

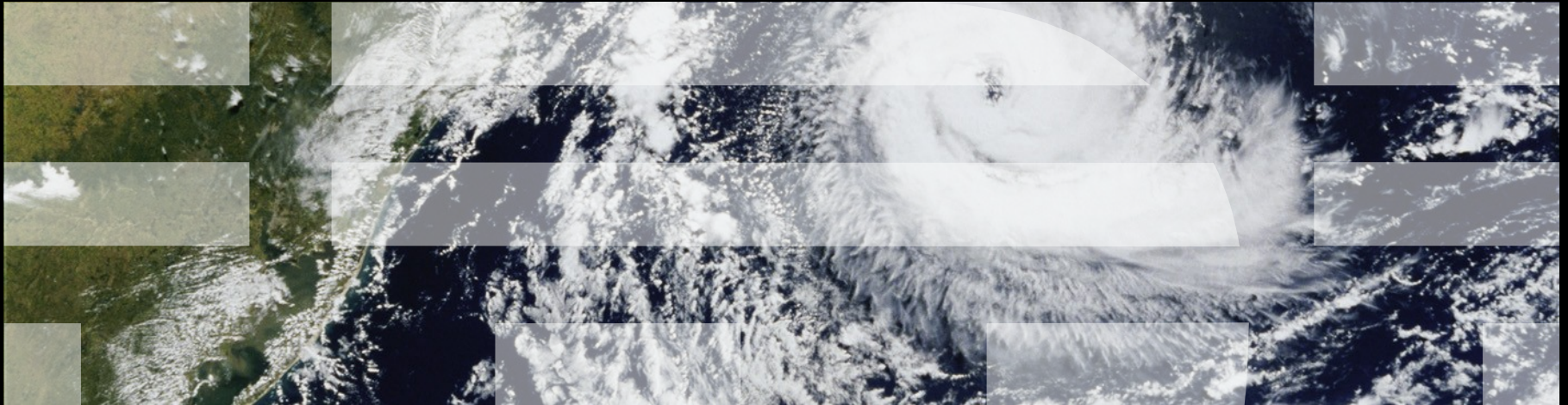
Paul E. McKenney, IBM Distinguished Engineer, Linux Technology Center

4 April 2012



# Validating Core Parallel Software

*Linux Collaboration Summit*



## Overview

- Avoiding Debugging The Open-Source Way
- Avoiding Debugging By Design
- Avoiding Debugging By Process
- Avoiding Debugging By Mechanical Proofs
- Avoiding Debugging By Statistical Analysis
- Coping With Schedule Pressure
- But I Did All This And There Are Still Bugs!!!
- Summary and Conclusions

## But First, Some Limitations...

- The only bug-free program is a trivial program
- A reliable programs has no known bugs

## But First, Some Limitations...

- The only bug-free program is a trivial program
- A reliable programs has no known bugs
- Therefore, any non-trivial reliable program has at least one bug that you do not know about

## But First, Some Limitations...

- The only bug-free program is a trivial program
- A reliable programs has no known bugs
- Therefore, any non-trivial reliable program has at least one bug that you do not know about
- Fortunately, the size and complexity of trivial programs has been steadily increasing over the decades, but unfortunately almost all useful programs are decidedly non-trivial

## But First, Some Limitations...

- The only bug-free program is a trivial program
- A reliable programs has no known bugs
- Therefore, any non-trivial reliable program has at least one bug that you do not know about
- Fortunately, the size and complexity of trivial programs has been steadily increasing over the decades, but unfortunately almost all useful programs are decidedly non-trivial
- The Linux kernel is a non-trivial program

## But First, Some Limitations...

- The only bug-free program is a trivial program
- A reliable programs has no known bugs
- Therefore, any non-trivial reliable program has at least one bug that you do not know about
- Fortunately, the size and complexity of trivial programs has been steadily increasing over the decades, but unfortunately almost all useful programs are decidedly non-trivial
- The Linux kernel is a non-trivial program
- In this imperfect real world, validation is more about reliability than about complete freedom from bugs

# Avoiding Debugging The Open-Source Way

Strengths and Weaknesses?



## Avoiding Debugging The Open-Source Way

- To 10,000 eyes, all bugs are shallow
  - But how many of those eyes are going to actually look at your patch?
  - When will they look at it?
  - How many will be experienced/clever enough to find your bugs?
- Many people test the Linux kernel
  - Some test random patches (perhaps even yours)
  - Some test -next
  - Some test maintainer trees
  - Some test mainline
  - Some test distro kernels
  - Will they run the HW/SW config and workload that will find your bug?
- Open source review/testing is wonderful, but for core parallel software you might want to be more proactive...

# Avoiding Debugging By Design

## Avoiding Debugging By Design

- Understand the Hardware
- Understand the Software Environment

# Performance is a Key Correctness Criterion for Parallel Software

If you don't care about performance,  
why on earth are you bothering with parallelism???

# Performance of Synchronization Mechanisms

4-CPU 1.8GHz AMD Opteron 844 system

Need to be here!  
(Partitioning/RCU)

Operation	Cost (ns)	Ratio
Clock period	0.6	1
Best-case CAS	37.9	63.2
Best-case lock	65.6	109.3
Single cache miss	139.5	232.5
CAS cache miss	306.0	510.0

Heavily optimized reader-writer lock might get here for readers (but too bad about those poor writers...)

Typical synchronization mechanisms do this a lot

# Performance of Synchronization Mechanisms

4-CPU 1.8GHz AMD Opteron 844 system

Need to be here!  
(Partitioning/RCU)

Operation	Cost (ns)	Ratio
Clock period	0.6	1
Best-case CAS	37.9	63.2
Best-case lock	65.6	109.3
Single cache miss	139.5	232.5
CAS cache miss	306.0	510.0

Heavily optimized reader-writer lock might get here for readers (but too bad about those poor writers...)

Typical synchronization mechanisms do this a lot

But this is an old system...

# Performance of Synchronization Mechanisms

4-CPU 1.8GHz AMD Opteron 844 system

Need to be here!  
(Partitioning/RCU)

Operation	Cost (ns)	Ratio
Clock period	0.6	1
Best-case CAS	37.9	63.2
Best-case lock	65.6	109.3
Single cache miss	139.5	232.5
CAS cache miss	306.0	510.0

Heavily optimized reader-writer lock might get here for readers (but too bad about those poor writers...)

Typical synchronization mechanisms do this a lot

But this is an old system...

And why low-level details???

## Why All These Low-Level Details???

- Would you trust a bridge designed by someone who did not understand strengths of materials?
  - Or a bridge designed by someone who did not understand that concrete, though quite strong under compression, is extremely weak under tension?
  - Or a car designed by someone who did not understand the corrosion properties of the metals used in the exhaust system?
  - Or a space shuttle designed by someone who did not understand the temperature limitations of O-rings?
  
- So why trust algorithms from someone ignorant of the properties of the underlying hardware???



# But Isn't Hardware Just Getting Faster?

## Performance of Synchronization Mechanisms

16-CPU 2.8GHz Intel X5550 (Nehalem) System

Operation	Cost (ns)	Ratio
Clock period	0.4	1
“Best-case” CAS	12.2	33.8
Best-case lock	25.6	71.2
Single cache miss	12.9	35.8
CAS cache miss	7.0	19.4

**What a difference a few years can make!!!**

## Performance of Synchronization Mechanisms

16-CPU 2.8GHz Intel X5550 (Nehalem) System

Operation	Cost (ns)	Ratio
Clock period	0.4	1
"Best-case" CAS	12.2	33.8
Best-case lock	25.6	71.2
Single cache "miss"	12.9	35.8
CAS cache "miss"	7.0	19.4
Single cache miss (off-core)	31.2	86.6
CAS cache miss (off-core)	31.2	86.5

**Not *quite* so good... But still a 6x improvement!!!**

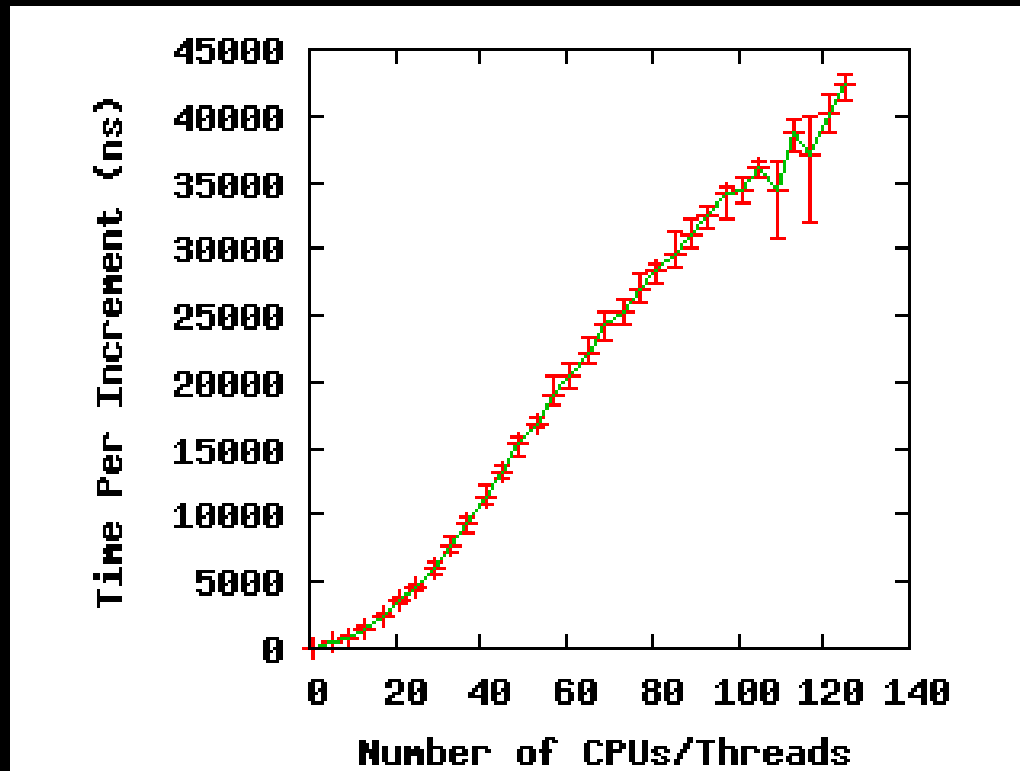
# Performance of Synchronization Mechanisms

## 16-CPU 2.8GHz Intel X5550 (Nehalem) System

Operation	Cost (ns)	Ratio
Clock period	0.4	1
"Best-case" CAS	12.2	33.8
Best-case lock	25.6	71.2
Single cache miss	12.9	35.8
CAS cache miss	7.0	19.4
Single cache miss (off-core)	31.2	86.6
CAS cache miss (off-core)	31.2	86.5
Single cache miss (off-socket)	92.4	256.7
CAS cache miss (off-socket)	95.9	266.4

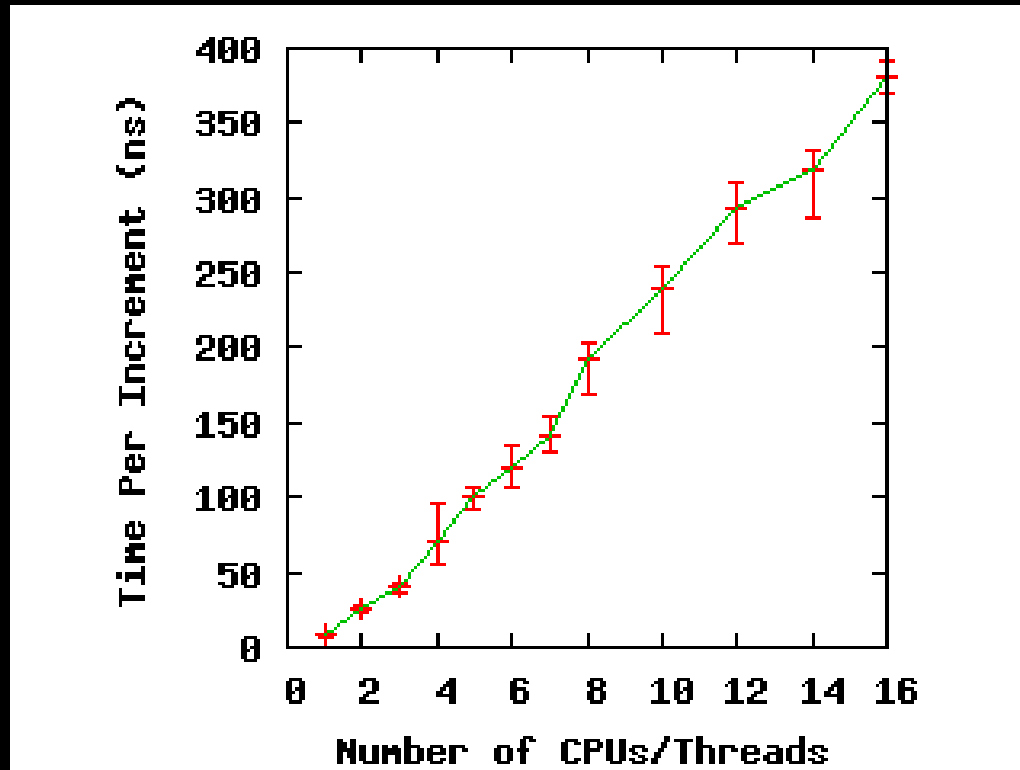
**Maybe not such a big difference after all...  
And these are best-case values!!! (Why?)**

# Performance of Synchronization Mechanisms



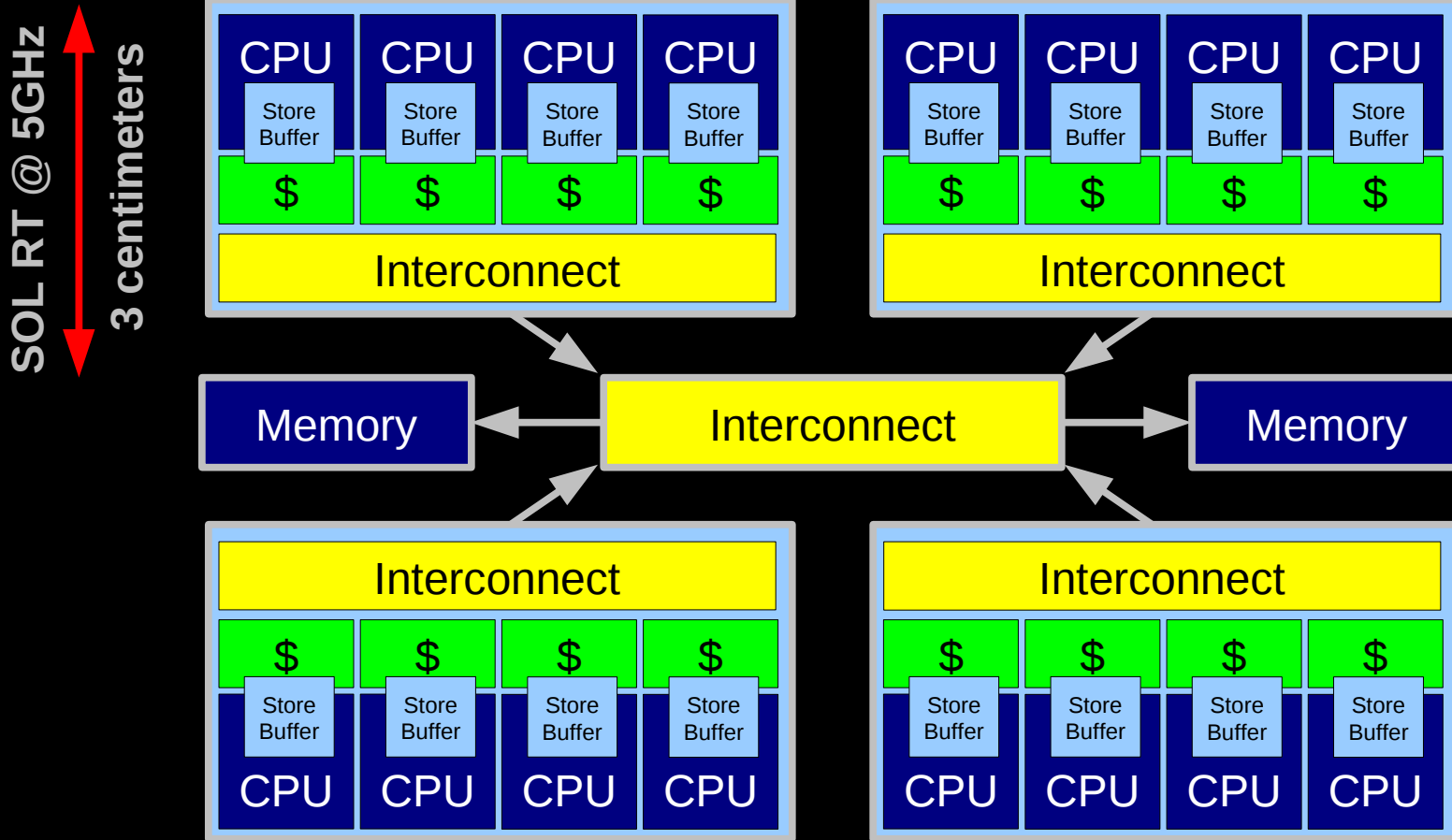
If you thought a *single* atomic operation was slow, try lots of them!!!  
(Parallel atomic increment of single variable on 1.9GHz Power 5 system)

# Performance of Synchronization Mechanisms



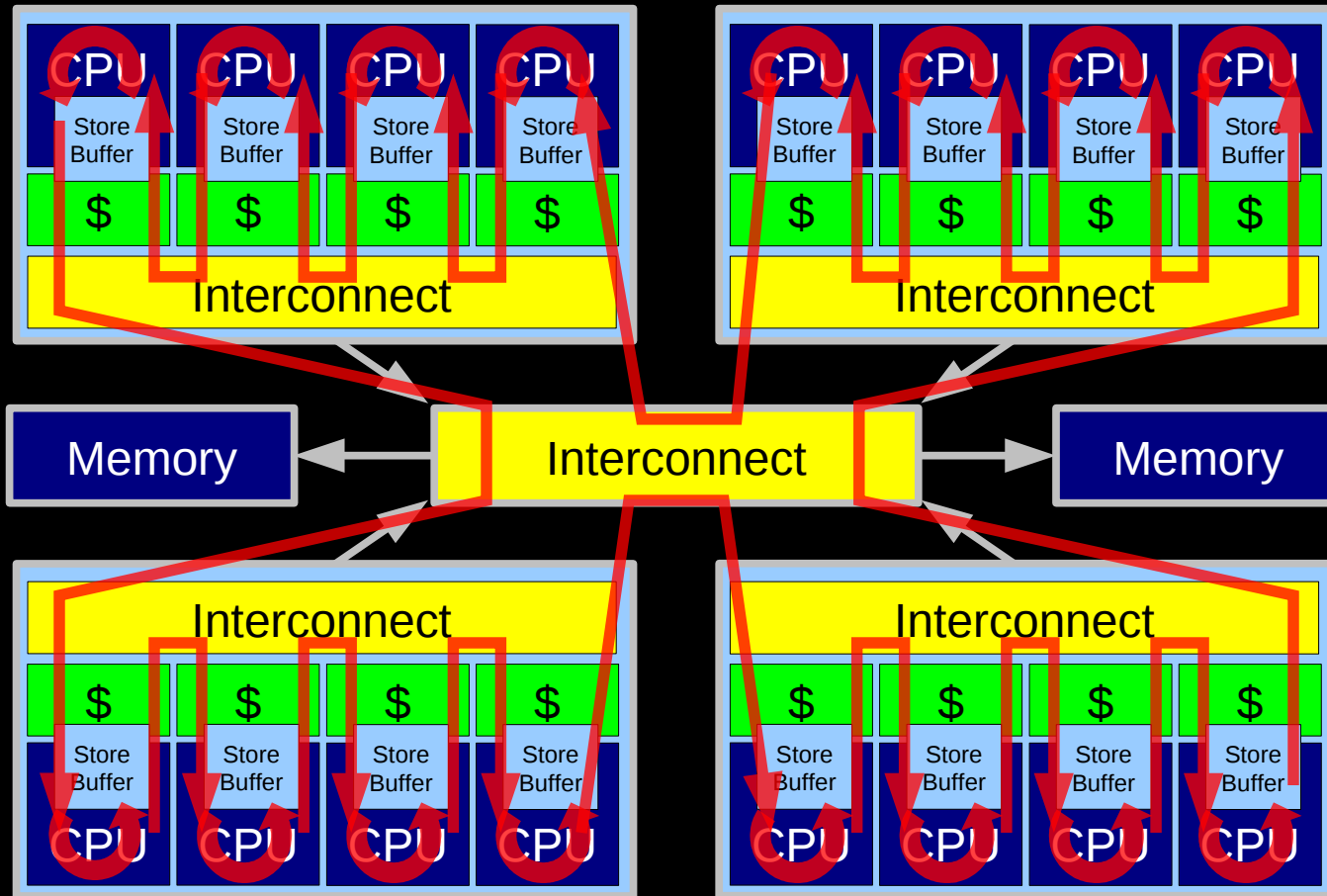
Same effect on a 16-CPU 2.8GHz Intel X5550 (Nehalem) system

# System Hardware Structure



Electrons move at 0.03C to 0.3C in transistors and, so lots of waiting. 3D???

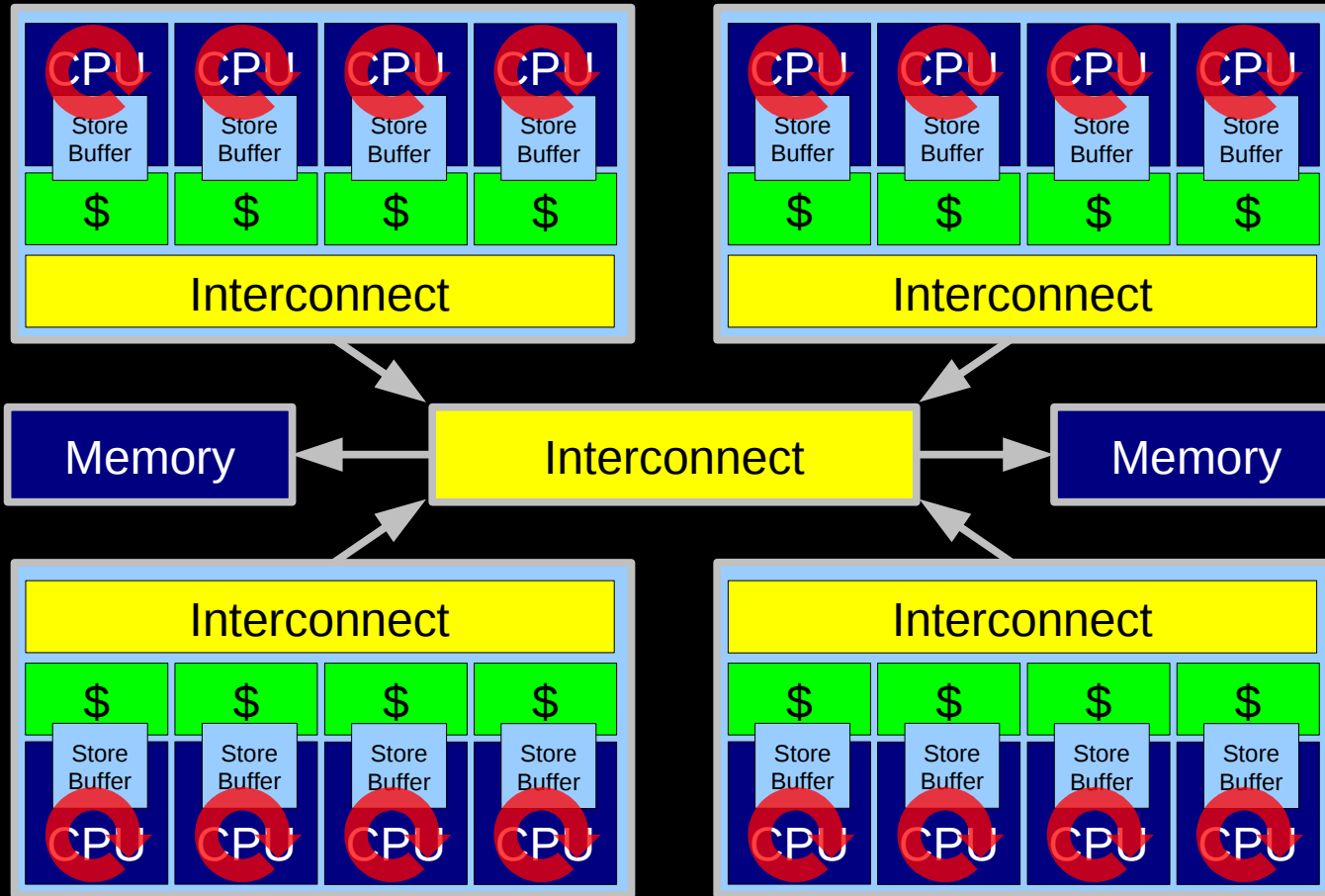
# Atomic Increment of Global Variable



Lots and Lots of Latency!!!

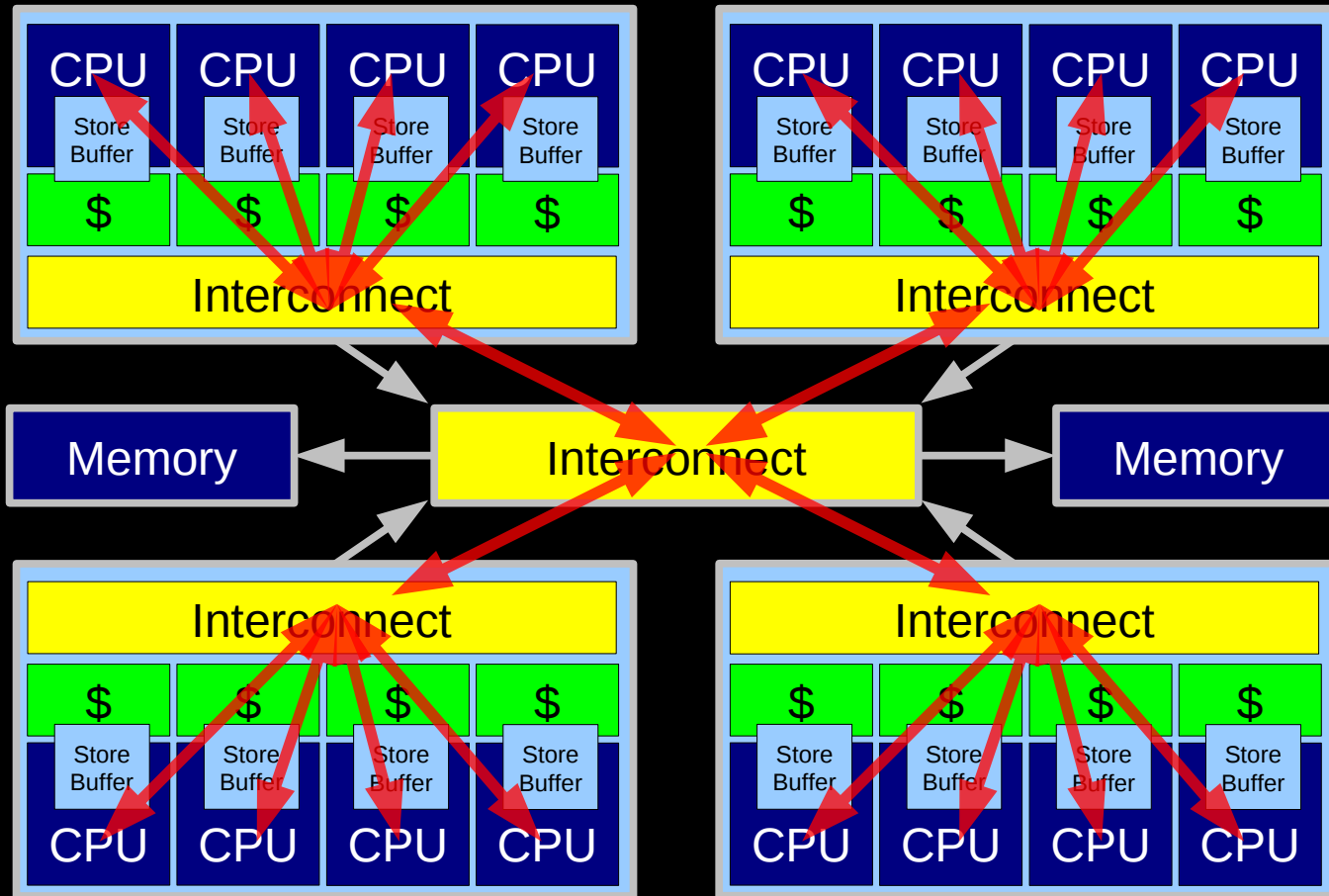


# Atomic Increment of Per-CPU Variable



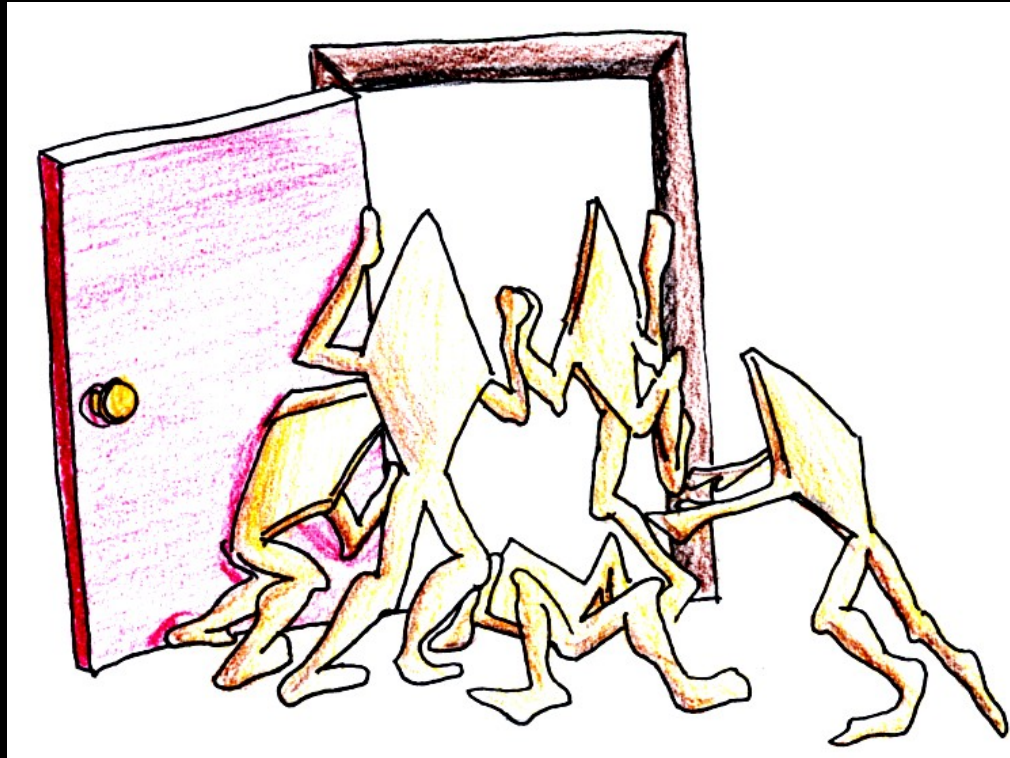
Little Latency, Lots of Increments at Core Clock Rate

# HW-Assist Atomic Increment of Global Variable



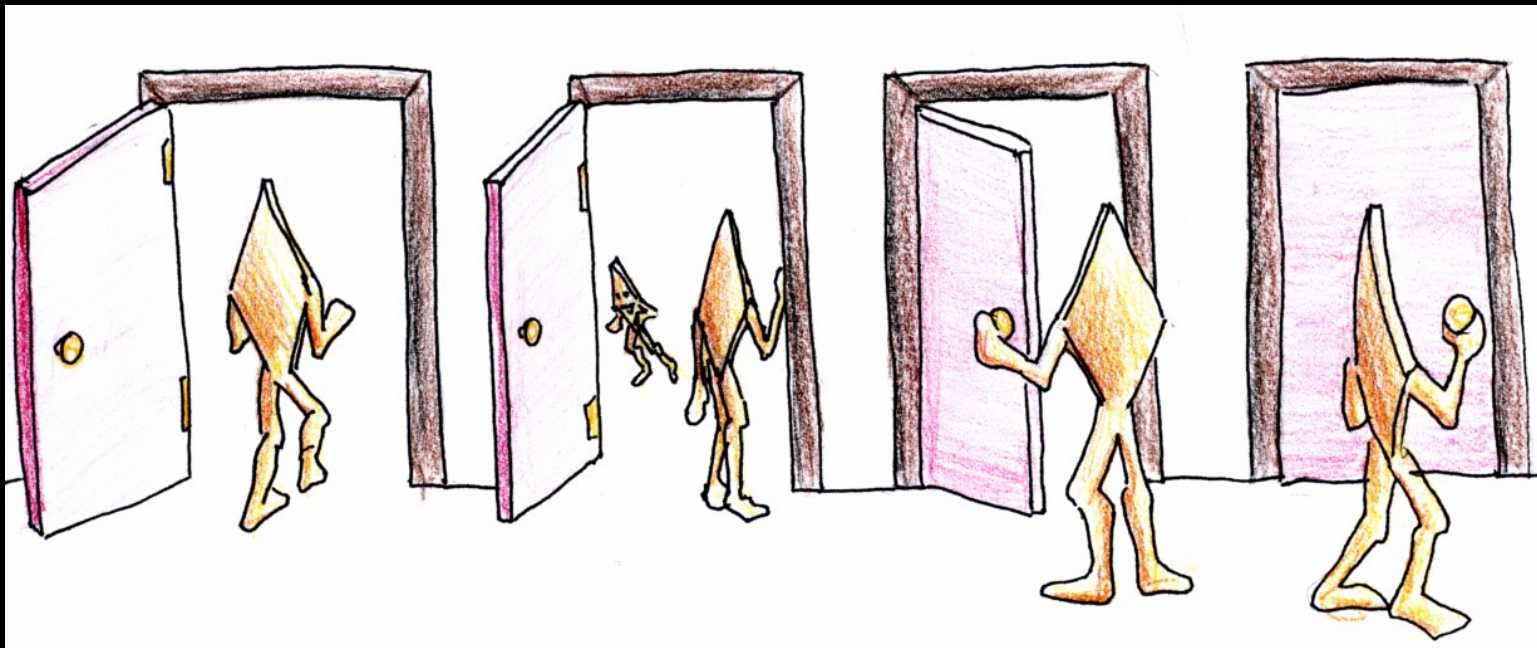
SGI systems used this approach in the 1990s, expect modern micros to pick it up.  
Yes, HW changes. SW must change with it.

## Design Principle: Avoid Bottlenecks



**Only one of something: bad for performance and scalability.  
Also typically results in high complexity.**

## Design Principle: Avoid Bottlenecks



**Many instances of something good!  
Avoiding tightly coupled interactions is an excellent way to avoid bugs.**

## Understand the Hardware: Summary

- A strong understanding of the hardware helps rule out infeasible designs early in process
- Understanding hardware trends helps reduce the amount of future rework required
- Ditto for low-level software that your code depends on
  - In my case, things like `in_irq()` and user-mode helpers...
- Of course both the hardware and low-level software will change with time...
  - Even if it is now correct, your code will eventually be wrong...

## Understand the Hardware: Summary

- A strong understanding of the hardware helps rule out infeasible designs early in process
- Understanding hardware trends helps reduce the amount of future rework required
- Ditto for low-level software that your code depends on
  - In my case, things like `in_irq()` and user-mode helpers...
- Of course both the hardware and low-level software will change with time...
  - Even if it is now correct, your code will eventually be wrong...
  - This is one reason why we do regression testing

## Understand the Software Environment

- Understand the Workloads
  - Which for Linux means a great many of them
  - Your code must handle whatever shows up
- Google-Search LWN
  - But you knew this already
- Test Unfamiliar Primitives
  - And complain on LKML if they break
  - Preferably accompanying the complaint with a fix
- Review Others' Code
  - See Documentation directory for how-to info
- Make a Map
  - See next slides...

# Making a Map of Software



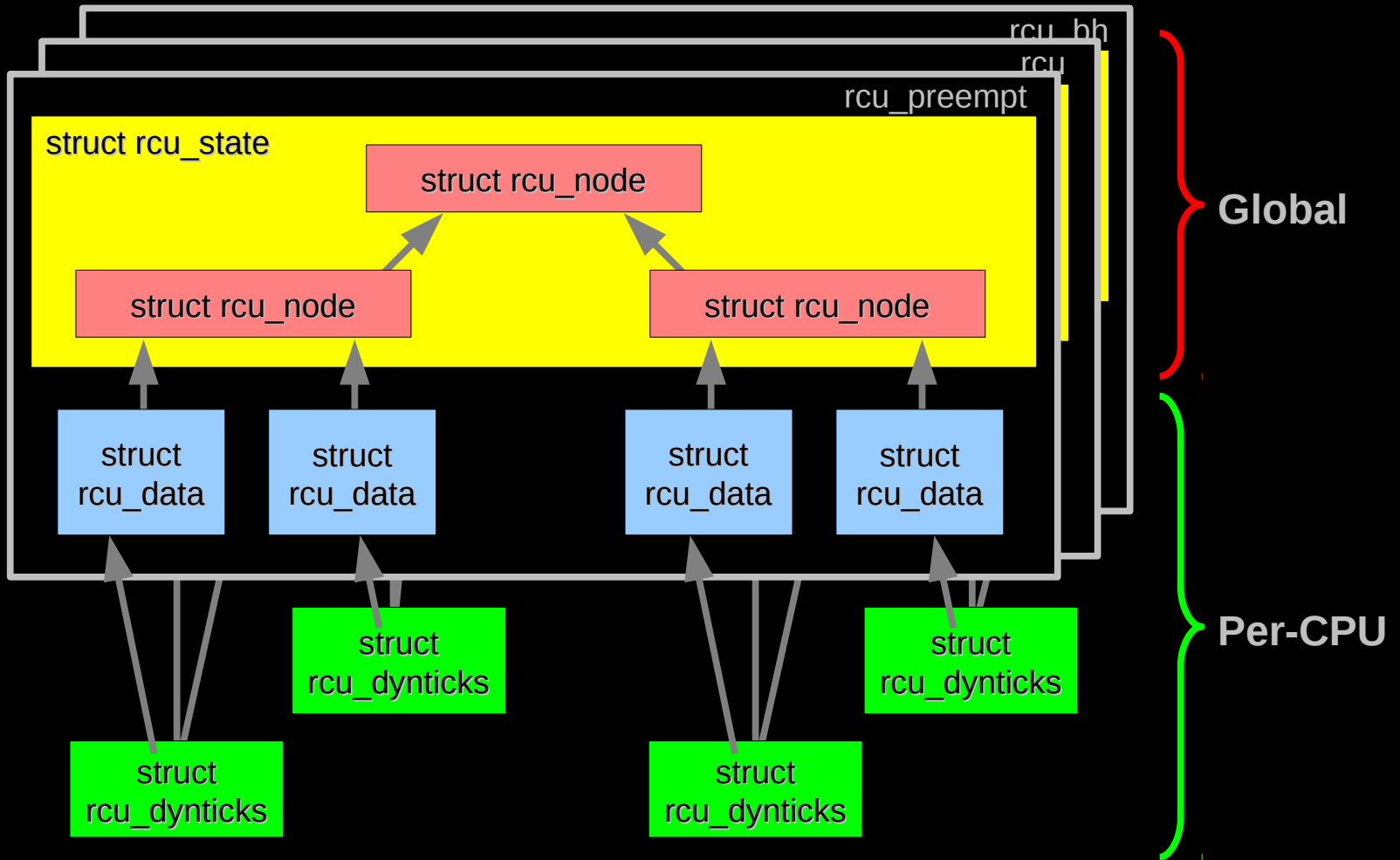
# Hierarchical RCU Data Structures

```

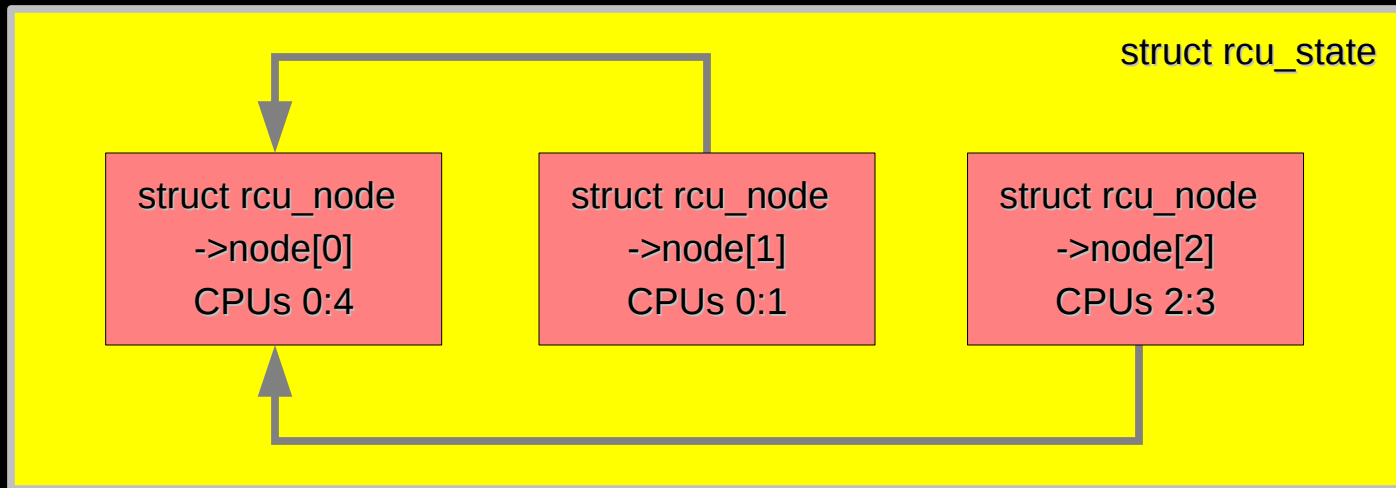
1 struct rcu_dynticks {
2     int dynticks_nesting;
3     int dynticks;
4     int dynticks_nmi;
5 };
6
7 struct rcu_node {
8     spinlock_t lock;
9     long gpnum;
10    long completed;
11    unsigned long qsmask;
12    unsigned long qsmaskinit;
13    unsigned long grpmask;
14    int grplo;
15    int grphi;
16    u8 grpnum;
17    u8 level;
18    struct rcu_node *parent;
19    struct list_head blocked_tasks[2];
20 }
21
22 struct rcu_data {
23     long    completed;
24     long    gpnum;
25     long    passed_quiesc_completed;
26     bool    passed_quiesc;
27     bool    qs_pending;
28     bool    beenonline;
29     bool    preemptable;
30     struct rcu_node *mynode;
31     unsigned long grpmask;
32     struct rcu_head *nxtlist;
33     struct rcu_head **nxttail[RCU_NEXT_SIZE];
34     long    qlen;
35     long    qlen_last_fqs_check;
36     unsigned long n_force_qs_snap;
37     long    blimit;
38 #ifdef CONFIG_NO_HZ
39     struct rcu_dynticks *dynticks;
40     int dynticks_snap;
41     int dynticks_nmi_snap;
42 #endif CONFIG_NO_HZ
43     unsigned long dynticks_fqs;
44 #endif /* #ifdef CONFIG_NO_HZ */
45     unsigned long offline_fqs;
46     unsigned long resched_ipi;
47     long n_rcu_pending;
48     long n_rp_qs_pending;
49     long n_rp_cb_ready;
50     long n_rp_cpu_needs_gp;
51     long n_rp_gp_completed;
52     long n_rp_gp_started;
53     long n_rp_need_fqs;
54     long n_rp_need_nothing;
55     int cpu;
56 };
57
58 struct rcu_state {
59     struct rcu_node node[NUM_RCU_NODES];
60     struct rcu_node *level[NUM_RCU_LVLLS];
61     u32 levelcnt[MAX_RCU_LVLLS + 1];
62     u8 levelspread[NUM_RCU_LVLLS];
63     struct rcu_data *rda[NR_CPUS];
64     u8 signaled;
65     long gpnum;
66     long completed;
67     spinlock_t onofflock;
68     struct rcu_head *orphan_cbs_list;
69     struct rcu_head **orphan_cbs_tail;
70     long orphan_qlen;
71     spinlock_t fqslock;
72     unsigned long jiffies_force_qs;
73     unsigned long n_force_qs;
74     unsigned long n_force_qs_lh;
75     unsigned long n_force_qs_ngp;
76 #ifdef CONFIG_RCU_CPU_STALL_DETECTOR
77     unsigned long gp_start;
78     unsigned long jiffies_stall;
79 #endif /* #ifdef CONFIG_RCU_CPU_STALL_DETECTOR */
80     long dynticks_completed;
81 };

```

# Mapping Data Structures



## Placement of rcu\_node Within rcu\_state



# Avoiding Debugging By Process

## Avoiding Debugging By Process

- Review your own work carefully
  - See following slides
- Test early, test often, test in small pieces
  - Debugging is 2-3 times harder than writing code
  - Debugging effort rises as the square of the amount of new code added to the testing effort
- Where possible, use existing well-tested code
  - Even if it is a lot more fun to re-invent the wheel
- I would have rejected this advice as late as the early 1990s, but have since learned the hard way to accept it
- But I still sometimes has difficulty following it:
  - <http://paulmck.livejournal.com/14639.html>

## Review Your Own Code Carefully

- Paul E. McKenney's self-review rules for complex code:
  - Write the code long hand in pen on paper
  - Correct bugs as you go
  - Copy onto a clean sheet of paper
  - Repeat until the last two versions are identical
  
- What constitutes “not complex”?
  - Sequential code, *and*
  - You test it incrementally
    - For example, bash script or single-threaded C-code with gdb)

↖ just after "if" in "do-while"  
 in --rcu-offline-cpu()

```

static void rcu-preempt-offline-tasks(struct rcu_state *rsp,
                                     struct rcu_node *rnp)
    
```

⋮

acquire root rcu-node lock, migrate tasks,  
 updating their pointers... (blocked-tasks)

-> also fix lock arg. in rcu-read-unlock-speed  
 put them all on current index of root node

(safe, simple) ← comment which is  
 current. - (the one  
 indexed by low  
 bit of gpnum)

→ OT copy straight  
 same index!!!

→ and fix ctx-switch stuff  
 to check lock after acq.~

```

static void rcu_preempt_offline_tasks(struct rcu_state *rsp,
                                     struct rcu_node *rnp)
{
    struct rcu_node *rnp_root = rcu_get_root();
    if (!list_empty struct task_struct *tp; stet
    int i; struct list_head *lp; stet
    if (rnp == rnp_root) return;
    for (i = 0; i < 2; i++) { lp =
        if (!list_empty(rnp->blocked_tasks[i])) {
            continue;
            list_for_each_entry(lp, &rnp->
                lp = list_entry(

```



```

static void rcu-preempt-offline-tasks(struct rcu-state *rsp,
                                     struct rcu-node *rnp)
{
    int i;
    struct list_head *lp;
    struct list_head *lp-root = rcu-get-root(rsp);
    struct rcu-node *rnp-root = rcu-get-root(rsp);
    struct task_struct *tp;

    if (rnp == rnp-root)
        return;

    for (i = 0; i < 2; i++) {
        lp = &rnp->blocked-tasks[i];
        while (!list_empty(lp)) {
            lp-root = &rnp-root->blocked-tasks[i];
            tp = list_entry(lp->next, typeof(*tp),
                           rcu-node-entry);
            spin_lock(&rnp-root->lock); /* irqs disabled */
            list_del(&tp->rcu-node-entry);
            list_add(&tp->rcu-node-entry, lp-root);
            tp->rcu-blocked-node = rnp-root;
            spin_unlock(&rnp-root->lock); /* irqs disabled */
        }
    }
}

```

```

static void rcu-preempt-offline-tasks(struct rcu-state *rsp,
                                     struct rcu-node *rnp)
{
    int i;
    struct list-head *lp;
    struct list-head *lp-root;
    struct rcu-node *rnp-root = rcu-get-root(rsp);
    struct task_struct *tp;

    if (rnp == rnp-root)
        return;
    for (i = 0; i < 2; i++) {
        lp = &rnp->blocked-tasks[i];
        lp-root = &rnp-root->blocked-tasks[i];
        while (!list-empty(lp)) {
            tp = list-entry(lp->next, typeof(*tp), rcu-node-entry);
            spin_lock(&rnp-root->lock); /* irqs disabled */
            list_del(&tp->rcu-node-entry);
            list_add(&tp->rcu-node-entry, lp-root);
            tp->rcu-blocked-node = rnp-root;
            spin_unlock(&rnp-root->lock); /* irqs disabled */
        }
    }
}

```

# So, How Well Did I Do?

```
1 static void rcu_preempt_offline_tasks(struct rcu_state *rsp,
2         struct rcu_node *rnp,
3         struct rcu_data *rdp)
4 {
5     int i;
6     struct list_head *lp;
7     struct list_head *lp_root;
8     struct rcu_node *rnp_root = rcu_get_root(rsp);
9     struct task_struct *tp;
10
11     if (rnp == rnp_root) {
12         WARN_ONCE(1, "Last CPU thought to be offlined?");
13         return;
14     }
15     WARN_ON_ONCE(rnp != rdp->mynode &&
16         (!list_empty(&rnp->blocked_tasks[0]) ||
17         !list_empty(&rnp->blocked_tasks[1])));
18     for (i = 0; i < 2; i++) {
19         lp = &rnp->blocked_tasks[i];
20         lp_root = &rnp_root->blocked_tasks[i];
21         while (!list_empty(lp)) {
22             tp = list_entry(lp->next, typeof(*tp), rcu_node_entry);
23             spin_lock(&rnp_root->lock); /* irqs already disabled */
24             list_del(&tp->rcu_node_entry);
25             tp->rcu_blocked_node = rnp_root;
26             list_add(&tp->rcu_node_entry, lp_root);
27             spin_unlock(&rnp_root->lock); /* irqs remain disabled */
28         }
29     }
30 }
```

```
1 static int rcu_preempt_offline_tasks(struct rcu_state *rsp,
2                                     struct rcu_node *rnp,
3                                     struct rcu_data *rdp)
4 {
5     int i;
6     struct list_head *lp;
7     struct list_head *lp_root;
8     int retval;
9     struct rcu_node *rnp_root = rcu_get_root(rsp);
10    struct task_struct *tp;
11
12    if (rnp == rnp_root) {
13        WARN_ONCE(1, "Last CPU thought to be offlined?");
14        return 0; /* Shouldn't happen: at least one CPU online. */
15    }
16    WARN_ON_ONCE(rnp != rdp->mynode &&
17                (!list_empty(&rnp->blocked_tasks[0]) ||
18                 !list_empty(&rnp->blocked_tasks[1])));
19    retval = rcu_preempted_readers(rnp);
20    for (i = 0; i < 2; i++) {
21        lp = &rnp->blocked_tasks[i];
22        lp_root = &rnp_root->blocked_tasks[i];
23        while (!list_empty(lp)) {
24            tp = list_entry(lp->next, typeof(*tp), rcu_node_entry);
25            spin_lock(&rnp_root->lock); /* irqs already disabled */
26            list_del(&tp->rcu_node_entry);
27            tp->rcu_blocked_node = rnp_root;
28            list_add(&tp->rcu_node_entry, lp_root);
29            spin_unlock(&rnp_root->lock); /* irqs remain disabled */
30        }
31    }
32    return retval;
33 }
```

```

1 static int rcu_preempt_offline_tasks(struct rcu_state *rsp,
2                                     struct rcu_node *rnp,
3                                     struct rcu_data *rdp)
4 {
5     int i;
6     struct list_head *lp;
7     struct list_head *lp_root;
8     int retval;
9     struct rcu_node *rnp_root = rcu_get_root(rsp);
10    struct task_struct *tp;
11
12    if (rnp == rnp_root) {
13        WARN_ONCE(1, "Last CPU thought to be offlined?");
14        return 0; /* Shouldn't happen: at least one CPU online. */
15    }
16    WARN_ON_ONCE(rnp != rdp->mynode &&
17                (!list_empty(&rnp->blocked_tasks[0]) ||
18                 !list_empty(&rnp->blocked_tasks[1])));
19    retval = rcu_preempted_readers(rnp);
20    for (i = 0; i < 2; i++) {
21        lp = &rnp->blocked_tasks[i];
22        lp_root = &rnp_root->blocked_tasks[i];
23        while (!list_empty(lp)) {
24            tp = list_entry(lp->next, typeof(*tp), rcu_node_entry);
25            spin_lock(&rnp_root->lock); /* irqs already disabled */
26            list_del(&tp->rcu_node_entry);
27            tp->rcu_blocked_node = rnp_root;
28            list_add(&tp->rcu_node_entry, lp_root);
29            spin_unlock(&rnp_root->lock); /* irqs remain disabled */
30        }
31    }
32    return retval;
33 }

```

*More changes due to RCU priority boosting*

# When Does This Approach Fail to Avoid Bugs?

## When Does This Approach Fail to Avoid Bugs?

- Excessive schedule pressure rules out this approach
- Excessive optimism about understanding of surrounding hardware and software
  - Limitations of `in_irq()`
  - “Interesting” properties of user-mode helpers: Entering exception handlers then never leaving them
- Excessive optimism about understanding of requirements
  - “But it is only a few microseconds added latency!!!”
  - “But it is only a few extra scheduler-clock interrupts!!!”
  - “But it won't degrade scalability up to at least eight CPUs!!!”
- Excessive optimism about understanding of algorithm
  - “But that cannot possibly happen!!!”
- But without excessive optimism, we would never start anything...



# Avoiding Debugging By Mechanical Proofs

## Avoiding Debugging By Mechanical Proofs

- Works well for small, self-contained algorithms
  - <http://lwn.net/Articles/243851/> (QRCU)
  - <http://lwn.net/Articles/279077/> (RCU dynticks I/F)
  - <git://ltnng.org/userspace-rcu> formal-model (URCU)
  
- However, the need for formal proof often indicates an overly complex design!!!
  - Preemptible RCU's dynticks interface being an extreme case in point (<http://lwn.net/Articles/279077/>)
  
- And you cannot prove your proof's assumptions...

# Avoiding Debugging By Statistical Analysis

## Avoiding Debugging By Statistical Analysis

- Different kernel configuration options select different code
- Suppose that more failure occur with `CONFIG_FOO=y`
  - Focus inspection on code under `#ifdef CONFIG_FOO`
- But what exactly does “more failures” mean?

## Avoiding Debugging By Statistical Analysis

- Different kernel configuration options select different code
- Suppose that more failure occur with `CONFIG_FOO=y`
  - Focus inspection on code under `#ifdef CONFIG_FOO`
- But what exactly does “more failures” mean?
  - That is where the statistical analysis comes in
  - The “more failures” must be enough more to be statistically significant
  - One of the most useful classes I took as an undergraduate was a statistics course!

# Coping With Schedule Pressure

## Coping With Schedule Pressure

- When you are fixing a critical bug, speed counts
- The difference is level of risk
  - The code is *already* broken, so there is less benefit from using extremely dainty process steps
  - But *only* if you follow up with careful process
  - Which I repeatedly learn the hard way:  
<http://paulmck.livejournal.com/14639.html>
  - Failure to invest a few days in early 2009 cost me more than a month in late 2009!!!
- Long-term perspective required
  - And that means *you* – leave the “blame it on management” game to Dilbert cartoons
  - Align with management initiatives, for example, “agile development”

---

**But I Did All This And There Are Still Bugs!!!**



## But I Did All This And There Are Still Bugs!!!

- “Be Careful!!! It Is A Real World Out There!!!”
  - Might your program be non-trivial?
- The purpose of careful software-development practices is to reduce risk
  - Strive for perfection, but understand that this goal is rarely reached in this world

## But I Did All This And There Are Still Bugs!!!

- “Be Careful!!! It Is A Real World Out There!!!”
  - Might your program be non-trivial?
- The purpose of careful software-development practices is to reduce risk
  - Strive for perfection, but understand that this goal is rarely reached in this world
- **But you still need to fix your bugs!!!**

## Fixing Bugs

- The first challenge is locating the bugs

## Fixing Bugs

- The first challenge is locating the bugs
  - The computer knows where the bugs are

## Fixing Bugs

- The first challenge is locating the bugs
  - The computer knows where the bugs are
  - So your job is to make it tell you!
- Ways to make the computer tell you where the bugs are:

## Fixing Bugs

- The first challenge is locating the bugs
  - The computer knows where the bugs are
  - So your job is to make it tell you!
- Ways to make the computer tell you where the bugs are:
  - Debugging `printk()`s and assertions
  - Event tracing and `ftrace`
  - Lock dependency checker (`CONFIG_PROVE_LOCKING` and `CONFIG_PROVE_RCU`)
  - Static analysis (and pay attention to compiler warnings!!!)
  - Structured testing: Use an experimental approach
  - **Record all test results, including environment**

## Fixing Bugs

- The first challenge is locating the bugs
  - The computer knows where the bugs are
  - So your job is to make it tell you!
  - But getting another person's viewpoint can be helpful
    - To 10,000 *educated and experienced* eyes, all bugs are shallow
- Gaining other people's viewpoints

## Fixing Bugs

- The first challenge is locating the bugs
  - The computer knows where the bugs are
  - So your job is to make it tell you!
  - But getting another person's viewpoint can be helpful
    - To 10,000 *educated and experienced* eyes, all bugs are shallow
- Gaining other people's viewpoints
  - Have other people review your code
  - Explain your code to someone else
  - Special case of explaining code: Document it
    - Think of questions you might ask if someone else showed you the code
    - Focus on the parts of the code you are most proud of: Most likely buggy!
    - Try making a copy of the code, removing the comments, and then documenting it: Perhaps the comments are confusing you



## But What If The Computer Knows Too Much?

- Event tracing for RCU: 35MB of trace events for failure
- Way too much to read and analyze by hand
- What to do?

## But What If The Computer Knows Too Much?

- Event tracing for RCU: 35MB of trace events for each failure
- Way too much to read and analyze by hand all the time
- What to do? Scripting!!!
- How to generate useful scripts:
  - Do it by hand the first few times
  - But keep detailed notes on what you did and what you found
  - Incrementally construct scripts to carry out the most laborious tasks
  - Eventually, you will have a script that analyzes the failures
- But suppose you are working on many different projects?

## But What If The Computer Knows Too Much?

- Event tracing for RCU: 35MB of trace events for each failure
- Way too much to read and analyze by hand all the time
- What to do? Scripting!!!
- How to generate useful scripts:
  - Do it by hand the first few times
  - But keep detailed notes on what you did and what you found
  - Incrementally construct scripts to carry out the most laborious tasks
  - Eventually, you will have a script that analyzes the failures
- But suppose you are working on many different projects?
  - Script the common cases that occur in many projects
  - Take advantage of tools others have constructed

# Summary and Conclusions

## Summary and Conclusions

- Avoid Debugging The Open-Source Way
- Avoid Debugging By Design
- Avoid Debugging By Process
- Avoid Debugging By Mechanical Proofs
- Avoid Debugging By Statistical Analysis
- Avoid Schedule Pressure via Long-Term View
- But Even If You Do All This, There Will Still Be Some Bugs (<http://lwn.net/Articles/453002/>)
  - Yes, you are living in the real world!!!
  - Might be painful sometimes, but it sure beats all known alternatives...

## Legal Statement

- This work represents the view of the author and does not necessarily represent the view of IBM.
- IBM and IBM (logo) are trademarks or registered trademarks of International Business Machines Corporation in the United States and/or other countries.
- Linux is a registered trademark of Linus Torvalds.
- Other company, product, and service names may be trademarks or service marks of others.

# Questions?