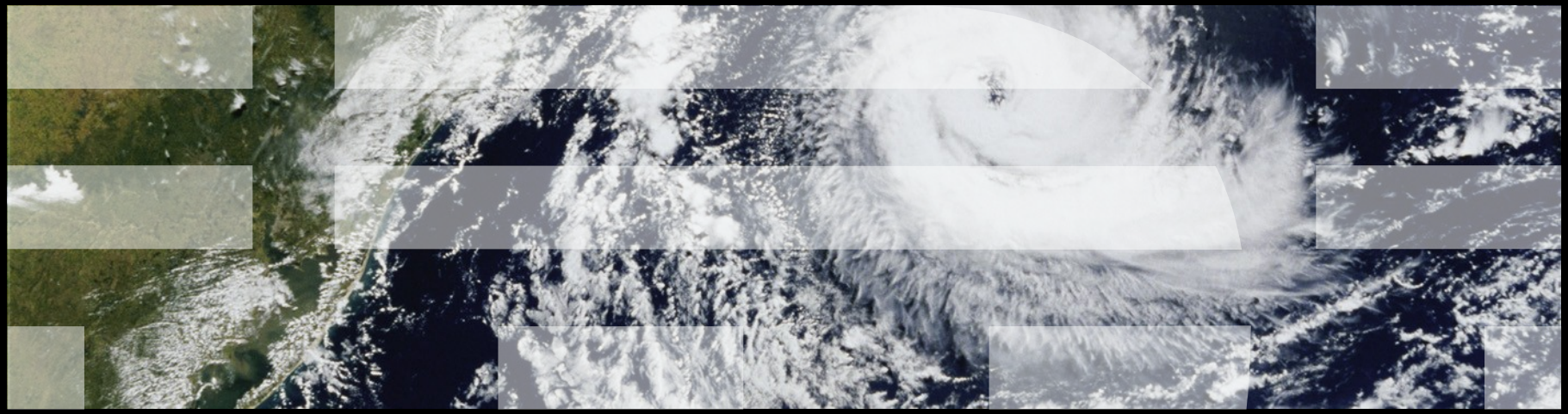
Paul E. McKenney, IBM Distinguished Engineer, Linux Technology Center (Linaro)

November 1, 2012

CPU Hotplug, RCU, and big.LITTLE



2012 Linux Plumbers Conference, Scheduling Microconference

© 2009 IBM Corporation

CPU Hotplug: Work In Progress

© 2009 IBM Corporation

CPU Hotplug: Work In Progress

- Reduce task creation/destruction overhead
  - Thomas Gleixner's generic-idle (mainline) and park/unpark (-tip)
  - Tejun Heo's fix for workqueues
- Numerous fixes from Srivatsa Bhat (3.4-3.6)
- Disturbance-free SRCU (Lai Jiangshan, 3.4)
- Wean RCU from __stop_machine() (PEM, 3.5-3.7)
- Wean scheduler from __stop_machine() (Peter, 3.7)
- Guard for_each_online_cpu() (Silas Boyd-Wickizer, 3.7-3.8)
- Switch hotplug from __stop_machine() (PEM, 3.8?)
- Tell RCU exactly when CPU goes offline (PEM, 3.8?)
- Reverse notifier order, reworking notifier concepts (TBD)
- Apply park/unpark everywhere (TBD)
- Good progress, some work left

© 2009 IBM Corporation

RCU Callbacks, HPC, Real Time and big.LITTLE

© 2009 IBM Corporation

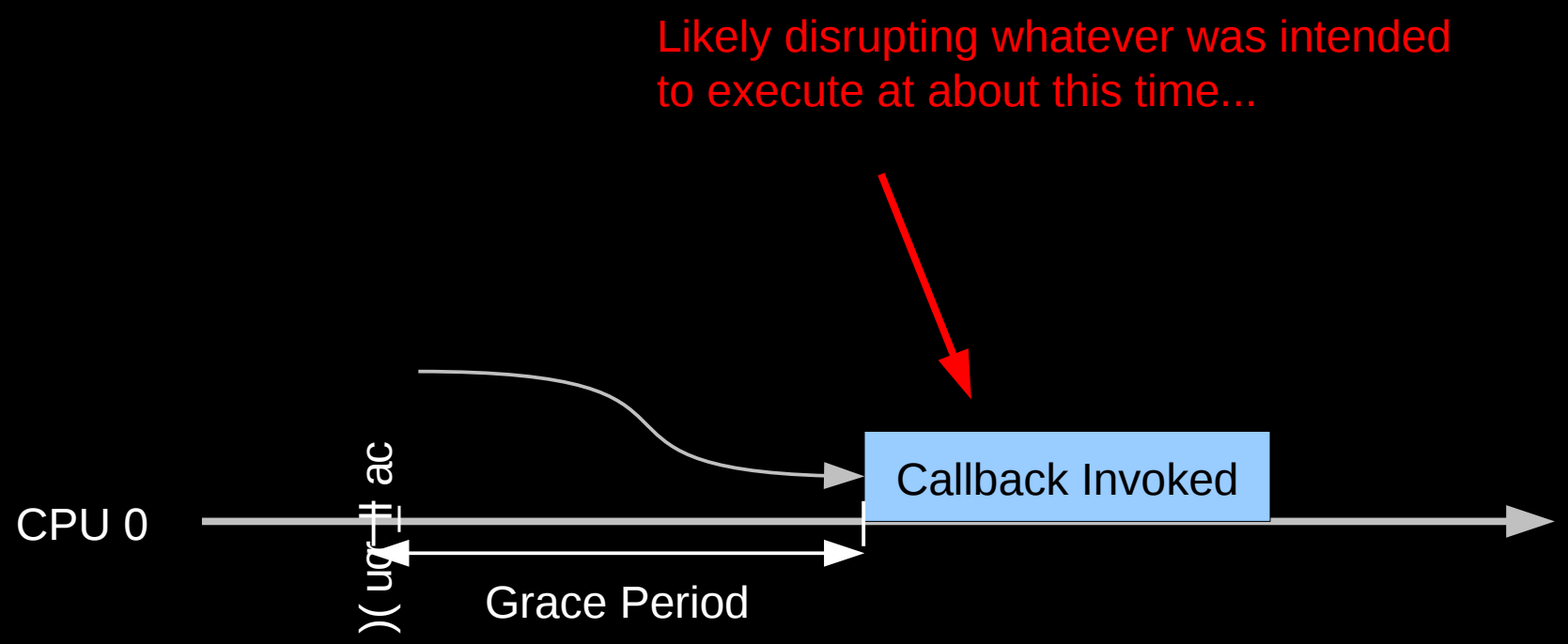
RCU Callbacks, HPC, Real Time and big.LITTLE

- For an overview, see http://lwn.net/Articles/262464/
- For the purposes of this presentation, think of RCU as something that defers work, with one work item per callback
  - Each callback has a function pointer and an argument
  - Callbacks are queued on per-CPU lists, invoked after grace period
    - Invocation can result in OS jitter and real-time latency
  - Global list handles callbacks from offlined CPUs: adopted quickly

© 2009 IBM Corporation

The Problem With RCU Callbacks

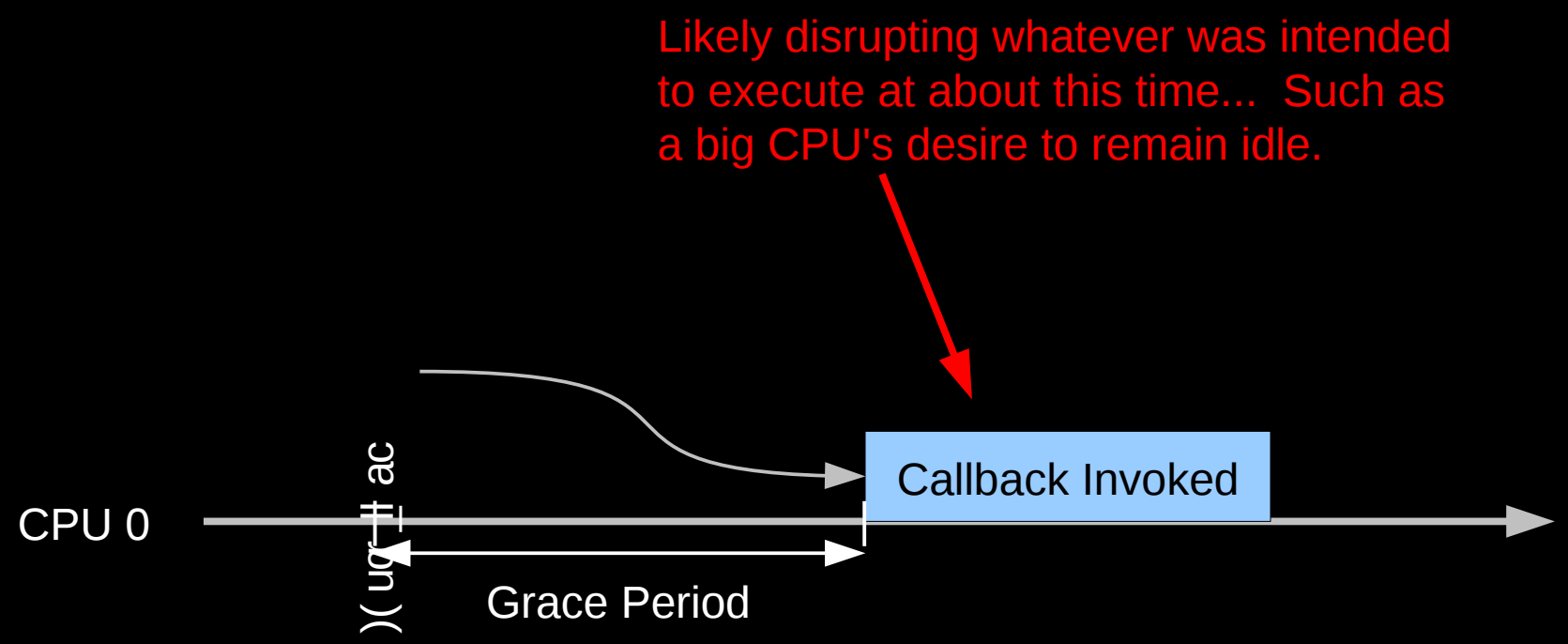
- The problem with RCU callbacks:



© 2009 IBM Corporation

The Problem With RCU Callbacks

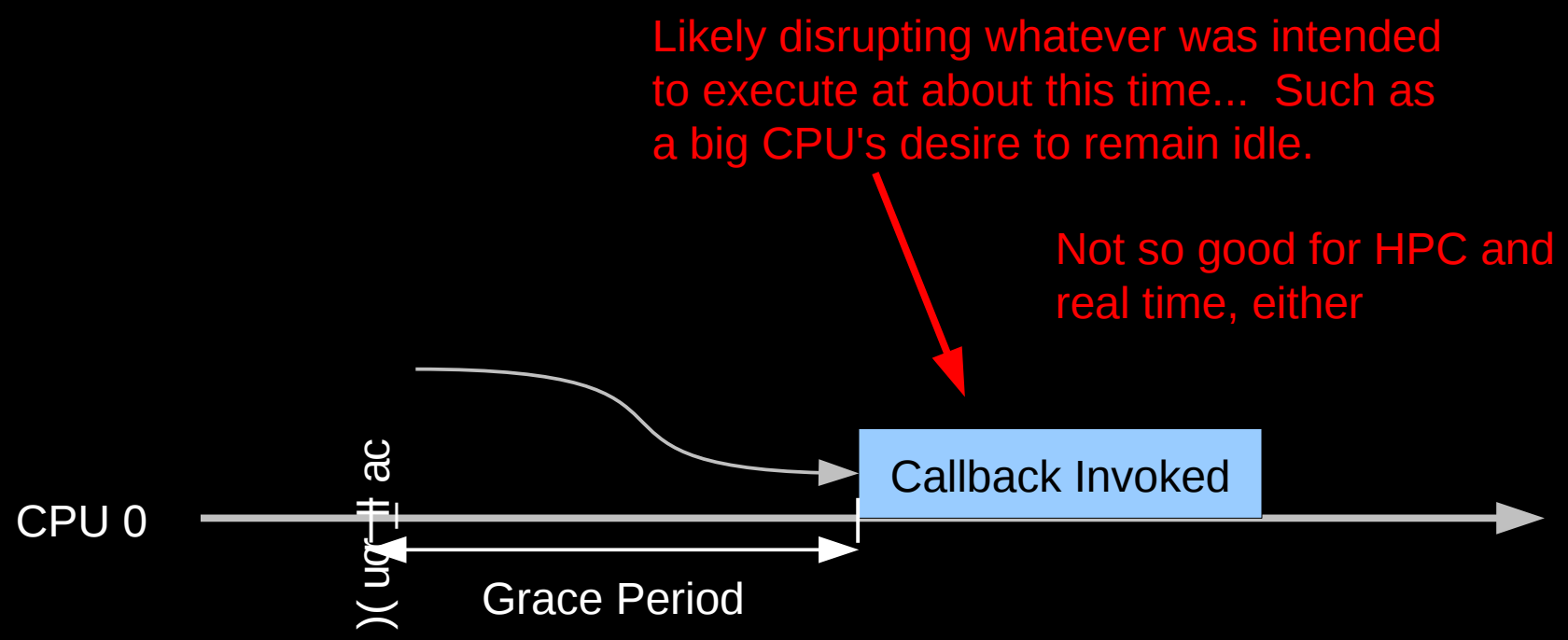
- The problem with RCU callbacks:



© 2009 IBM Corporation

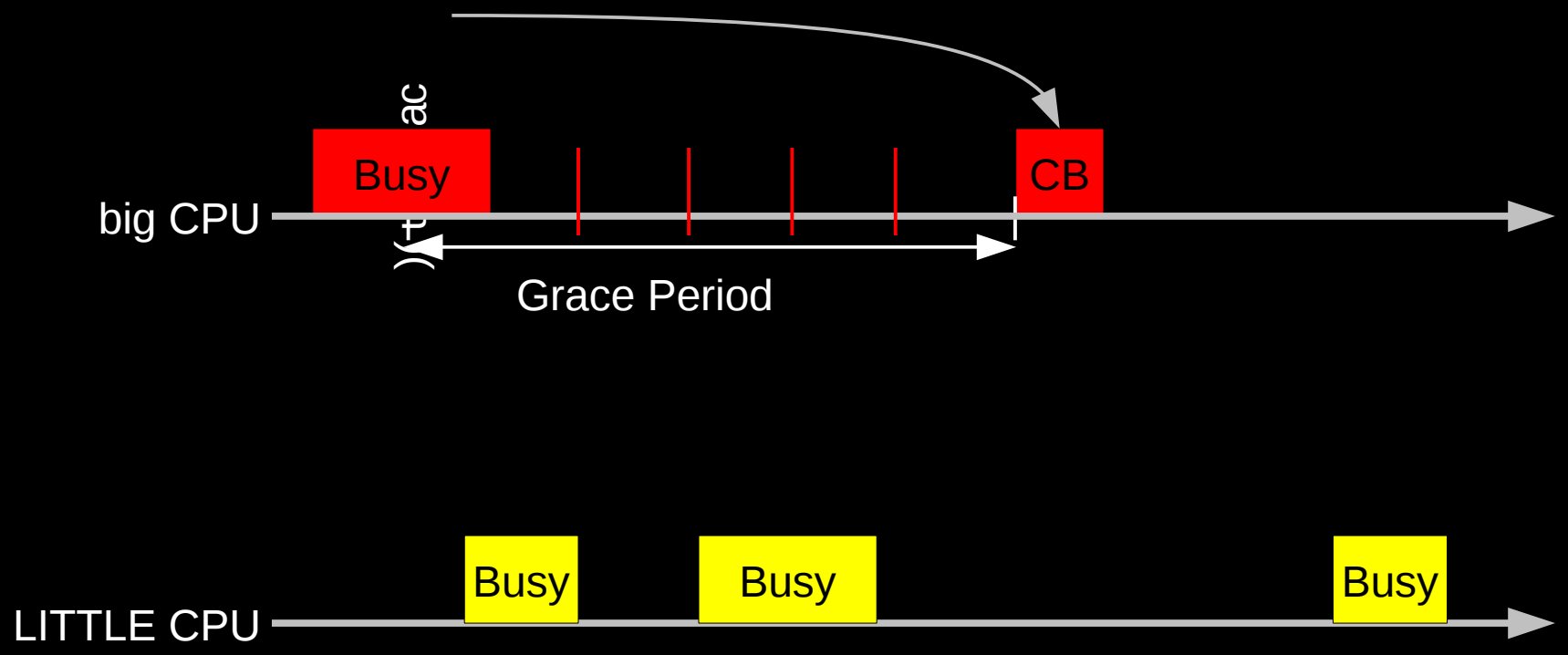
The Problem With RCU Callbacks

- The problem with RCU callbacks:



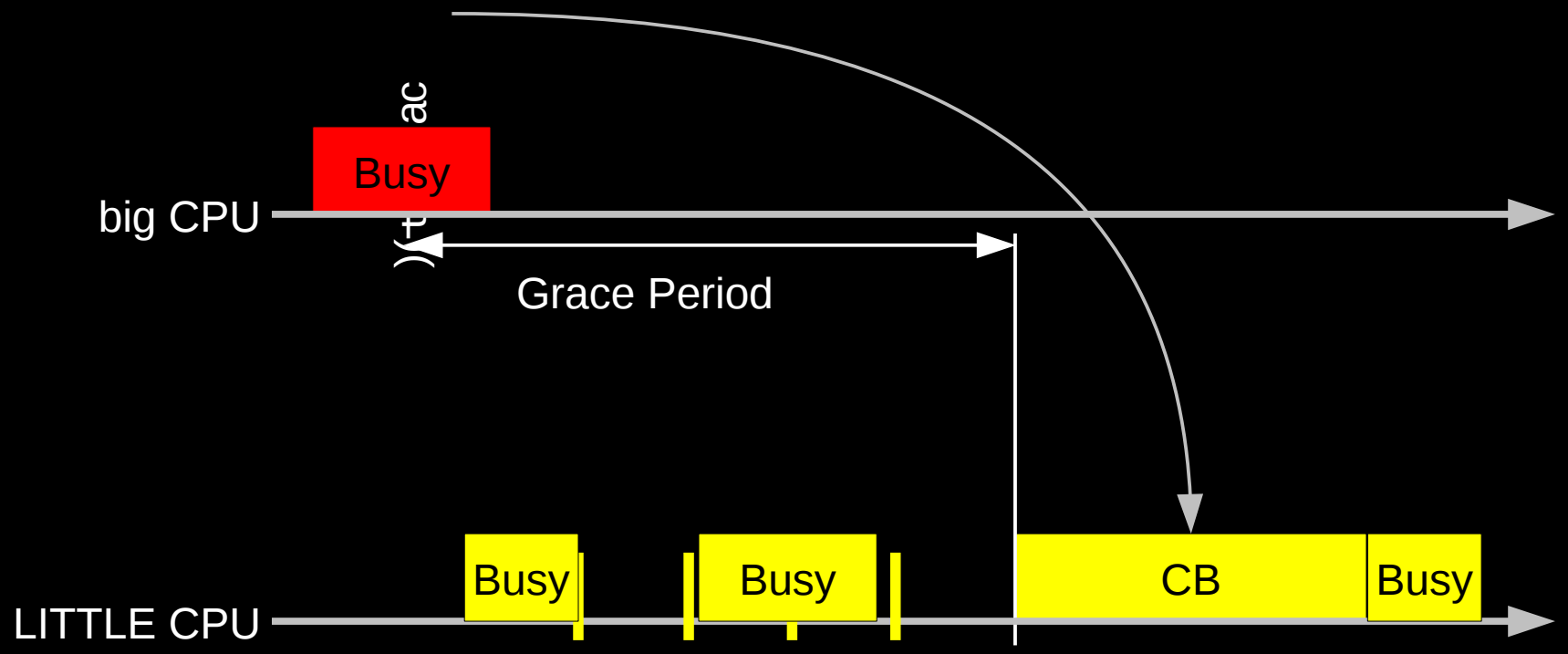
© 2009 IBM Corporation

The Problem With RCU Callbacks and big.LITTLE



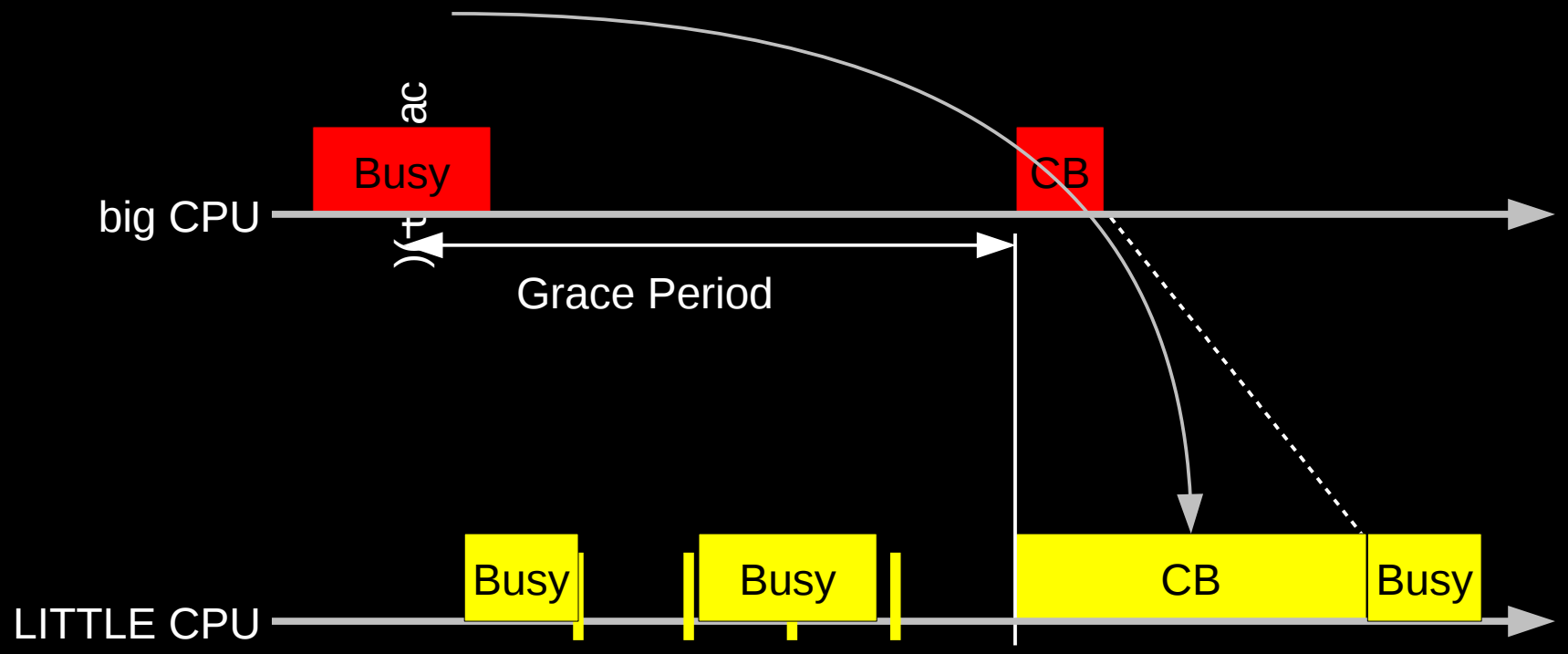
© 2009 IBM Corporation

What big.LITTLE Might Want Instead



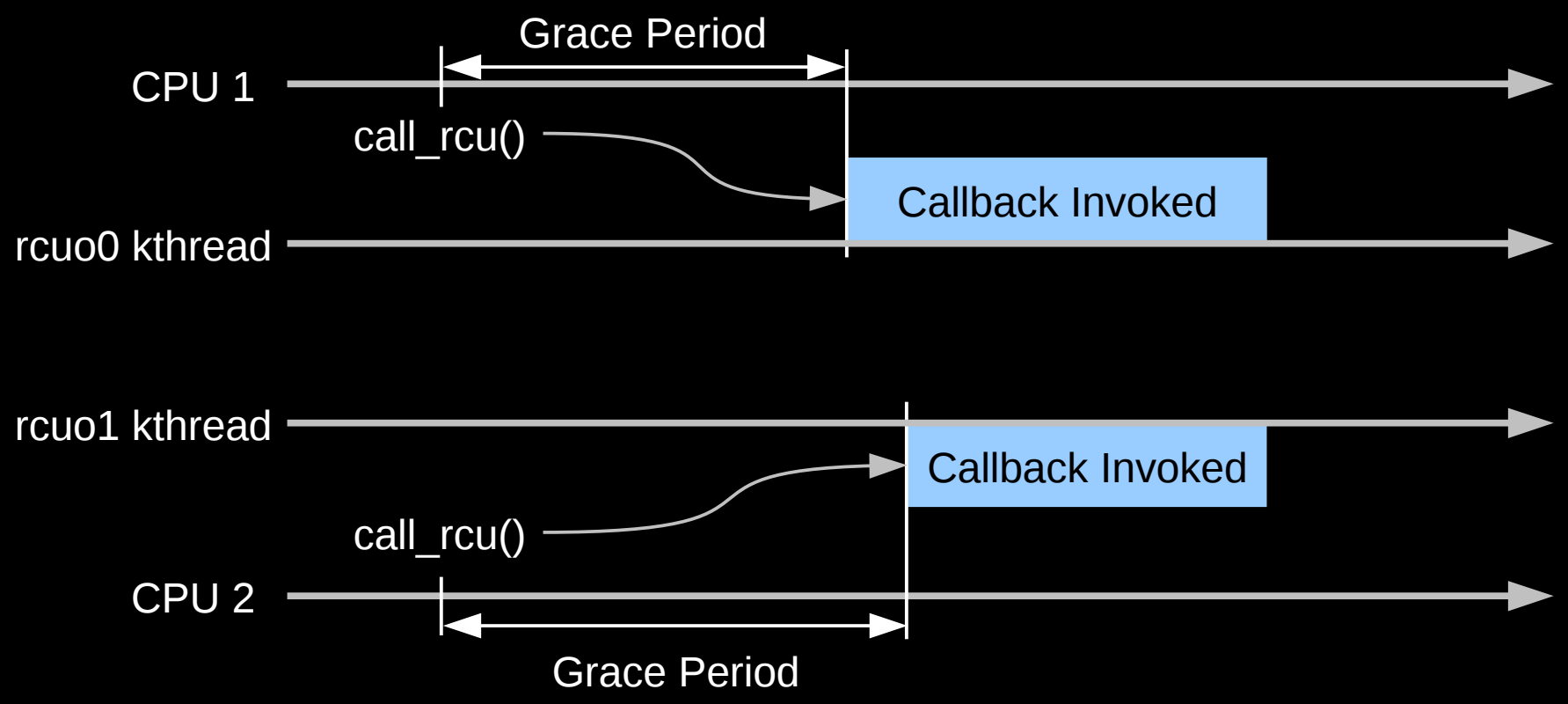
© 2009 IBM Corporation

What big.LITTLE Might Want Instead: But Not Free



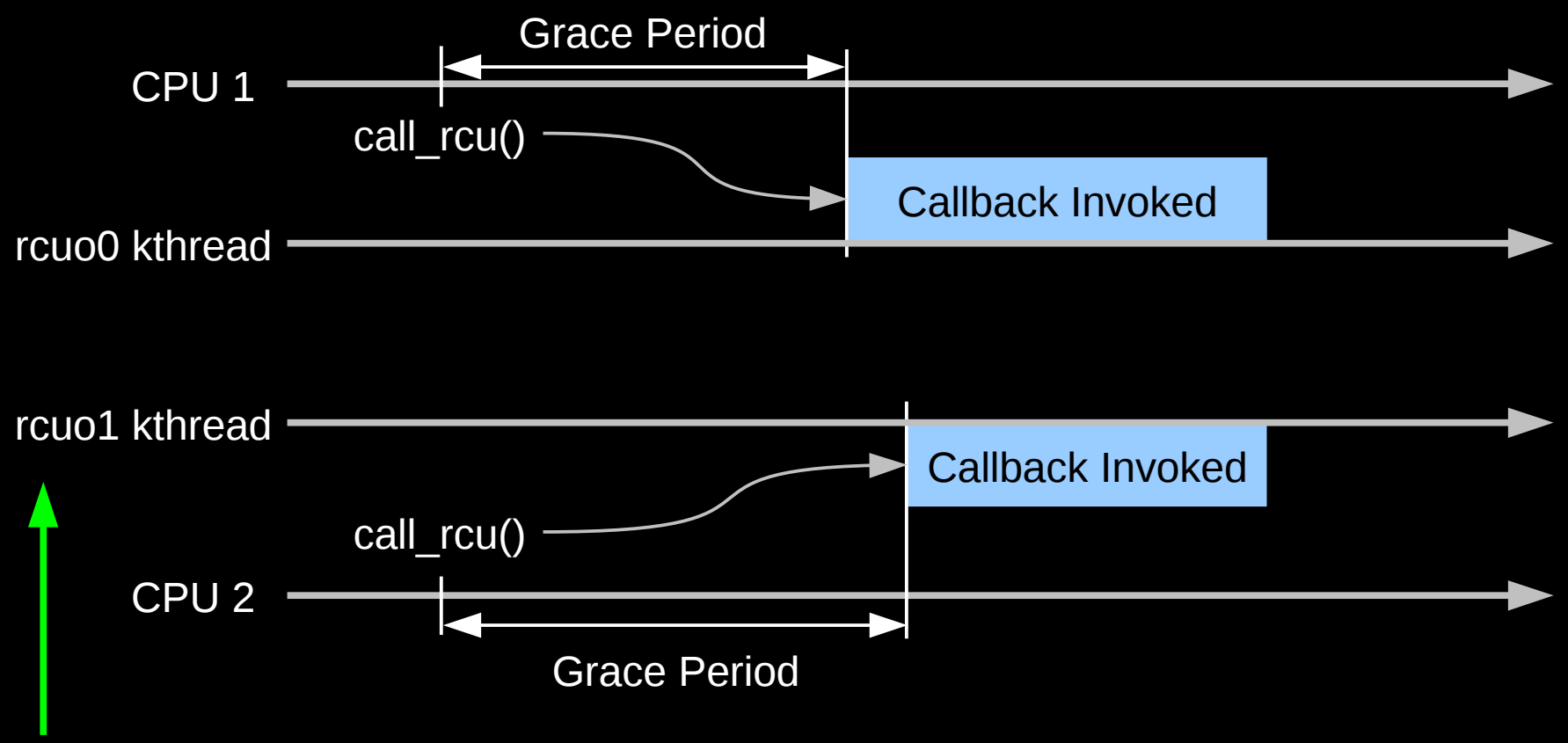
© 2009 IBM Corporation

RCU Callbacks, Houston/Korty for TREE_RCU



© 2009 IBM Corporation

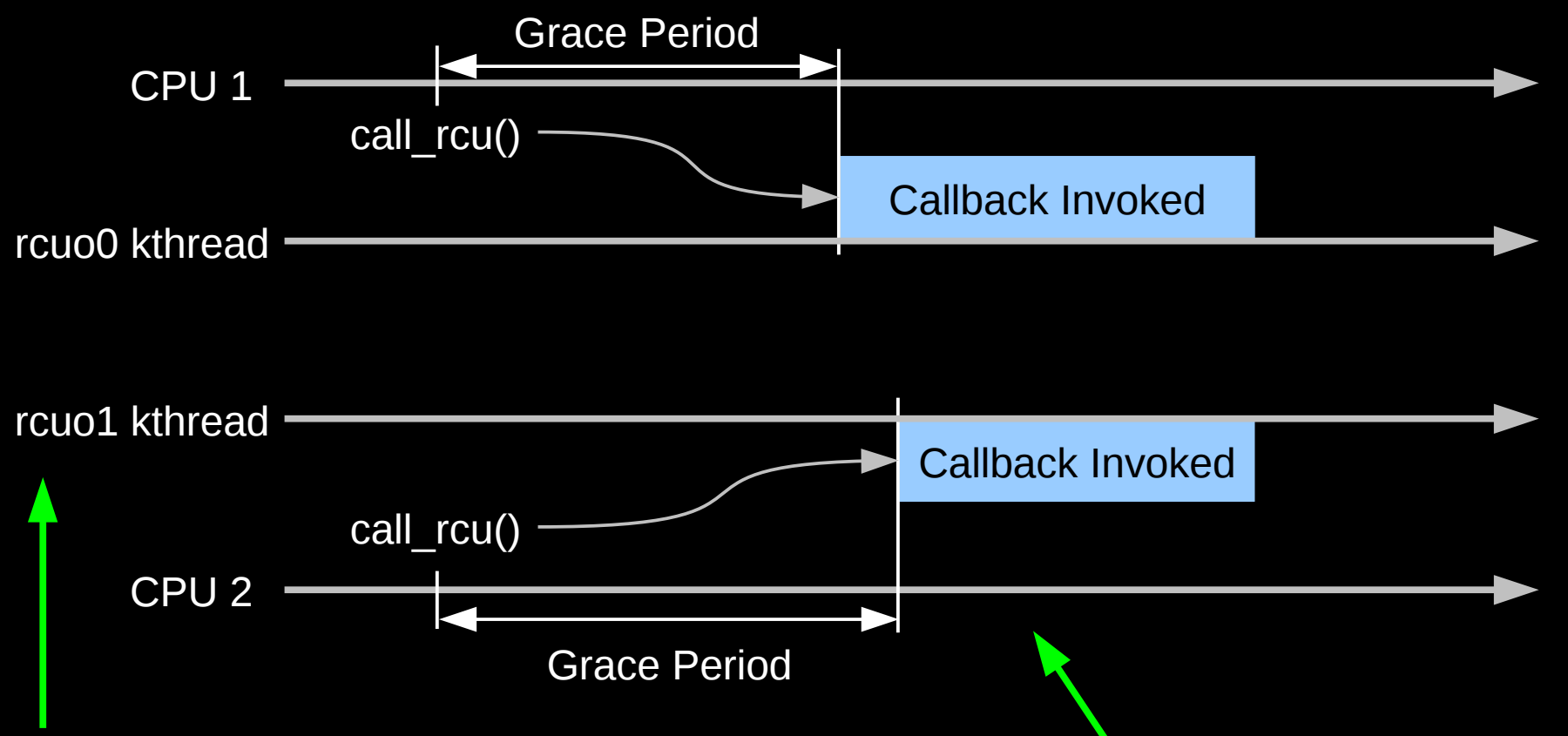
RCU Callbacks, Houston/Korty for TREE_RCU



Manually place rcuoN kthreads
on LITTLE CPUs

© 2009 IBM Corporation

RCU Callbacks, Houston/Korty for TREE_RCU



Manually place rcuoN kthreads on LITTLE CPUs

No big-CPU wakeups!!!

© 2009 IBM Corporation

Offloadable RCU Callbacks: Limitations and Futures

- Must reboot to reconfigure no-CBs CPUs
  - Should be just fine for big.LITTLE
  - Hopefully also OK for HPC and real-time workloads
- Scalability: 1,000 no-CBs CPUs would not do well
  - Not a problem for big.LITTLE, but must be solved for HPC
- Must be at least one non-no-CBs CPU (namely, CPU 0)
  - OK if CPU 0 is a LITTLE CPU, but expect to relax this constraint “real soon now”
- No energy-efficiency code: lazy & non-lazy CBs? Non-lazy!
  - But non-lazy Cbs are common case, so deferring interpretation of laziness.
- No-CBs CPUs' kthreads not subject to priority boosting
  - Probably not a near-term problem for big.LITTLE
- Setting all no-CBs CPUs' kthreads to RT prio w/out pinning them: bad!
  - At least on large systems: Probably OK near-term for big.LITTLE
- Note: I do not expect no-CBs path to completely replace current CB path

© 2009 IBM Corporation

Offloadable RCU Callbacks: Plans

- Initial version in 3.8 or 3.9, depending on testing results
  - Thanks to Robin Randhawa and Paul Gortmaker for testing efforts thus far

© 2009 IBM Corporation

To Probe More Deeply...

- What Is New In RCU for Real Time (RTLWS 2012)
  - http://www.rdrop.com/users/paulmck/realtime/paper/RTLWS2012occcRT.2012.10.19e.pdf
    - Slides 21-on
- Getting RCU Further Out of the Way (Plumbers 2012)
  - http://www.rdrop.com/users/paulmck/realtime/paper/nocb.2012.08.31a.pdf
- Cleaning Up Linux's CPU Hotplug For Real Time and Energy Management (ECRTS 2012)
  - http://www.rdrop.com/users/paulmck/realtime/paper/hotplug-ecrts.2012.06.11a.pdf
- And **hot off the press**: "Relocating RCU callbacks"
  - http://lwn.net/Articles/522262/

© 2009 IBM Corporation

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.

© 2009 IBM Corporation

Questions?

© 2009 IBM Corporation

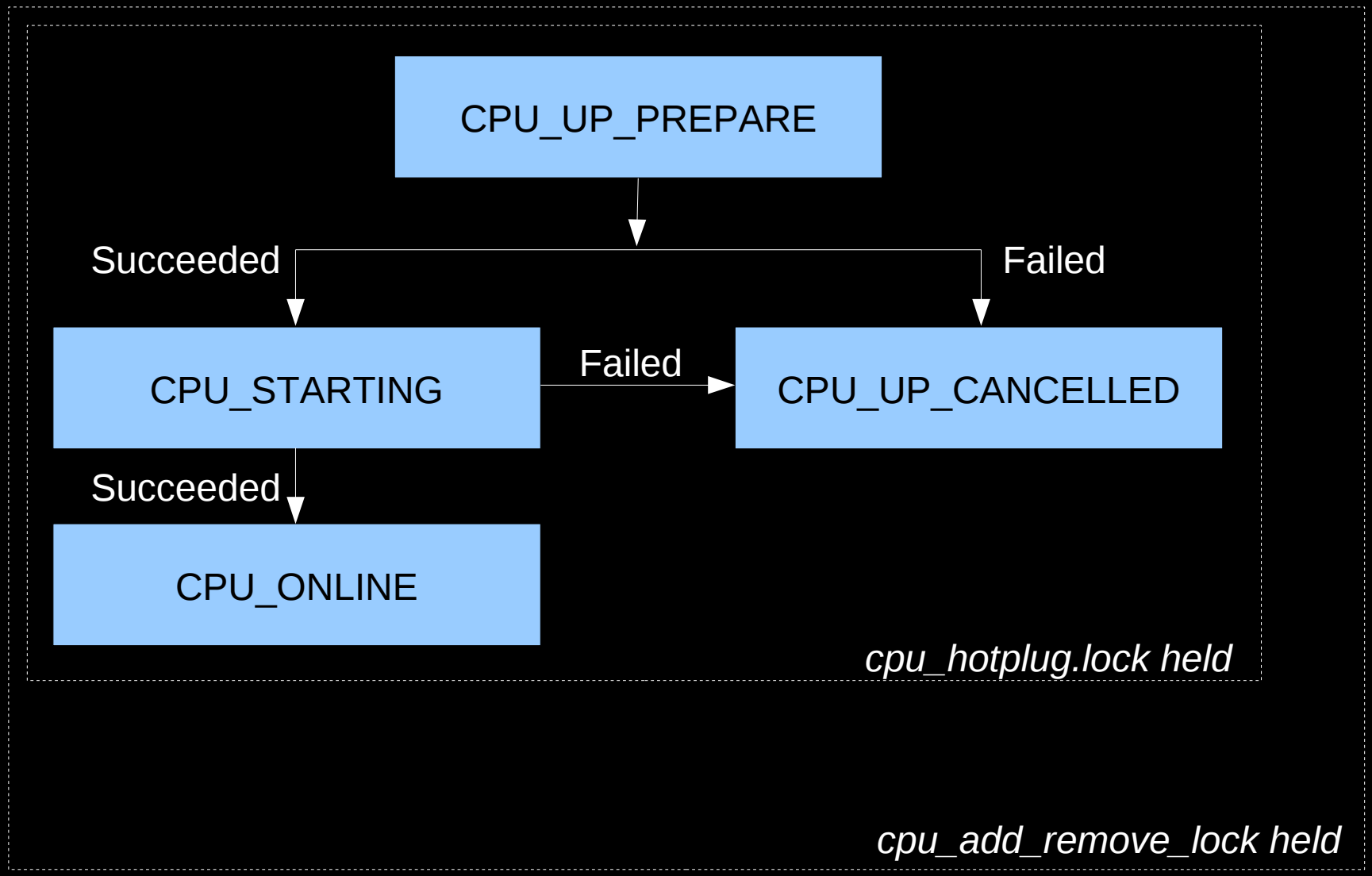
Scheduling Topics

- Target CPU Selection (Vincent Guittot)
- Sharing Scheduler Information (Vincent Guittot)
- Task Placement for Asymmetric Cores (Morten Rasmussen)
- Scheduling and the big.LITTLE Architecture ***(you are here!)***
- Dynamic CPU Core Management (Peter De Schrijver and Antti Miettinen)
- Application of Deadline Scheduling for Power-Saving Strategies (Juri Lelli)

- Main focus of this talk is progress on CPU hotplug
  - See Peter's and Antti's talk for alternatives

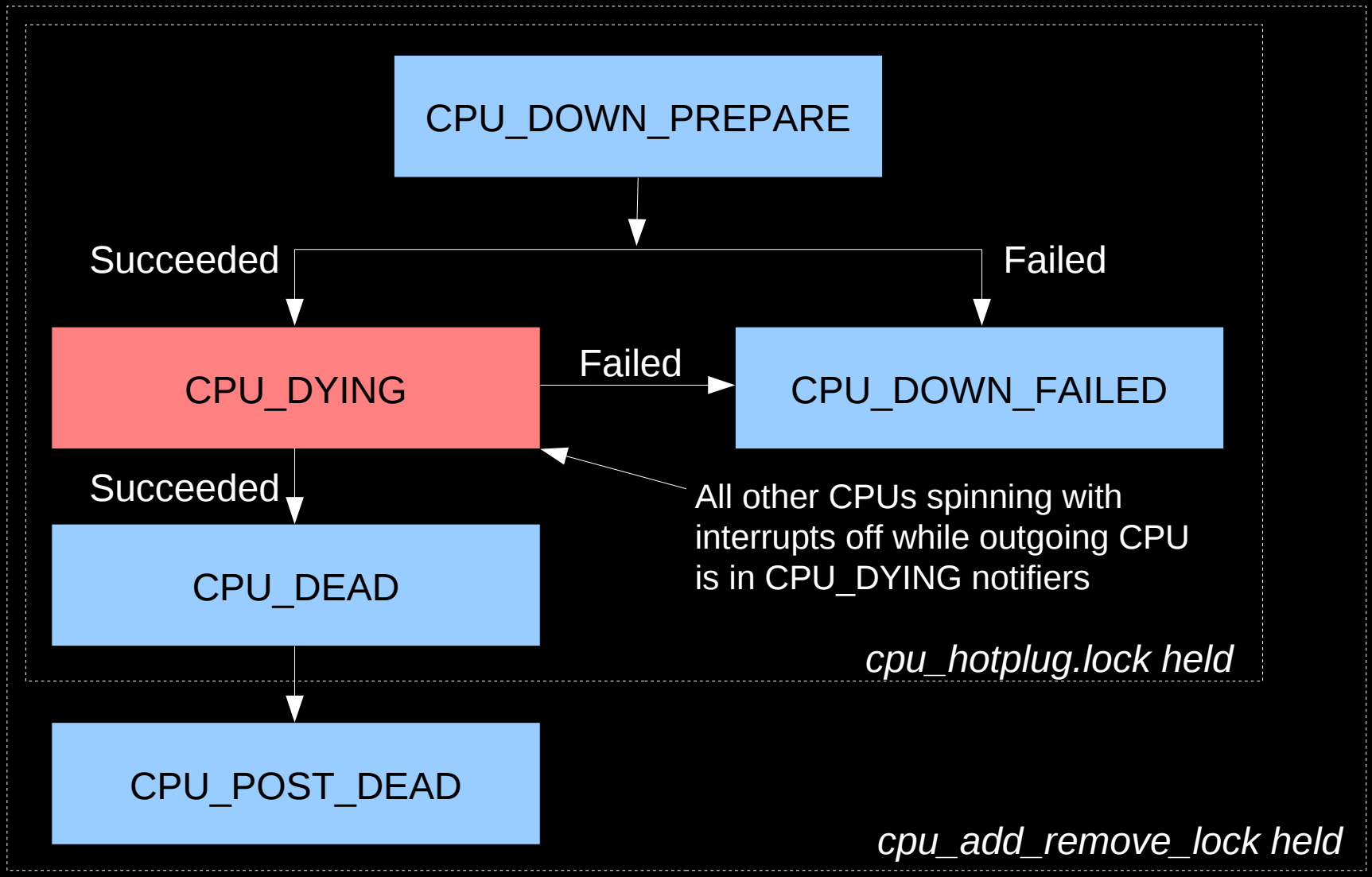
© 2009 IBM Corporation

CPU Hotplug Primer: Online Process



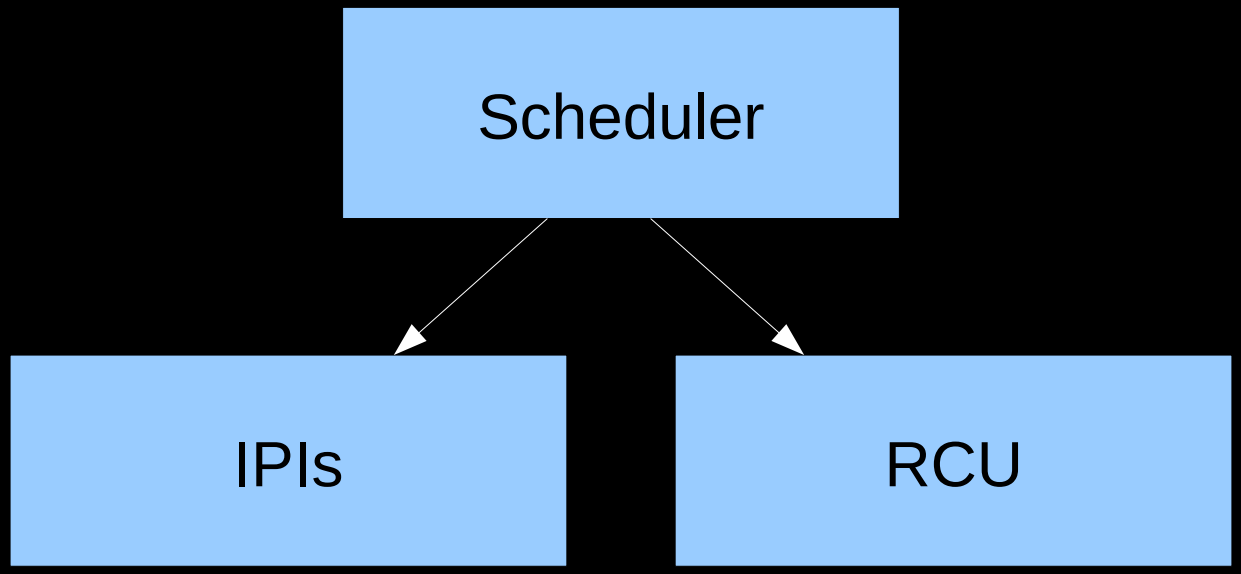
© 2009 IBM Corporation

CPU Hotplug Primer: Offline Process



© 2009 IBM Corporation

CPU Hotplug Is Not Atomic



Valid notifier order for online:
- IPIs
- RCU
- Scheduler

Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

CPU Hotplug: Valid Online Order (1/3)

IPIs

Valid notifier order for online:
- IPIs
- RCU
- Scheduler

Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

CPU Hotplug: Valid Online Order (2/3)

IPIs

RCU

Valid notifier order for online:
- IPIs
- RCU
- Scheduler

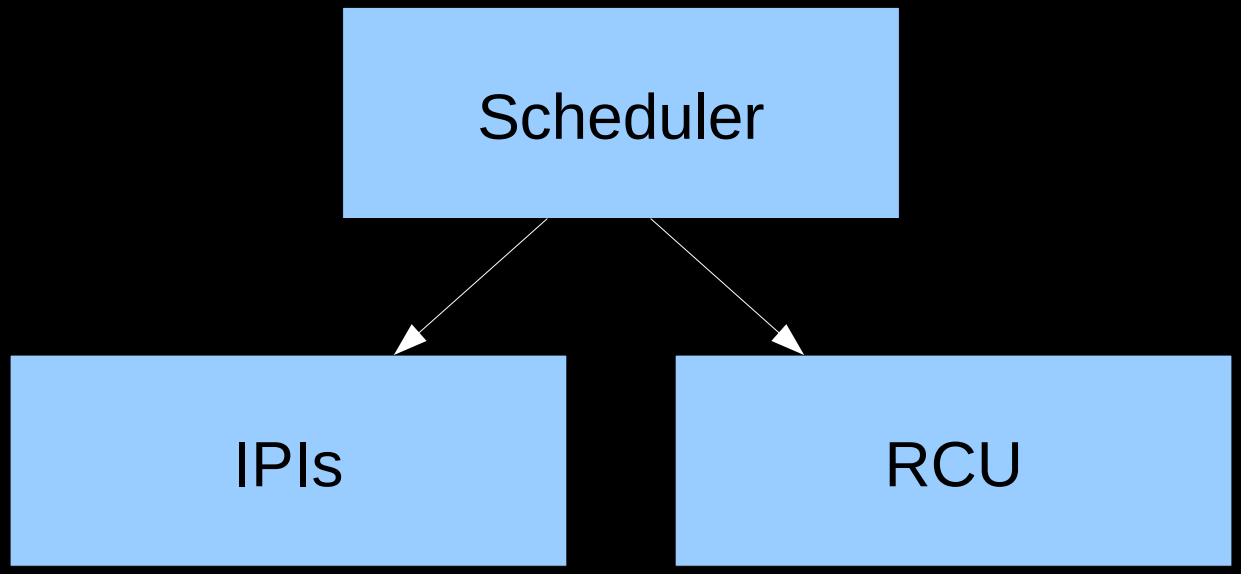
Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

CPU Hotplug: Valid Online Order (3/3)



Valid notifier order for online:
- IPIs
- RCU
- Scheduler

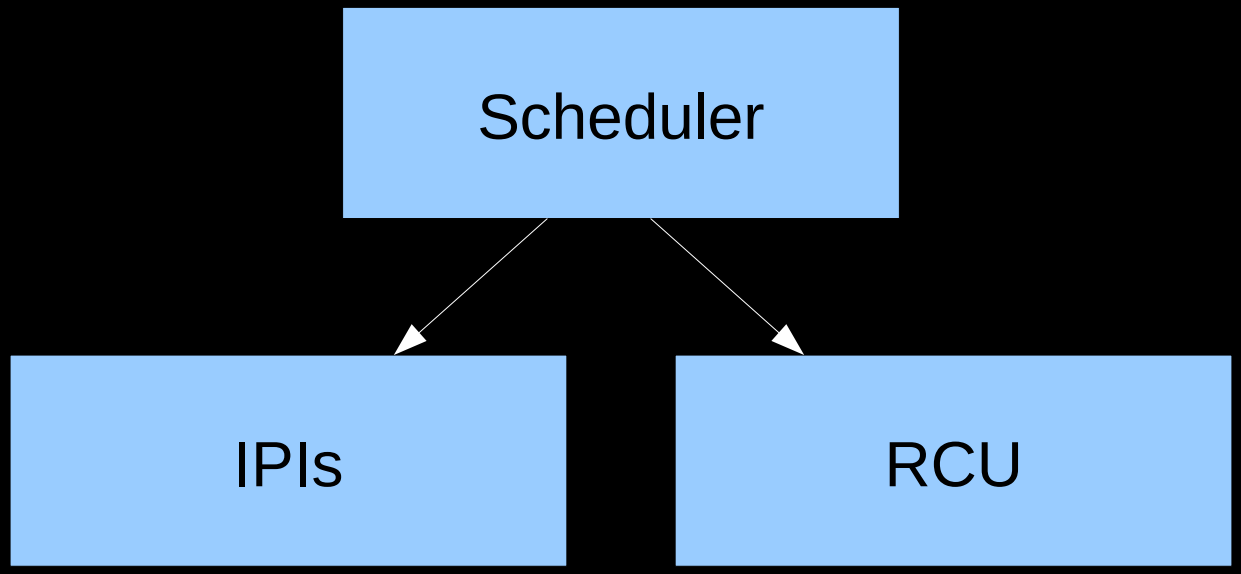
Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

Suppose We Offline In The Same Order...



Valid notifier order for online:
- IPIs
- RCU
- Scheduler

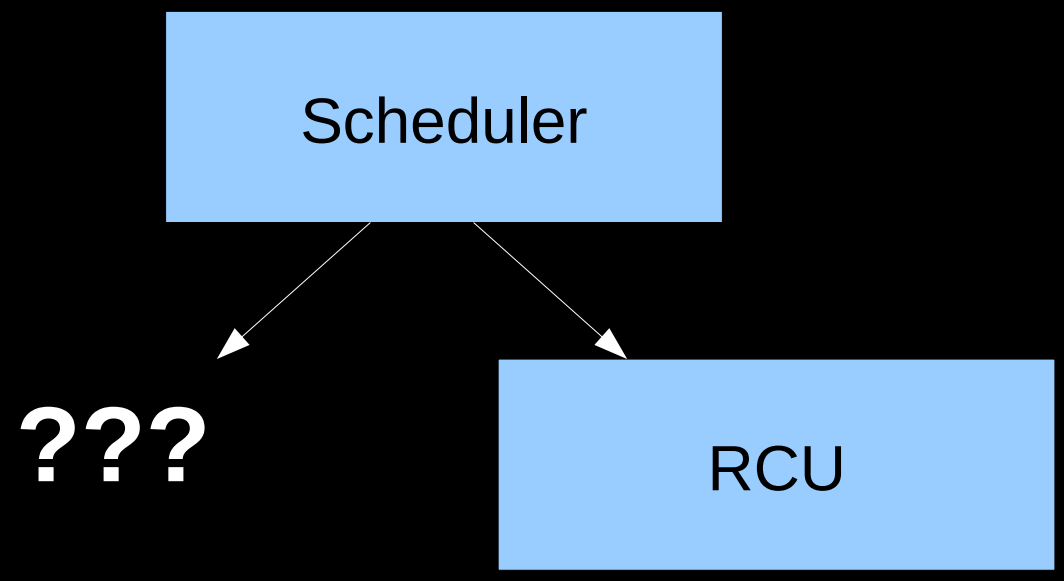
Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

CPU Hotplug: Invalid Offline Order



Valid notifier order for online:
- IPIs
- RCU
- Scheduler

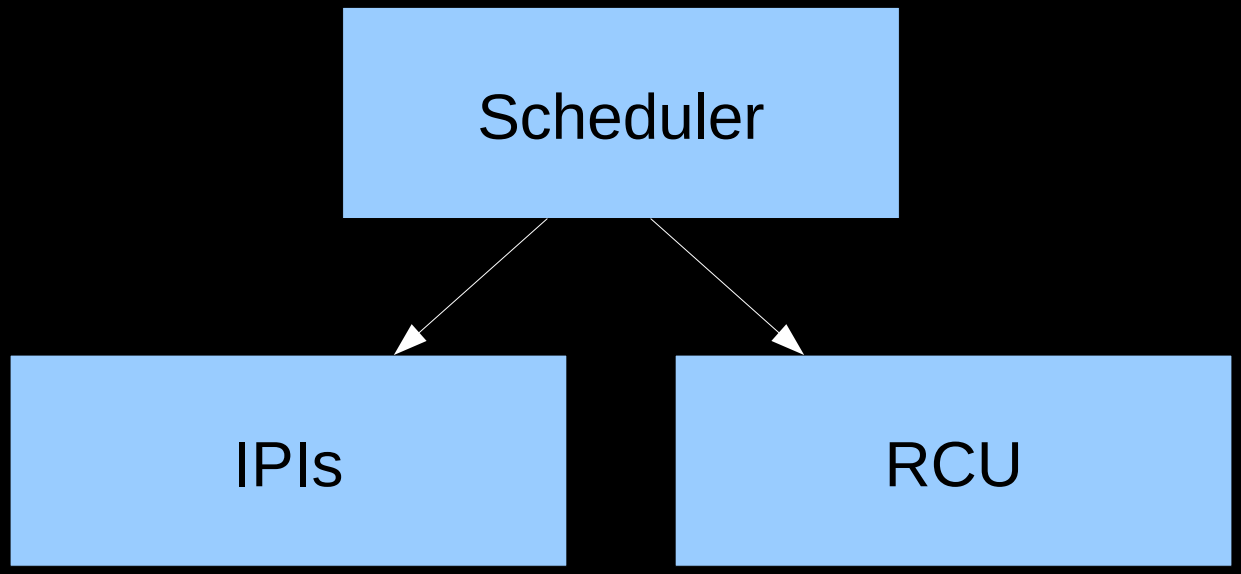
Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

CPU Hotplug: Valid Offline Order (1/3)



Valid notifier order for online:
- IPIs
- RCU
- Scheduler

Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

CPU Hotplug: Valid Online Order (2/3)

IPIs

RCU

Valid notifier order for online:
- IPIs
- RCU
- Scheduler

Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

CPU Hotplug: Valid Offline Order (3/3)

IPIs

Valid notifier order for online:
- IPIs
- RCU
- Scheduler

Must reverse order for offline

Reality will intrude...
- RCU depends on scheduler
- Circular dependency!

Must further decompose RCU and scheduler interaction

© 2009 IBM Corporation

Other Topics

- Synthetic Mobile Workloads
  - Progress, but no nice solution yet
- Emulating big.LITTLE on commodity systems
  - An embarrassment of riches: http://lwn.net/Articles/501501/

© 2009 IBM Corporation

Summary

- Excellent progress on many fronts
- But much work still to be done!

© 2009 IBM Corporation