China Linux Kernel Conference

Carbon neutral – we can help

- Power Management in Linux

Zhang Rui (rui.zhang@intel.com)

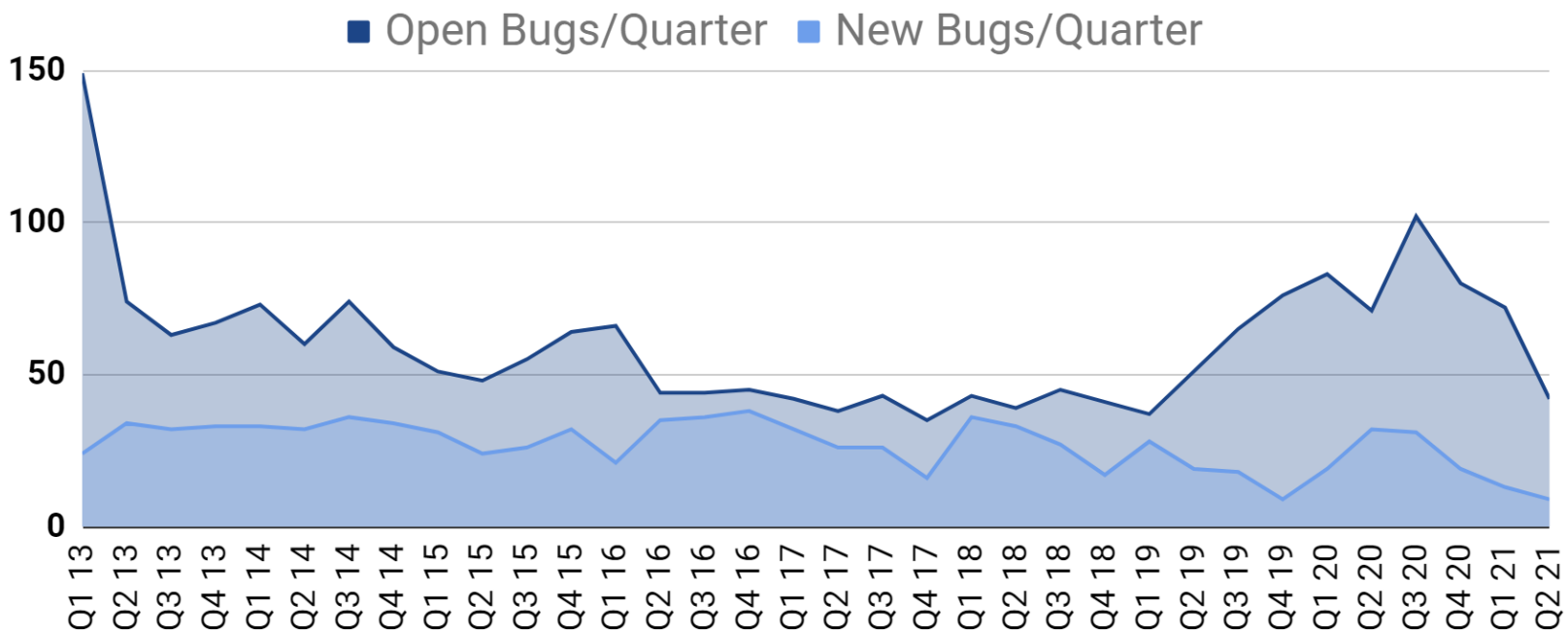
Intel Linux OS Kernel Engineer

Team introduction

Team Introduction

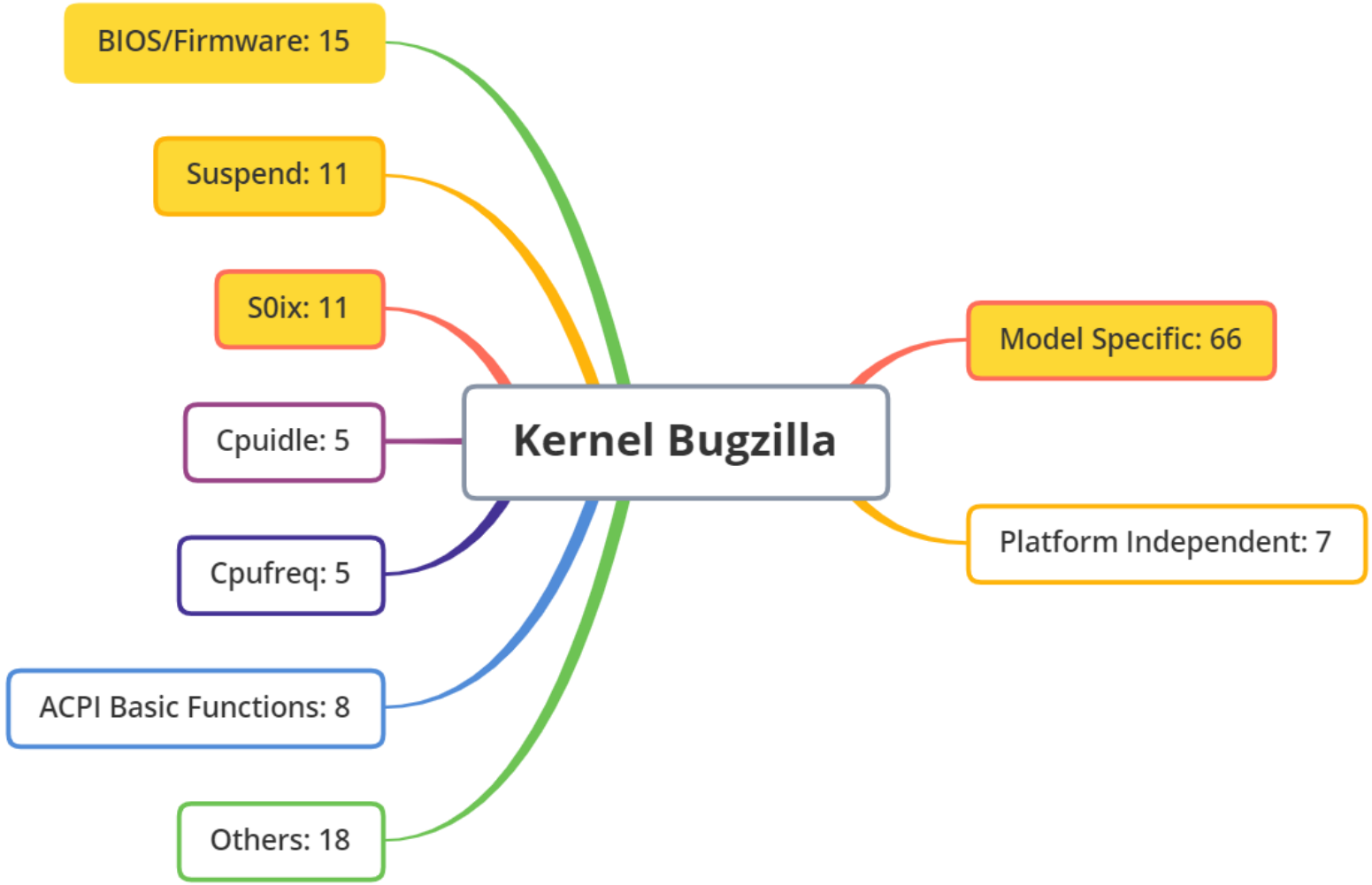
Kernel Bugzilla Status

- Top 1 priority
- 15+ years
- ~25 new bugs per quarter



https://bugzilla.kernel.org/enter_bug.cgi

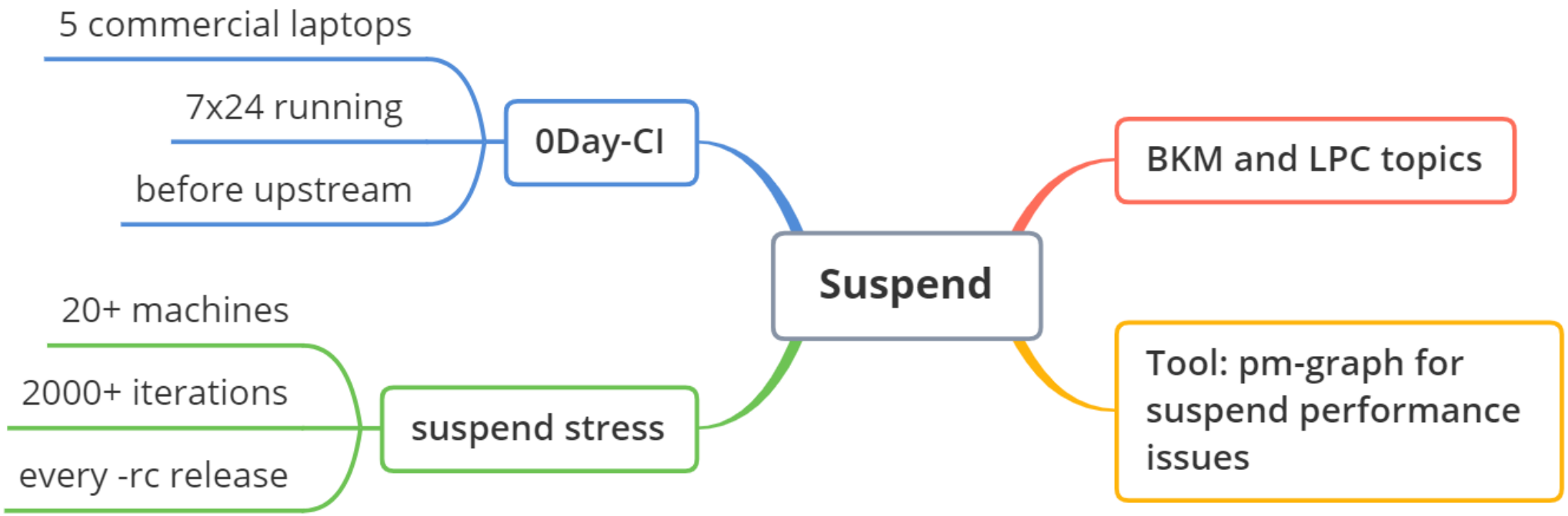
Kernel Bugzilla Analysis



Kernel Bugzilla Challenges/Actions

Model Specific	• Limited hardware coverage and hard to debug remotely
BIOS/Firmware	• Hard to do low level debug without OEM support
Suspend	• Actions to improve suspend quality and performance
S0ix	• Actions to improve S0ix quality

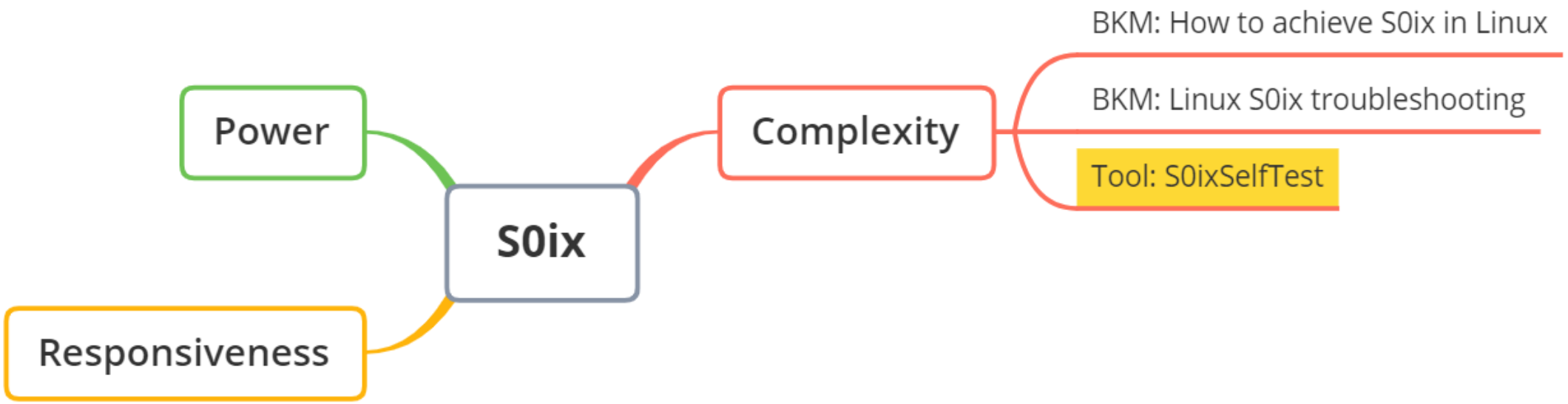
Suspend Quality Improvement



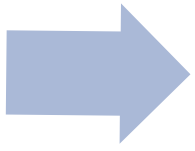
Limited hardware coverage

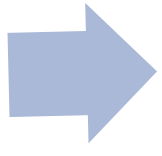
Lack for full stack knowledge

S0ix


BKM: How to achieve S0ix in Linux
BKM: Linux S0ix troubleshooting
Tool: S0ixSelfTest
Complexity
Power
S0ix
Responsiveness

S0ixSelftTestTool





- **Time: days**
- **Linux engineer**

- turbostat/Powertop
- Ramp up on S0ix requirements
- Kernel debug interface and workarounds

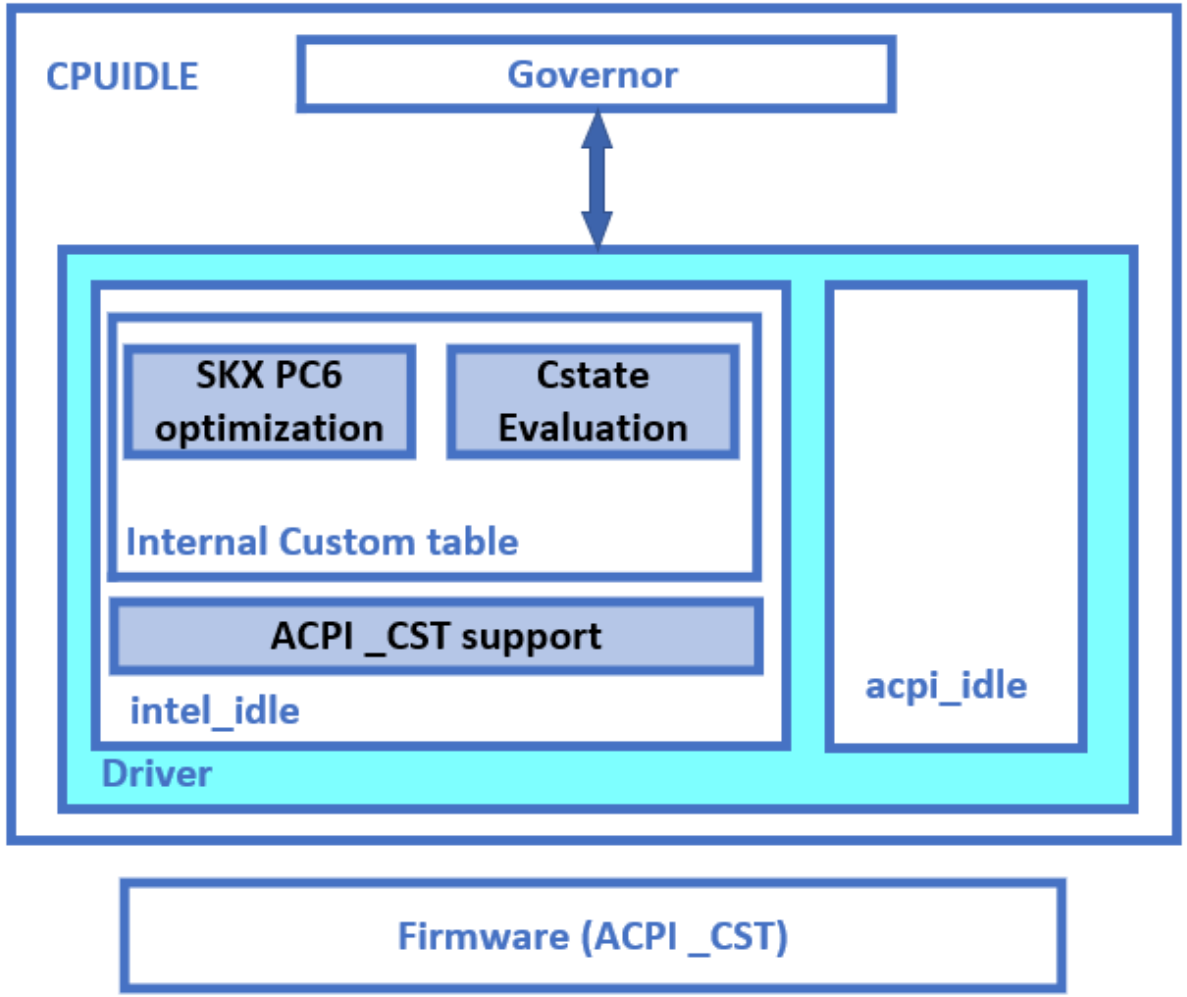
- **Time: hours**
- **Experienced Linux PC10/S0ix engineer**

- intel_pmc_core sysfs
- PCI D3 status check
- Kernel workarounds

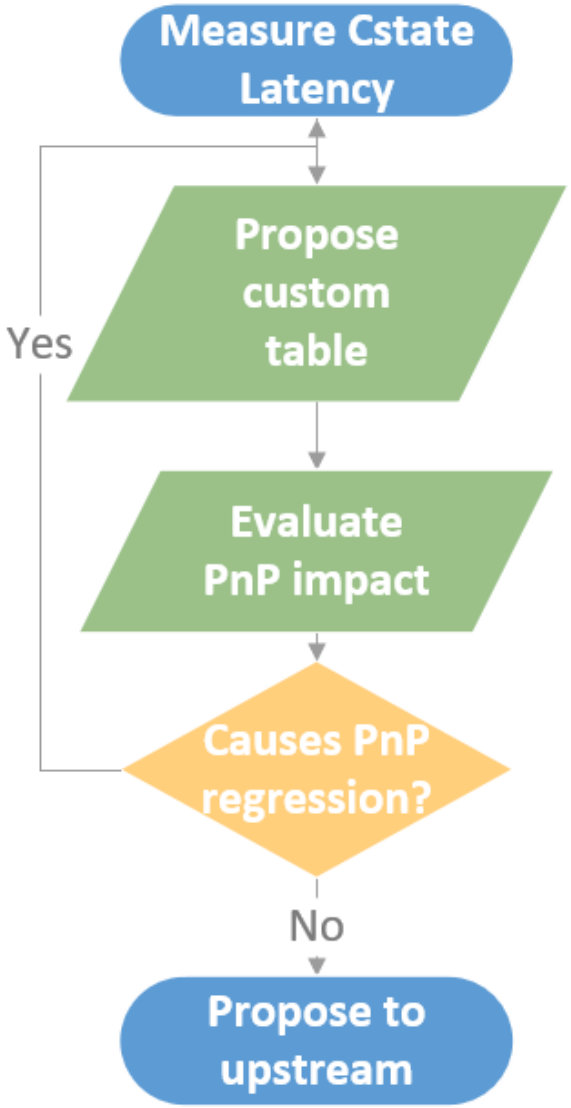
- **Time: 1~3 minutes**
- **Linux user**

- One line command
- kernel workarounds applied automatically
- Quick feedback with potential failure reason and solutions

CPU Power – intel_idle



CPU Power – intel_idle c-state evaluation

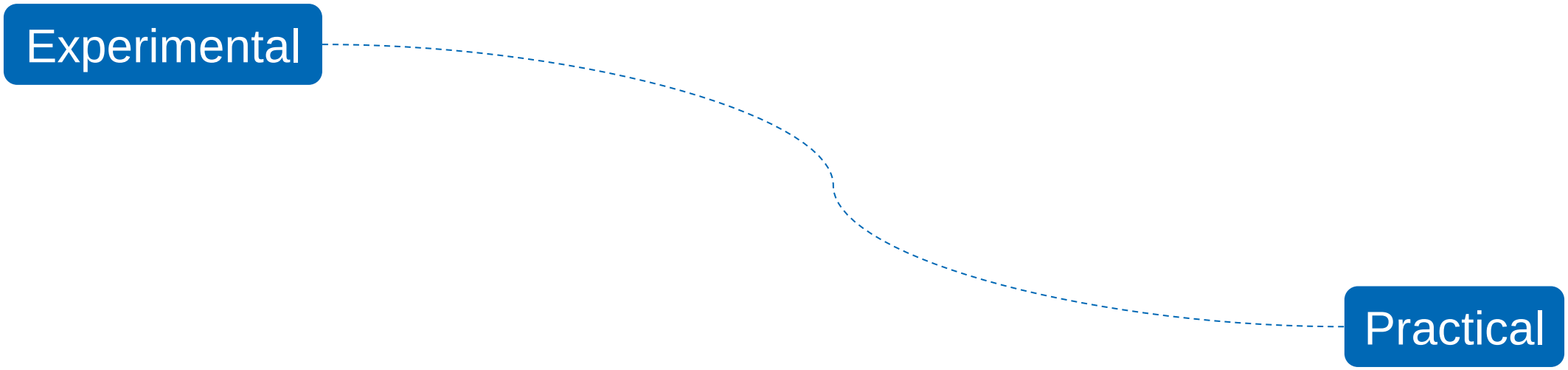


- WULT
  - Schedule a delayed interrupt to happen at future time *LaunchTime*
  - Put CPU into a c-state
  - At *LaunchTime*, the delayed interrupt fires and the CPU starts exiting the C-state
  - When the CPU starts executing instructions, take the time after idle timestamp (*TAI*)
  - The C-state exit latency is *TAI* - *LaunchTime*

Running microbenchmarks, not aligned with production environment

CPU power - optimization opportunities

- Intel CPU Power tunning handbook
  - Motivation
    - Various of CPU power features
    - May behave differently on different generations
    - CPU Power features may get missed or misused (waste extra power/hurt performance)
  - Challenges



Call For Action

- We can improve the Linux PnP altogether
  - Bug reports are appreciated – You help us find issues, for free!
  - Low level support from OEMs would be great.
  - To make your system/product works better in Linux, use S0ixSelfTestTool/pm-graph to help improve the S0ix/suspend quality and performance
  - Feedbacks on the power/latency requirements for server platforms are appreciated

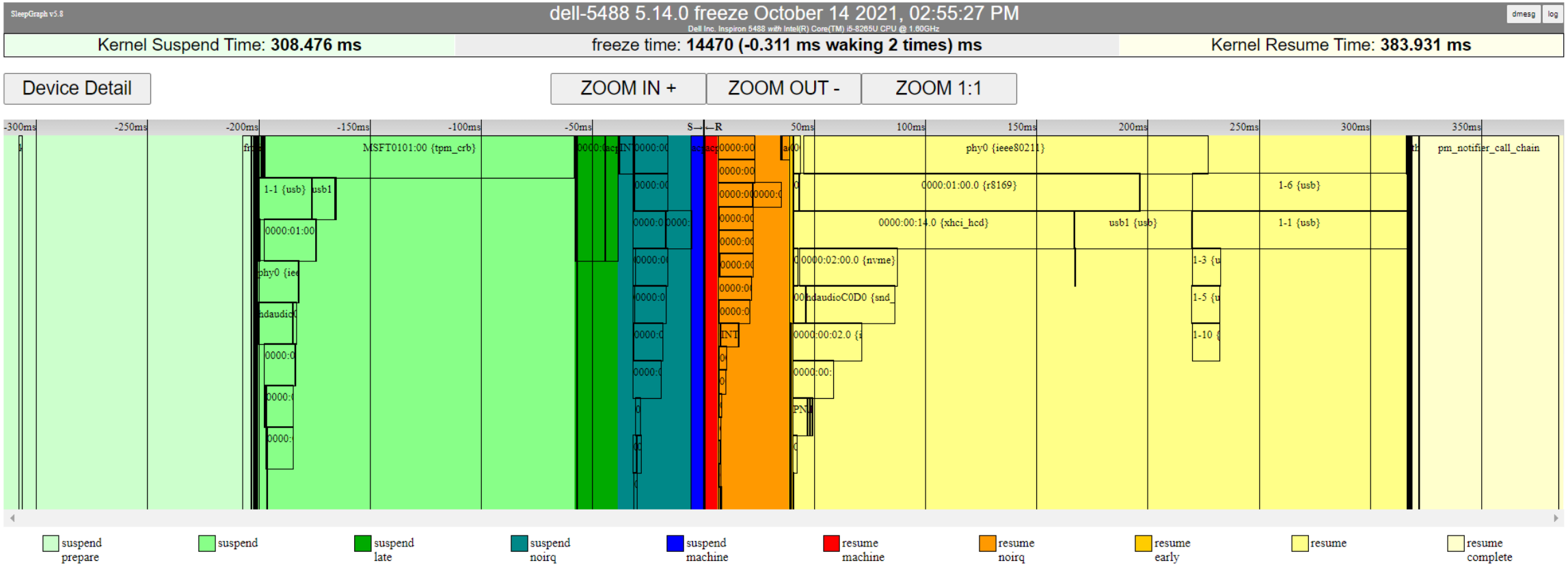
Reference (1/2)

- BKM: Best practice to debug Linux suspend/hibernate issues
  - https://01.org/blogs/rzhang/2015/best-practice-debug-linux-suspend/hibernate-issues
- LPC topic: Linux suspend/resume ..... at the speed of light
  - https://events.static.linuxfound.org/sites/events/files/slides/Brown-Linux-Suspend-at-Speed-of-Light-LC-EU-2015.pdf
- LPC topic: Taking suspend/resume validation into next level
  - https://www.linuxplumbersconf.org/event/4/contributions/499/
- LPC topic: Suspend/resume quality and performance
  - https://www.linuxplumbersconf.org/event/7/contributions/832/
- Tool: pm-graph
  - https://01.org/pm-graph
  - https://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux.git/tree/tools/power/pm-graph

Reference (2/2)

- Meta bug for suspend performance issues
  - https://bugzilla.kernel.org/show_bug.cgi?id=178231
- BKM: How to achieve S0ix states in Linux
  - https://01.org/blogs/qwang59/2018/how-achieve-s0ix-states-linux
- BKM: Linux S0ix troubleshooting
  - https://01.org/blogs/qwang59/2020/linux-s0ix-troubleshooting
- Tool: S0ixSelfTestTool
  - https://github.com/intel/S0ixSelftestTool
- Tool: WULT
  - https://intel.github.io/wult/

Backup: typical pm-graph output



Backup: suspend stress summary using pm-graph (1/2)

host	otcpl-dell-7386-whl	*hostname of the machine where the tests were run*
mode	freeze	*low power mode requested with write to /sys/power/state*
kernel	5.15.0-rc5+	*kernel version or release candidate used (+ means code is newer than the rc)*
summary	summary	*html summary from sleepgraph*
issues	8	*number of unique kernel issues found in test dmesg logs*
target	1440m	*target time or count*
total	2280	*total number of tests run*
pass	2279 (100.0%)	*percent of tests where freeze was entered successfully*
fail in wifi_resume	1 (0.0%)	*percent of tests where freeze NOT entered (aborted in wifi_resume)*
pkgpc10	2280 (100.0%)	*percent of tests where PC10 was entered (disabled means PC10 is not supported, hence 0 percent)*
syslpi	2278 (99.9%)	*percent of tests where S0IX mode was entered (disabled means S0IX is not supported, hence 0 percent)*

Backup: suspend stress summary using pm-graph (2/2)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
#	Mode	Host	Kernel	Test Start	Result	Kernel Issues	Suspend	Resume	Worst Suspend Device	ms	Worst Resume Device	ms	PkgPC10	SysLPI
1	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:05:42	pass	S2WAKEx2	356.479	1116.628	0000:00:02.0 {i915} (async)	288.301	hdaudioC0D0 {snd_hda_codec_realtek} (async)	688.800	90.66%	90.57%
2	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:06:20	pass	S2WAKEx1	380.258	1208.434	0000:00:02.0 {i915} (async)	296.192	hdaudioC0D0 {snd_hda_codec_realtek} (async)	693.780	89.72%	89.68%
3	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:06:58	pass	S2WAKEx1	383.827	1116.109	0000:00:02.0 {i915} (async)	292.827	hdaudioC0D0 {snd_hda_codec_realtek} (async)	693.030	90.68%	90.53%
4	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:07:36	pass	S2WAKEx1	380.465	1112.932	0000:00:02.0 {i915} (async)	300.860	hdaudioC0D0 {snd_hda_codec_realtek} (async)	691.756	90.17%	90.04%
5	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:08:13	pass	S2WAKEx1	370.450	1126.403	0000:00:02.0 {i915} (async)	296.100	hdaudioC0D0 {snd_hda_codec_realtek} (async)	690.785	90.04%	90.03%
6	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:08:50	pass	S2WAKEx1	376.345	1120.668	0000:00:02.0 {i915} (async)	289.237	hdaudioC0D0 {snd_hda_codec_realtek} (async)	689.019	90.60%	90.47%
7	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:09:28	pass	S2WAKEx1	370.382	1115.690	0000:00:02.0 {i915} (async)	289.243	hdaudioC0D0 {snd_hda_codec_realtek} (async)	692.352	90.79%	90.70%
8	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:10:06	pass	S2WAKEx1	364.438	1212.776	0000:00:02.0 {i915} (async)	288.282	hdaudioC0D0 {snd_hda_codec_realtek} (async)	692.497	89.92%	89.91%
9	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:10:44	pass	S2WAKEx1	372.092	1116.318	0000:00:02.0 {i915} (async)	292.362	hdaudioC0D0 {snd_hda_codec_realtek} (async)	692.228	90.39%	90.41%
10	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:11:22	pass	S2WAKEx1	389.921	1121.059	0000:00:02.0 {i915} (async)	305.190	hdaudioC0D0 {snd_hda_codec_realtek} (async)	692.700	90.62%	90.50%
11	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:12:00	pass	S2WAKEx1	365.048	1112.350	0000:00:02.0 {i915} (async)	287.846	hdaudioC0D0 {snd_hda_codec_realtek} (async)	689.126	90.71%	90.72%
12	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:12:38	pass	S2WAKEx4	362.534	1114.687	0000:00:02.0 {i915} (async)	288.519	hdaudioC0D0 {snd_hda_codec_realtek} (async)	691.401	90.71%	90.73%
13	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:13:16	pass	S2WAKEx1	379.242	1112.502	0000:00:02.0 {i915} (async)	297.212	hdaudioC0D0 {snd_hda_codec_realtek} (async)	689.316	90.76%	90.64%
14	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:13:54	pass	S2WAKEx1	381.840	1129.781	0000:00:02.0 {i915} (async)	297.043	hdaudioC0D0 {snd_hda_codec_realtek} (async)	690.967	90.66%	90.58%
15	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:14:32	pass	S2WAKEx1	380.520	1110.600	0000:00:02.0 {i915} (async)	296.609	hdaudioC0D0 {snd_hda_codec_realtek} (async)	689.504	90.79%	90.69%
16	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:15:10	pass	S2WAKEx1	373.902	1133.495	0000:00:02.0 {i915} (async)	293.515	hdaudioC0D0 {snd_hda_codec_realtek} (async)	690.060	90.57%	90.48%
17	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:15:48	pass	S2WAKEx2	377.564	1128.312	0000:00:02.0 {i915} (async)	297.299	hdaudioC0D0 {snd_hda_codec_realtek} (async)	690.720	90.60%	90.51%
18	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:16:26	pass	S2WAKEx1	374.605	1190.605	0000:00:02.0 {i915} (async)	292.138	hdaudioC0D0 {snd_hda_codec_realtek} (async)	692.005	90.19%	90.20%
19	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:17:04	pass	S2WAKEx1	374.739	1122.703	0000:00:02.0 {i915} (async)	289.164	hdaudioC0D0 {snd_hda_codec_realtek} (async)	690.980	90.62%	90.48%
20	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:17:42	pass	S2WAKEx1	372.728	1120.494	0000:00:02.0 {i915} (async)	288.747	hdaudioC0D0 {snd_hda_codec_realtek} (async)	691.586	90.59%	90.46%
21	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:18:20	pass	S2WAKEx1	375.121	1114.242	0000:00:02.0 {i915} (async)	289.497	hdaudioC0D0 {snd_hda_codec_realtek} (async)	693.619	90.76%	90.62%
22	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:18:58	pass	S2WAKEx1	378.179	1113.586	0000:00:02.0 {i915} (async)	301.375	hdaudioC0D0 {snd_hda_codec_realtek} (async)	692.688	90.66%	90.53%
23	freeze	otcpl-dell-7386-whl	5.15.0-rc5+	2021/10/15 18:19:36	pass	S2WAKEx2	368.802	1124.547	0000:00:02.0 {i915} (async)	289.209	hdaudioC0D0 {snd_hda_codec_realtek} (async)	689.008	90.50%	90.36%

Backup: typical S0ixSelfTest output (Pass)

```
CPU Core C7 residency after S2idle is: 39.18
GFX RC6 residency after S2idle is: 1342.28
CPU Package C-state 2 residency after S2idle is: 1.04
CPU Package C-state 3 residency after S2idle is: 0.16
CPU Package C-state 8 residency after S2idle is: 0.00
CPU Package C-state 9 residency after S2idle is: 0.00
CPU Package C-state 10 residency after S2idle is: 79.20
S0ix residency after S2idle is: 77.75

S0ix substate after S2idle:
  S0i2.0 S0i3.0

S0ix substate residency after S2idle:
  1745088 10930041

S0ix substates residency delta value: S0i2.0 1745088

S0ix substates residency delta value: S0i3.0 10930041

Congratulations! Your system achieved the deepest S0ix substate!
Here is the S0ix substates status:
Substate   Residency
S0i2.0     1745088
S0i3.0     10930041
```

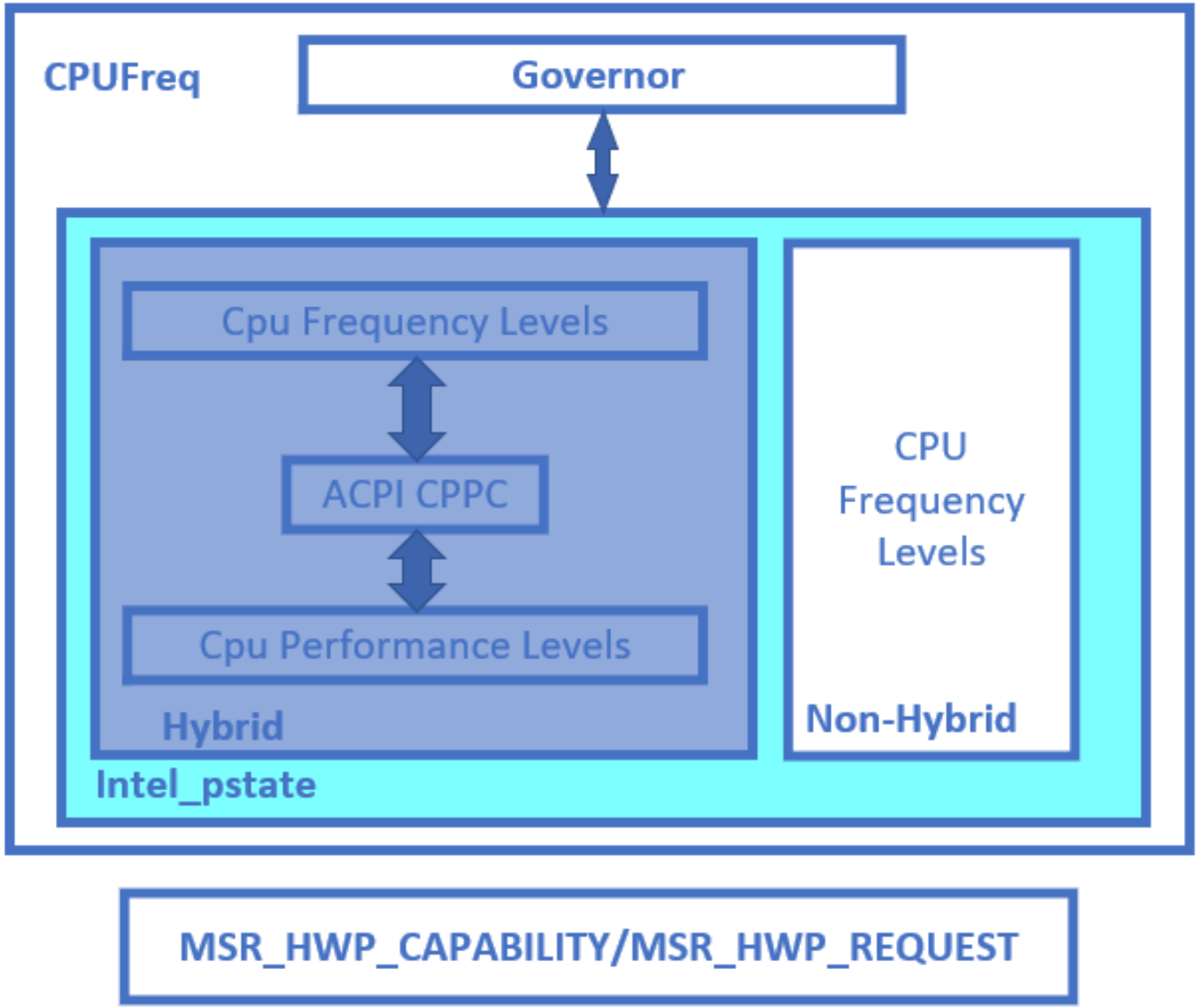
Backup: typical S0ixSelfTest output (Fail)

```
CPU%c1  CPU%c6  CPU%c7  Pkg%pc2 Pkg%pc3 Pkg%pc6 Pkg%pc7 Pkg%pc8 Pkg%pc9 Pk%pc10 SYS%LPI
4.47    32.45   62.81   88.59   0.00    0.00    0.00    0.00    0.00    0.00    0.00
4.51    3.91    91.35   88.59   0.00    0.00    0.00    0.00    0.00    0.00    0.00
4.53
3.96    0.92    95.05
3.98
4.01    0.35    95.61
3.97
4.76    0.26    94.95
4.76
7.32    0.33    92.32
3.71
7.40    0.21    92.34
7.42
3.67    0.11    96.17
3.69
3.60    0.38    95.97
3.63
3.61    96.35   0.00
3.53    96.44   0.00
3.50    93.91   0.00
3.78    96.18   0.00

Your CPU Core C7 residency is available: 62.81

^[[31mPlease check if Intel graphic i915 driver is not loaded
or Intel graphics controller has been disabled
or the 3rd party graphics device is installed.^[[0m
```

Backup: CPU Freq on Hybrid platform




intel