2lkup[]` variable. But this is not the end of story. You also need to give the compiler a "helping hand" in actually reading the bytes from the program memory. The compiler comes with the macro `pgm_read_byte_near()`, which you need to use for every `PROGMEM` access.

If you want to see how this works, you might want to download the open source QP-nano event-driven framework and a development kit for AVR:

www.quantum-leaps.com/downloads

www.quantum-leaps.com/avr/

The last website contains ports for IAR and for WinAVR toolsets.

Miro

RomHuang

Teacher

commented on Feb 26, 2008 2:02:43 AM

I am enjoyed reading this article. The process before ending IRET is great.

1.I thought one event can only post to one particular task, can not to multitask,if tasks can register the events what the task want, It's will more great.

2.The queque for each task will take memory space just like the private stack of traditional kernel(linux,uc/os II). how can we sure not overflowing of the queque?

P.S. sorry for my poor english.

Best Regards,
Rom Huang

Guest

commented on Feb 26, 2008 11:26:22 AM

Good points.

I believe that your point (1) refers to the publish-subscribe event delivery model. In this model multiple tasks can subscribe to certain events. When the event occurs, the kernel is then supposed to multicast (some people say incorrectly broadcast) the event to all subscribers. This model is also known as the "Observer" pattern (see the seminal Gang-of-Four book) and is very popular in bigger event-driven systems/frameworks, such as the Java event model and many others.

As we note in the article, SST is intentionally simple (maybe oversimplified) to clearly demonstrate the core ideas. However, the state-machine-based QP framework supports publish-subscribe event delivery. QP also contains a small preemptive kernel, called QK, which works exactly like the SST described in the article. QP is open source and is available online at www.quantum-leaps.com/downloads and also from SourceForge at <http://sourceforge.net/projects/qpc/>.

As to your point (2), you are absolutely right. Event queues are an additional burden that offsets somewhat the memory gains from less stack usage and simplified TCB structure. Event queues are necessary in all event-driven systems (see any Windows GUI application) and a precise explanation why is a bit out of the scope of this post. In real-time applications, event queues should be viewed exactly like the execution stacks in that overflowing a queue is as big a crime as overflowing a stack. In other words, you must size your event queues adequately, just like you must adequately size the stacks in a traditional RTOS.

Miro

Guest

commented on Feb 26, 2008 11:42:40 AM

I'd like to add a general comment to the article. If you've never done event-driven programming before, the internal structure of SST must certainly represent a big paradigm shift for you. In fact,

I hope that it actually blows your mind, because otherwise I'm not sure that you really appreciate the complete reversal of control of the event-driven SST paradigm compared to the traditional sequential code, such as tasks structured as endless loops in a conventional RTOS. This reversal of control, known as the "Hollywood Principle" (Don't call us, we'll call you!), baffles many newcomers, who often find it "mind-boggling", "backwards", or "weird".

Michael Barr and I will be speaking exactly about this subject matter at the upcoming Embedded Systems Conference in San Jose, CA. Here is the presentation abstract for our 3-hour class "Event Driven Programming for Embedded Systems" (classes ESC-301 and ESC-321):

Presentation Abstract:

Most embedded systems are event-driven by nature. Yet, the conventional embedded software architectures, ranging from "superloops" to real-time operating systems (RTOS), are typically not truly event driven. This class helps embedded developers to make the transition from sequential to the modern event-driven programming, which can be one of the trickiest paradigm shifts. The class covers such concepts as the inversion of control ("Hollywood principle"), blocking versus non-blocking code, run-to-completion (RTC) execution semantics, events as objects, the role of event queues, dealing with time, use of state machines to maintain the context, and the concept of a generic, reusable, event-driven framework. The class uses diagrams, code, and examples running on an actual target CPU.

Here is the link to the ESC information about this class:
<https://www.cmpevents.com/ESCw08/a.asp?option=C&V=11&SessID=6697>

Miro Samek

RomHuang

Teacher

commented on Feb 28, 2008 1:12:24 AM

1. How about to design only ONE Queue for all Tasks, just like only one Stack for all Tasks in SST? The scheduler may be more difficult?

2. One Task (or Active Object) can handle many events, some events are important, and some are not. I think the priority should be binding to the Events but not Task (or AO) as OS now in embedded system. Is it right?

Best Regards,
Rom

RomHuang

Teacher

commented on Feb 28, 2008 1:56:26 AM

Yes, you recalled me the IOC. Assuredly there is IOC in SST, "if a task (an actor) posts an event to the kernel (Hollywood), then the kernel will CALL (phone) the task some time, the task will not poll the kernel".

Great! I like the idea.

Best Regards,
Rom

QL

Software Engineer

commented on Feb 28, 2008 4:04:24 PM

A single queue for all tasks would not allow any prioritization of work. An event for a high-priority task would have to wait until all events preceding it in the queue are processed. The multiple queues used by SST (and this class of preemptive kernels, in general) work as priority queues, whereas I mean here the standard priority queue algorithm.

Assigning priorities to events rather than tasks is problematic because of the run-to-completion (RTC) event processing semantics. Consider a task working on a low-priority event, which takes a long time to complete. If during this time the task receives a high-priority event, the high-priority event must still wait until the low-priority event is completely processed. During that time, the task could be preempted by tasks that receive medium-priority events. In other words, you have the classic priority inversion problem. Of course, you can try to do "priority inheritance", that is, boost the priority of a task whenever it receives a high-priority event, but this becomes complicated. (On the side note, the ROOM method assigns priorities to events rather than tasks, but then ROOM maintains multiple priority-queues per task, which is really complex and can change the sequence of events)

Assigning priorities to tasks (each task owning a priority queue) evades all these problems in a simple and elegant way.

Again, the aforementioned ESC classes (ESC-301 and ESC-321) will discuss all these issues and more.

Miro Samek

QL

Software Engineer

commented on Aug 13, 2009 11:50:07 AM

The article mentions that run-to-completion kernels like the SST are widely used in the industry, but does not give any examples. Here I'd like to provide such an example, which is the OSEK/VDX ISO17356 Standard (<http://www.osek-vdx.org/>).

Specifically, SST falls somewhere between the OSEK/VDX Basic Conformance Class 1 and 2 (BCC1 and BCC2), as described in the OSEK/VDX Operating System Specification 2.2.3 (<http://portal.osek-vdx.org/files/pdf/specs/os223.pdf>). Using the OSEK/VDX nomenclature, SST supports Basic Tasks with one priority level per task (like BCC1) and allows multiple activations per task (like BCC2).

The OSEK/VDX concept of a Basic Task corresponds exactly to the run-to-completion task described in the SST article. Basic Tasks cannot block in the middle of execution and must always run to completion. However, this does not preclude preemptive scheduling and Basic Tasks can support full preemptive scheduling (OSEK/VDX Specification 2.2, Section 4.6.1), which is exactly how SST tasks work.

Interestingly, the OSEK/VDX also uses the Priority Ceiling Protocol (see OSEK/VDX Specification 2.2, Section 8.5), which is exactly the priority-ceiling mutex described in the SST article.

In summary, the SST kernel and article provides a case study for a specific open-source implementation of the very lightweight OSEK/VDX system. Of course, the specific API used in SST does not currently conform to the OSEK/VDX specification, but can be easily adapted by simply renaming the functions.

Miro Samek
<http://www.state-machine.com>

Ari Okkonen

CTO

commented on Sep 14, 2009 8:41:20 AM

For previous steps on this approach you may read the paper "ReaGOS - An Application Specific R/T-OS" presented 1993 in 5th Euromicro Workshop on Real-Time Systems. See <http://www.reagenix.com/eurom93k.pdf>

Please [login or register here](#) to post a comment or to get an email when other comments are made on this article

[Track This Thread](#)

Sends you email alerts when someone comments on this article