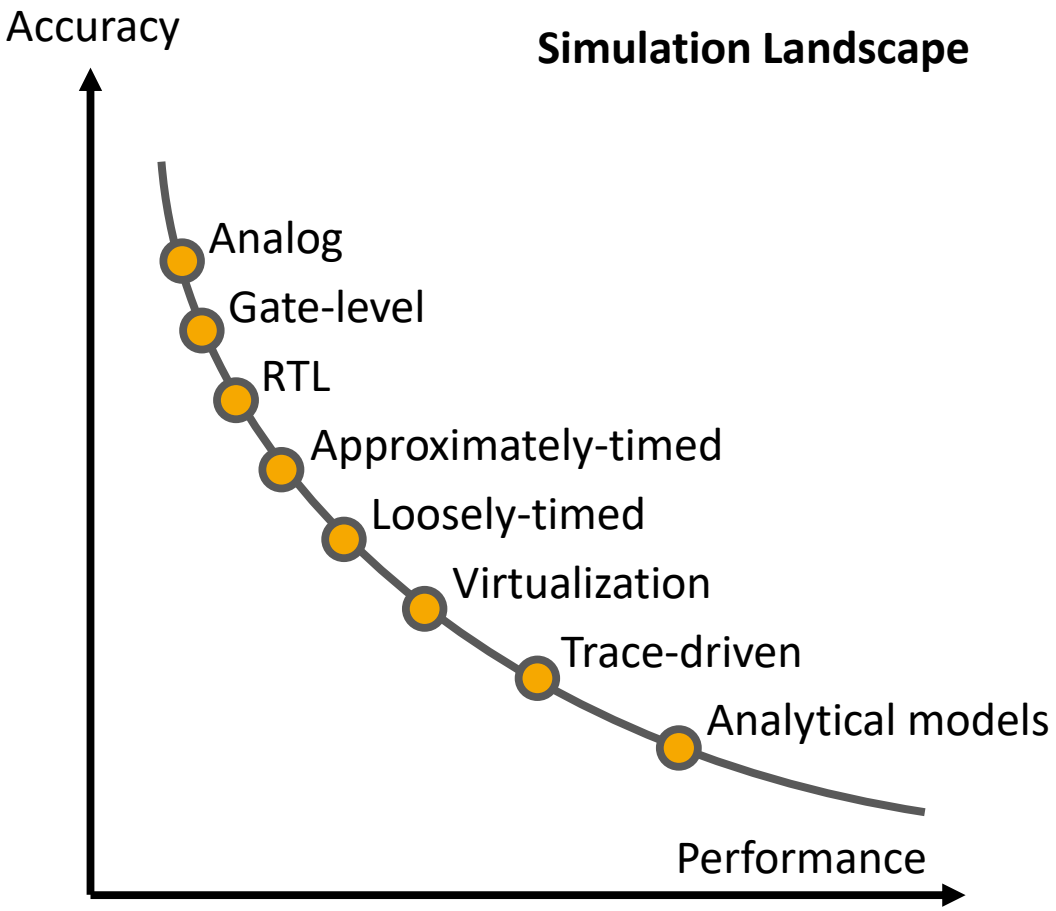
The Optimal Quantum of Temporal Decoupling

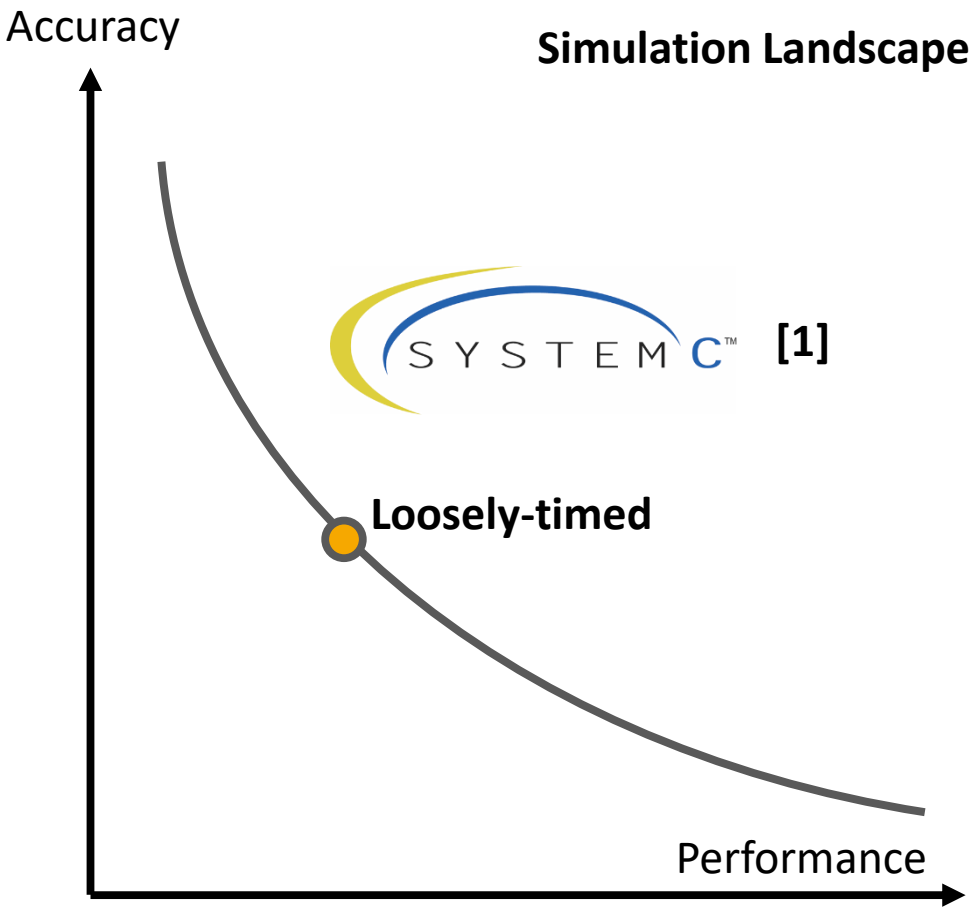
Niko Zurstraßen, Ruben Brandhofer, José Cubero-Cascante,
Lukas Jünger, Nils Bosbach, Rainer Leupers

Institute for Communication Technologies and Embedded Systems
RWTH AACHEN UNIVERSITY

Background: Simulations



Background: Simulations



Performance
Near-native (1-10 GIPS)

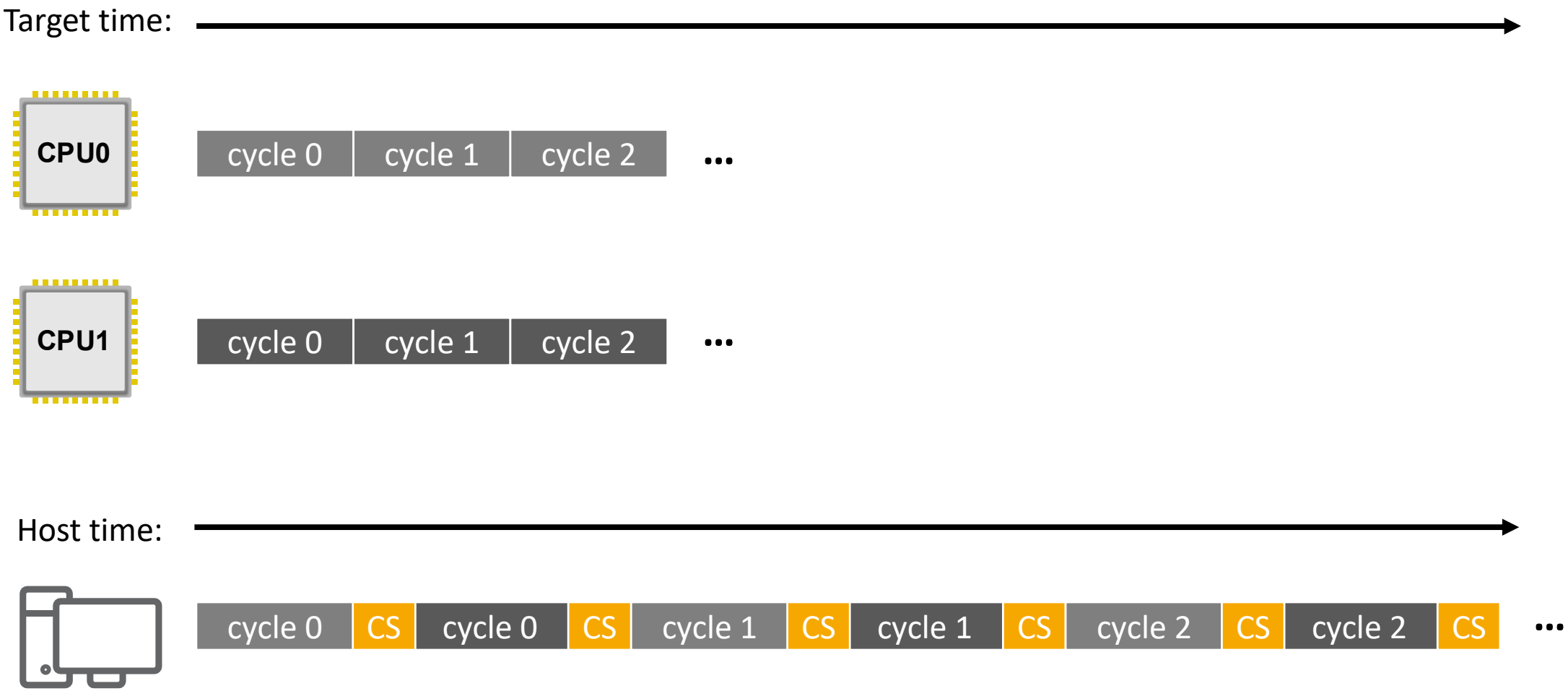
Accuracy
Instruction-accurate
1 instruction ≙ 1 cycle

Use for
Early SW development, SW analysis, ...

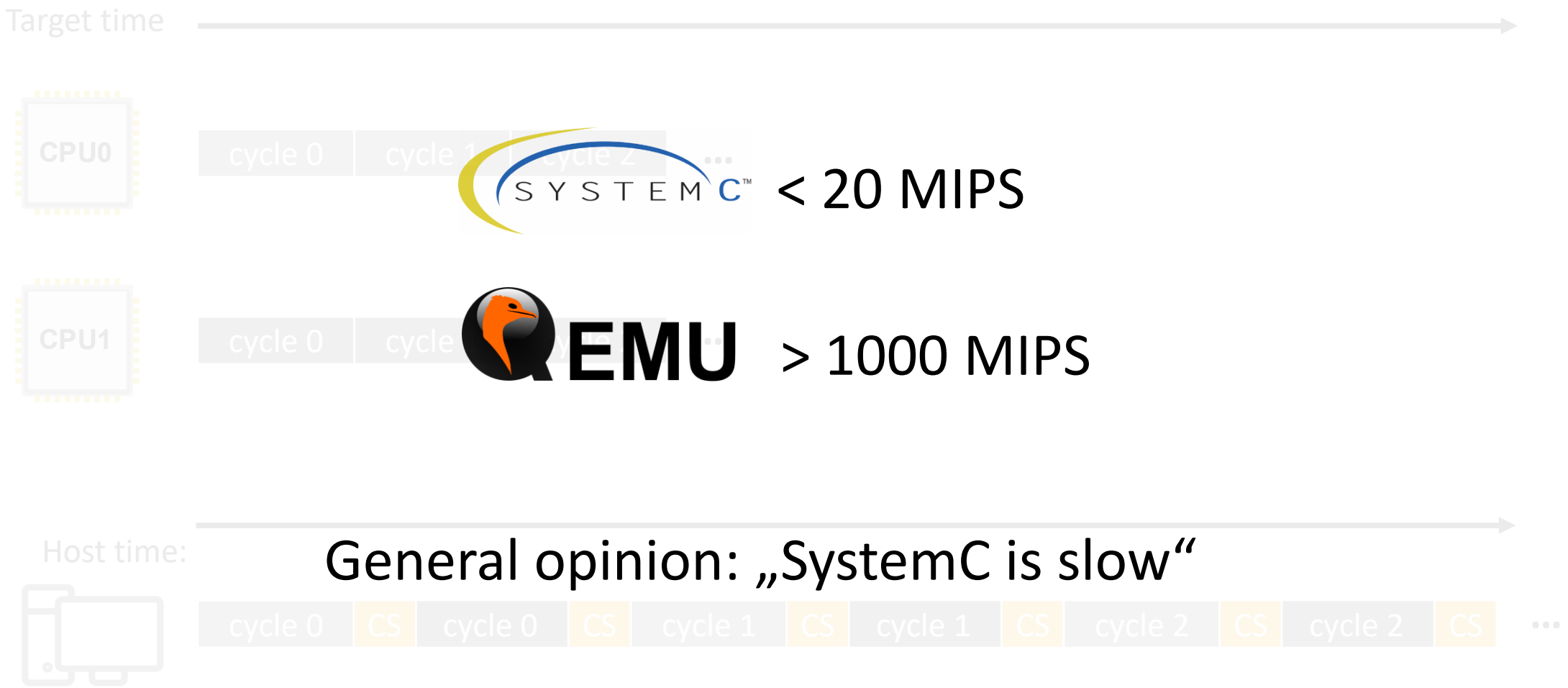
Do not use for
Design space exploration, power estimation, …

[1] "IEEE Standard for Standard SystemC Language Reference Manual," in *IEEE Std 1666-2011 (Revision of IEEE Std 1666-2005)*

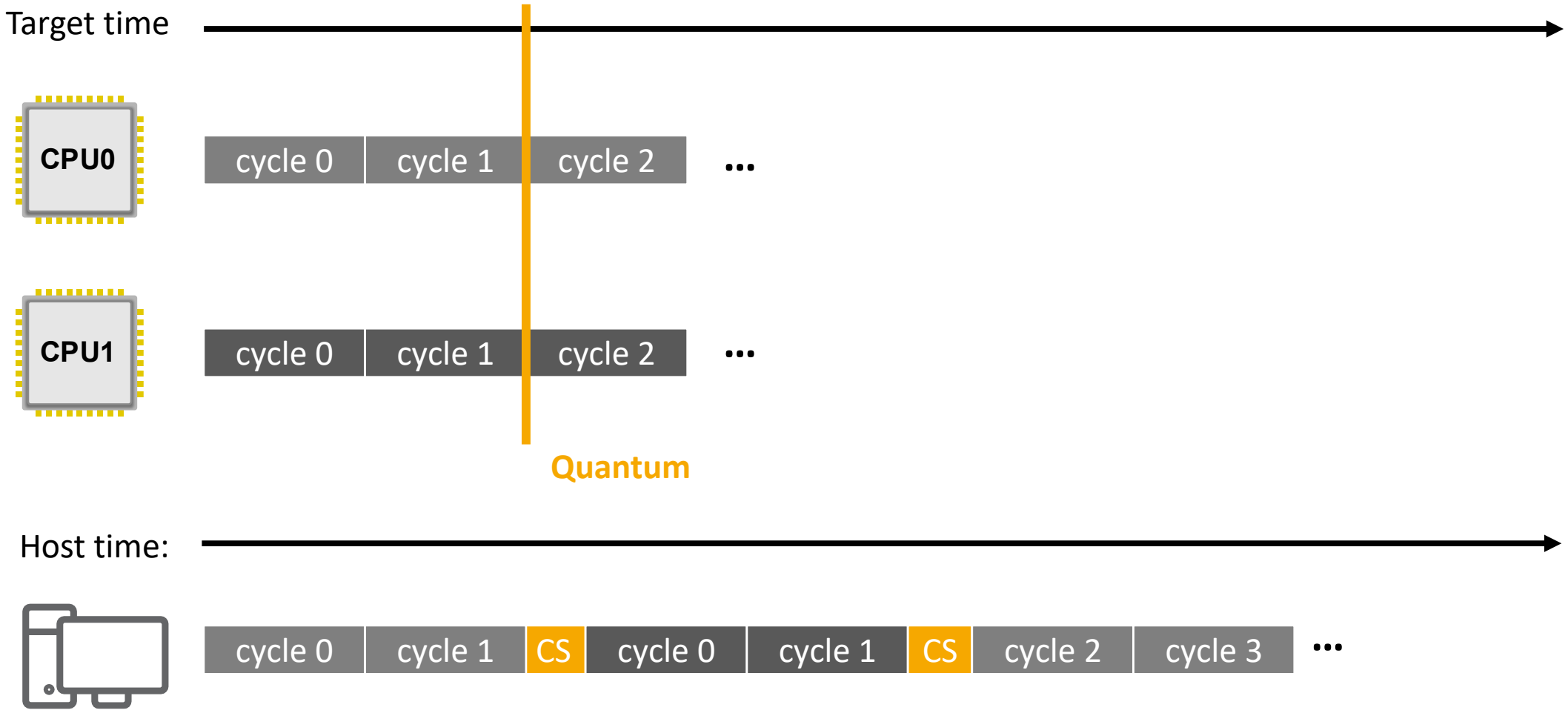
Background: Traditional SystemC Simulation



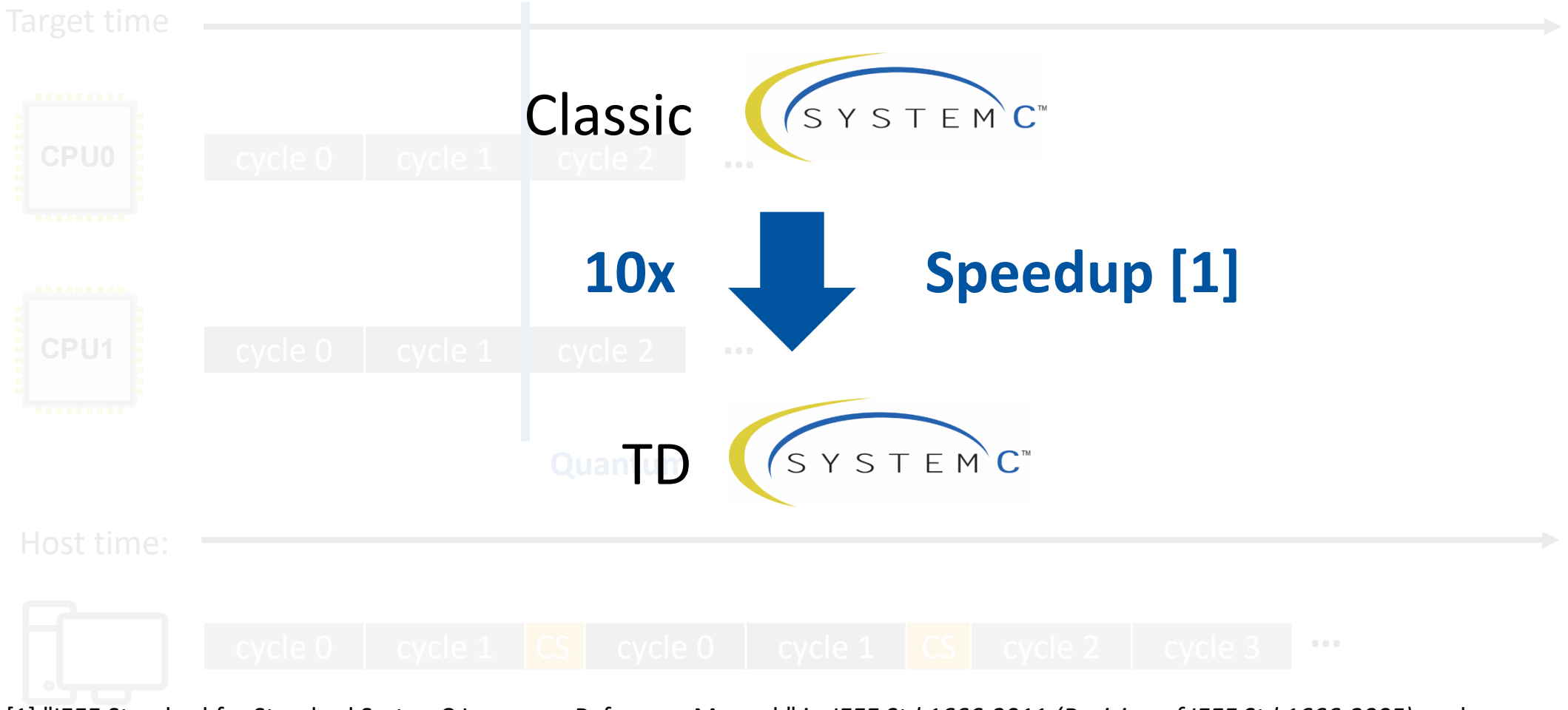
Background: Traditional SystemC Simulation



Background: Temporally-Decoupled Simulations

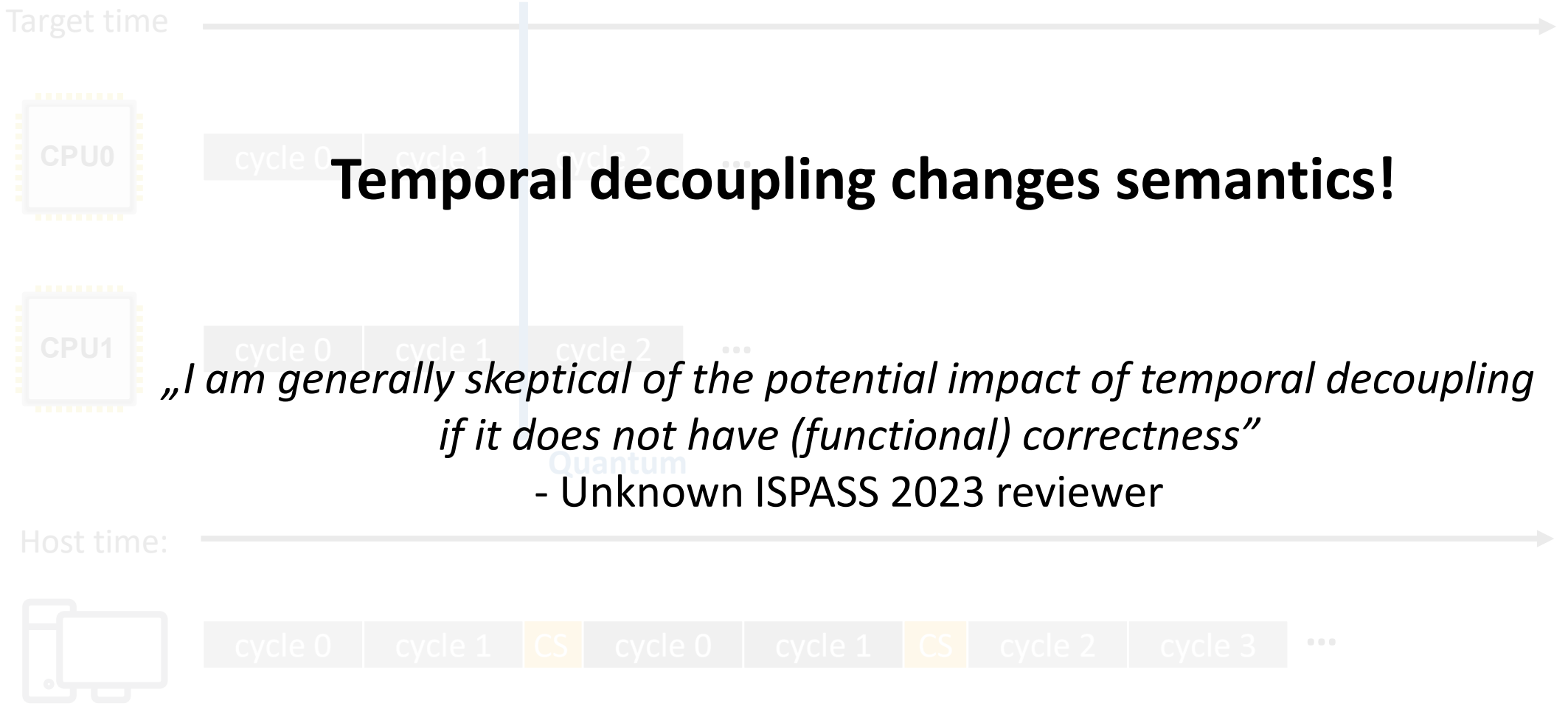


Background: Temporally-Decoupled Simulations



[1] "IEEE Standard for Standard SystemC Language Reference Manual," in *IEEE Std 1666-2011 (Revision of IEEE Std 1666-2005)* , vol., no., pp.1-638, 9 Jan. 2012, doi: 10.1109/IEEESTD.2012.6134619.

Background: Temporally-Decoupled Simulations



Temporal decoupling changes semantics!

„I am generally skeptical of the potential impact of temporal decoupling if it does not have (functional) correctness”

- Unknown ISPASS 2023 reviewer

The Story

The Story: Related Work (?)

"[…] Most of the time, software functionality and correctness are unaffected by temporal decoupling, and **the default should be to use long time quanta**"

- J. Engblom, 2018, Temporal Decoupling – Are "Fast" and "Correct" Mutually Exclusive?, 2018 Design and Verification Conference Europe (DVCon)

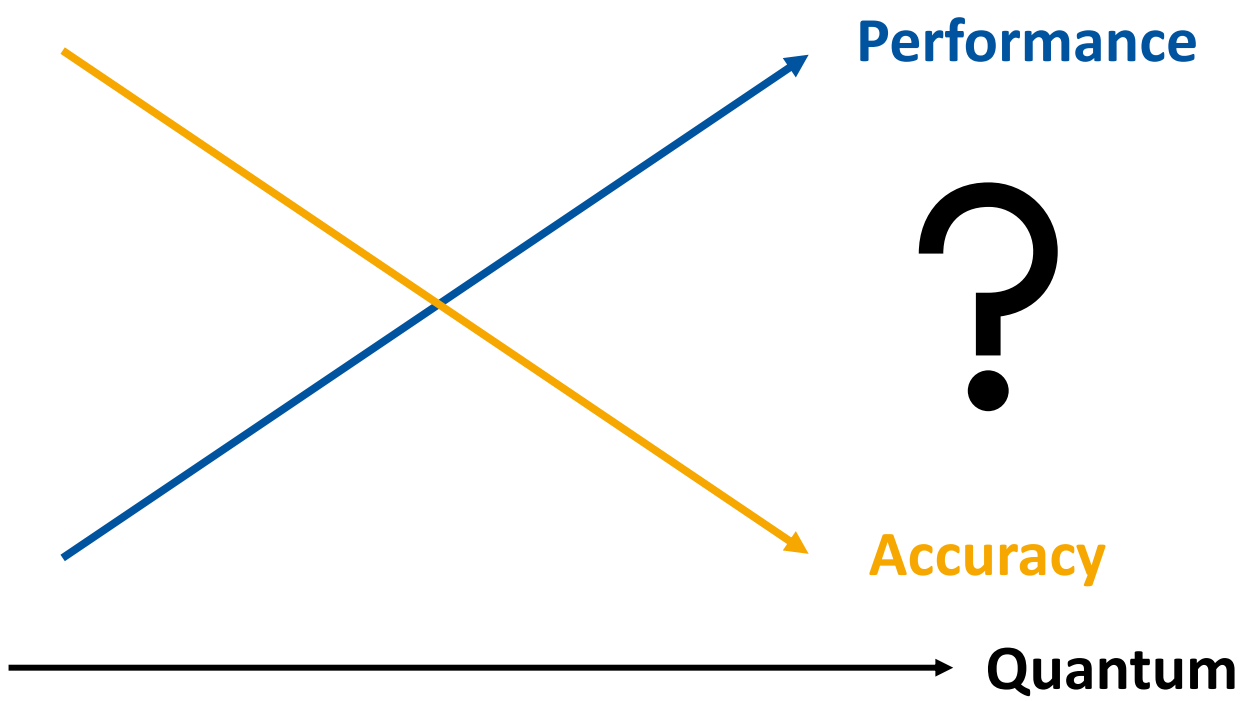
"Increasing the quantum can cause a thread to run for a longer time, thus reducing the context switching overhead. **This increases the simulation speed, but at the cost of accuracy**."

- J. Joy, 2020, Evaluating Temporal Decoupling in a Virtual Platform, Thesis

"To increase performance, **the quantum should be as large as possible** to reduce context switching. **However, a large quantum reduces simulation accuracy,** as events may be handled too late. Therefore, deploying **temporal decoupling is not trivial**."

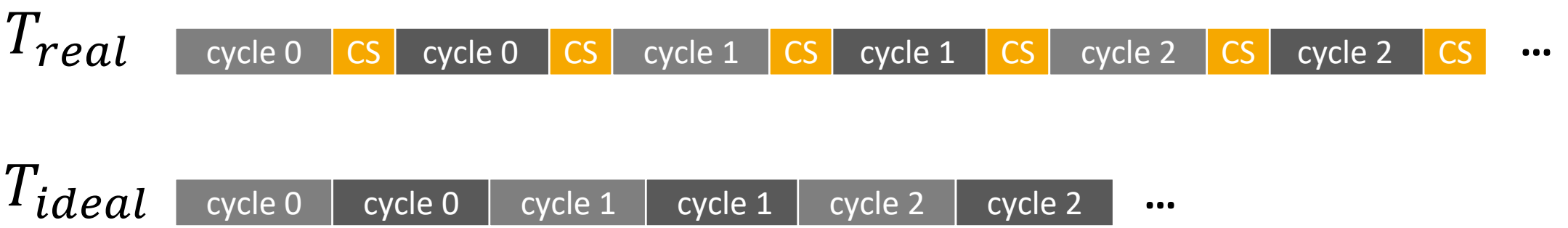
- Jünger et al., 2021, Optimizing Temporal Decoupling using Event Relevance, 2021 26th Asia and South Pacific Design Automation Conference (ASP-DAC)

The Story: Speedup vs. Accuracy



Methods

Methods: A Performance Model

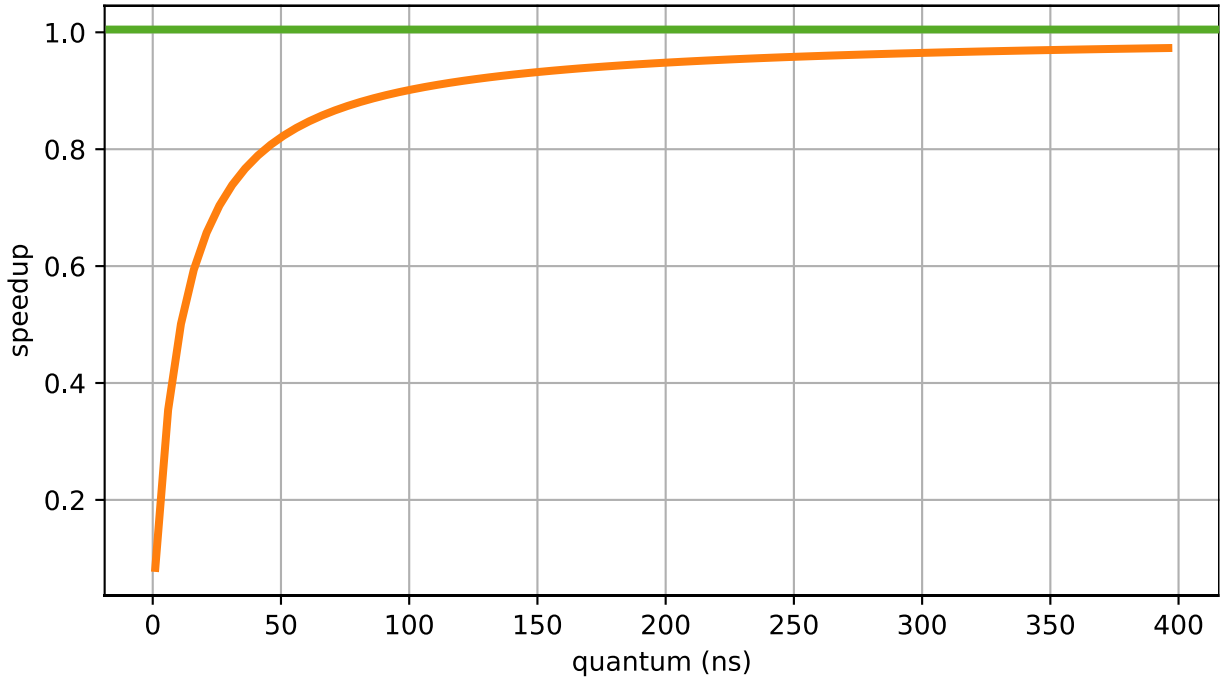


$$Speedup(t_{\Delta q}) = \frac{T_{ideal}}{T_{real}} = \cdots = \frac{t_{\Delta q}}{t_{\Delta q} + O_c{}'}$$

$t_{\Delta q} = \text{quantum}$ $O_c' = \text{overhead factor}$

Methods: A Performance Model

$$Speedup(t_{\Delta q}) = \frac{T_{ideal}}{T_{real}} = \cdots = \frac{t_{\Delta q}}{t_{\Delta q} + O_c{}'}$$



$$O_c' = 11ns$$

Methods: An Accuracy Model?

What is (in)accuracy?

Qualitative

Example: Program execution

Default: Success

```
niko@laptop ~> ./app
success
```

With TD: Segmentation Fault

```
niko@laptop ~> ./app
terminated by signal SIGSEGV
```

Quantitative

Example: Target time

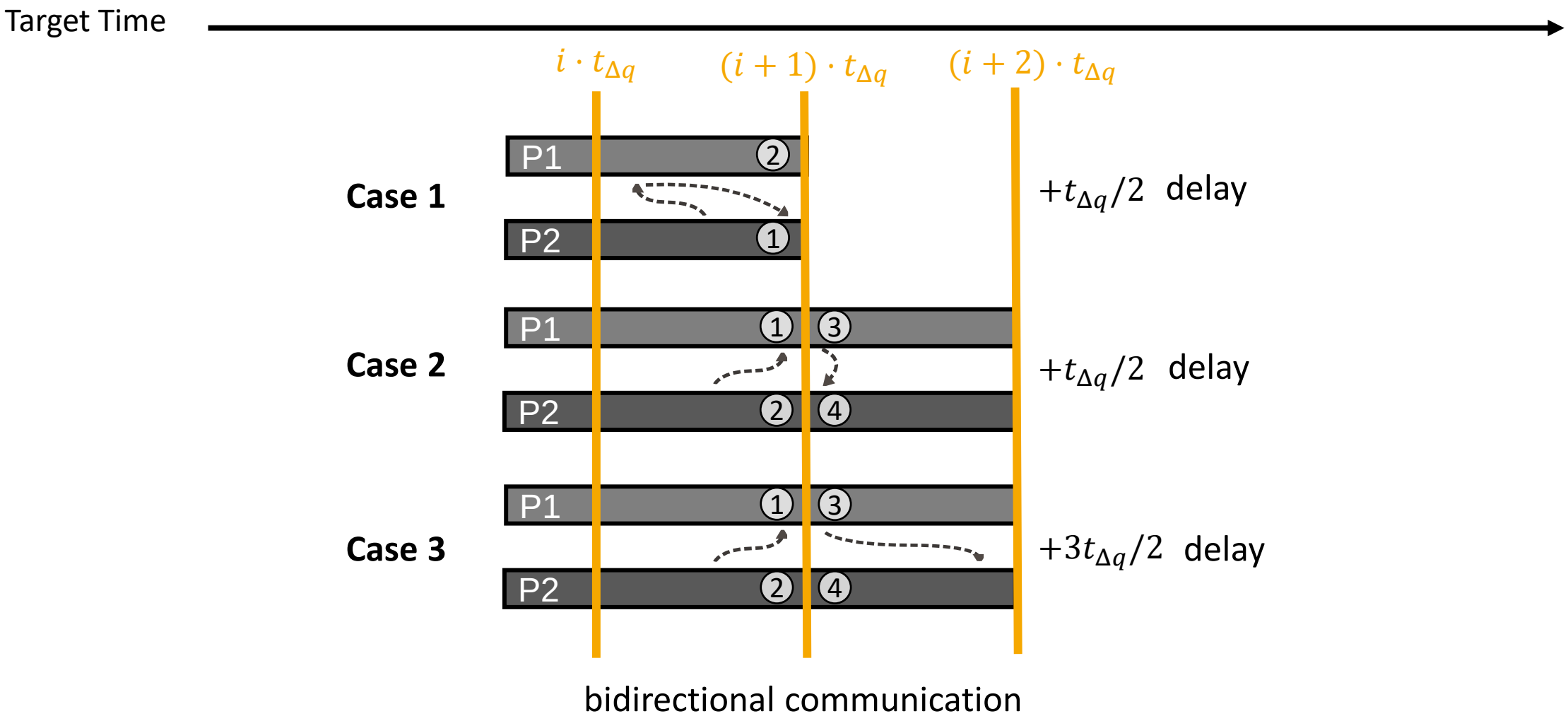
Default:

1s

With TD:

2s

Methods: Quantitative Accuracy



Methods: Quantitative Accuracy

$$t_d = t_{\Delta q} - E\left(X \middle| X \leq t_{\Delta q}\right) P\left(X < t_{\Delta q}\right) - t_{\Delta q} P(X > t_{\Delta q})$$

$$= t_{\Delta q} \int_0^{t_{\Delta q}} rte^{-rt}\, dt - \int_{t_{\Delta q}}^{\infty} rt_{\Delta q} e^{-rt} dt$$

$$= t_{\Delta q} - \frac{\left(1 - e^{-rt\Delta q}\right)}{r}$$

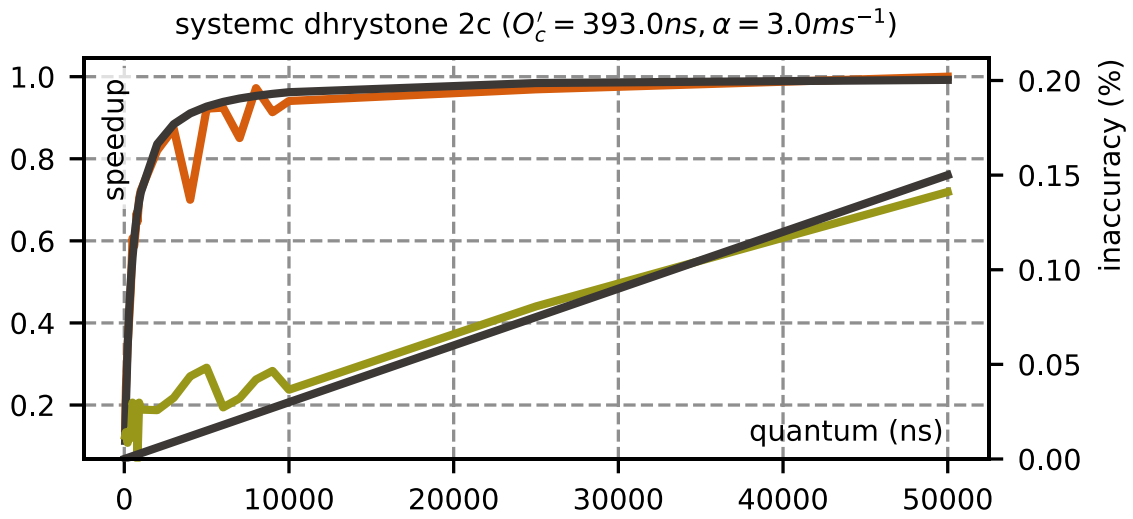
$$I = \frac{t_{t\Delta q}}{t_{t\Delta q} - t_d} - 1 = r \cdot \frac{t_{\Delta q}}{1 - e^{-rt_{\Delta q}}} - 1 \approx r \cdot t_{\Delta q}$$

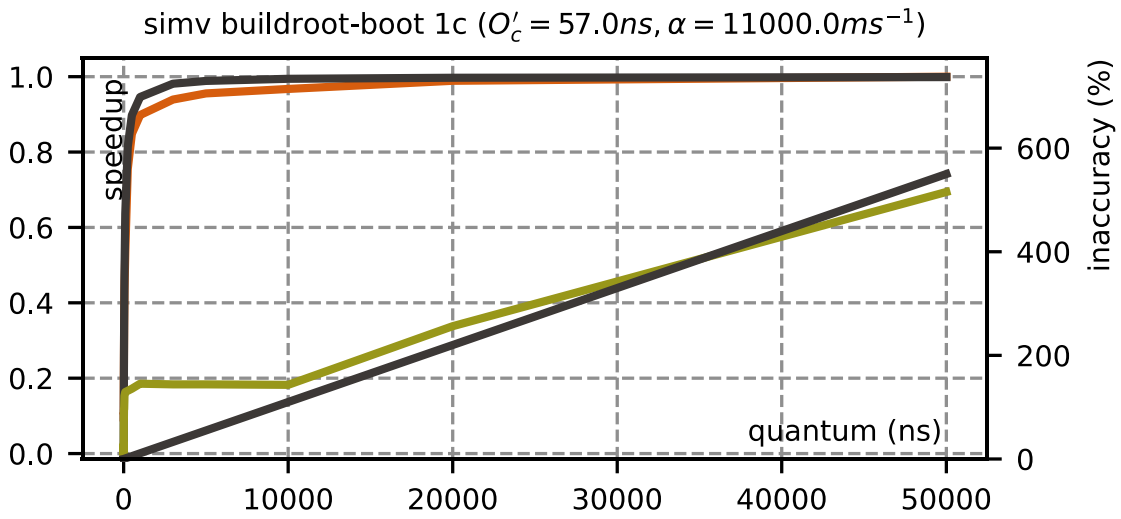
$$\boldsymbol{Inaccuracy(t_{\Delta q}) = \alpha \cdot t_{\Delta q}}$$

Timing inaccuracy grows linearly with the quantum ($t_{\Delta q}$)!

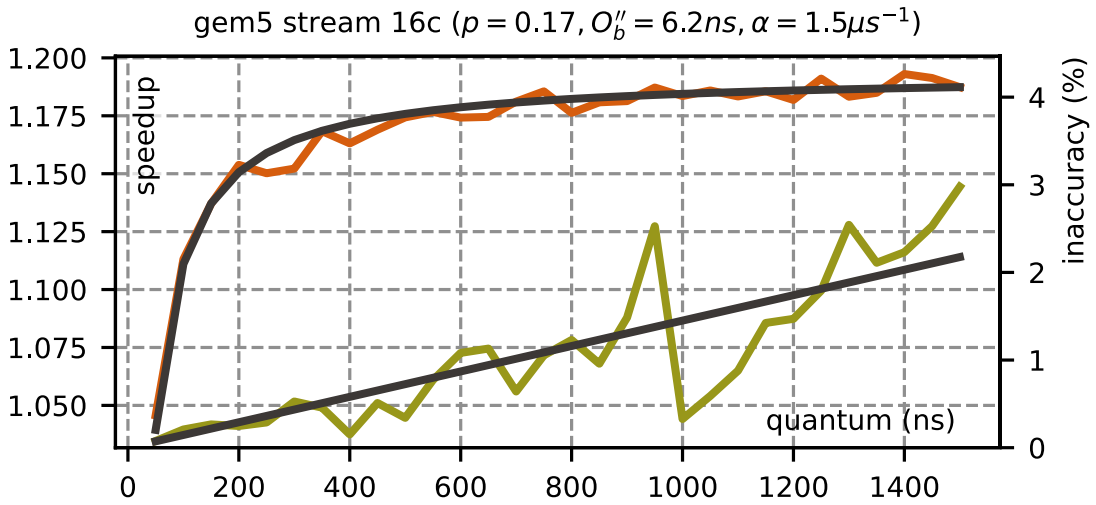
Results

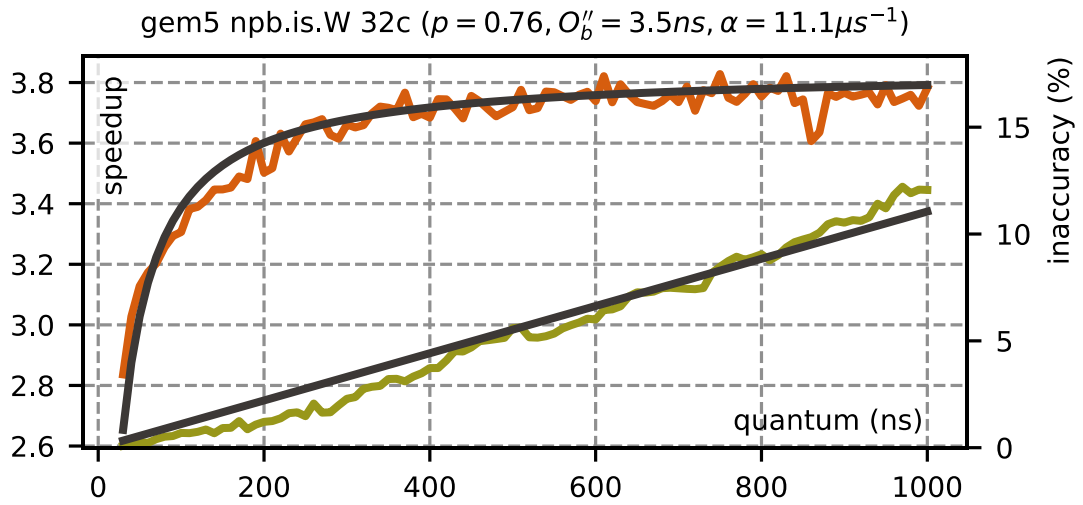
Results: SystemC Speedup vs. Accuracy



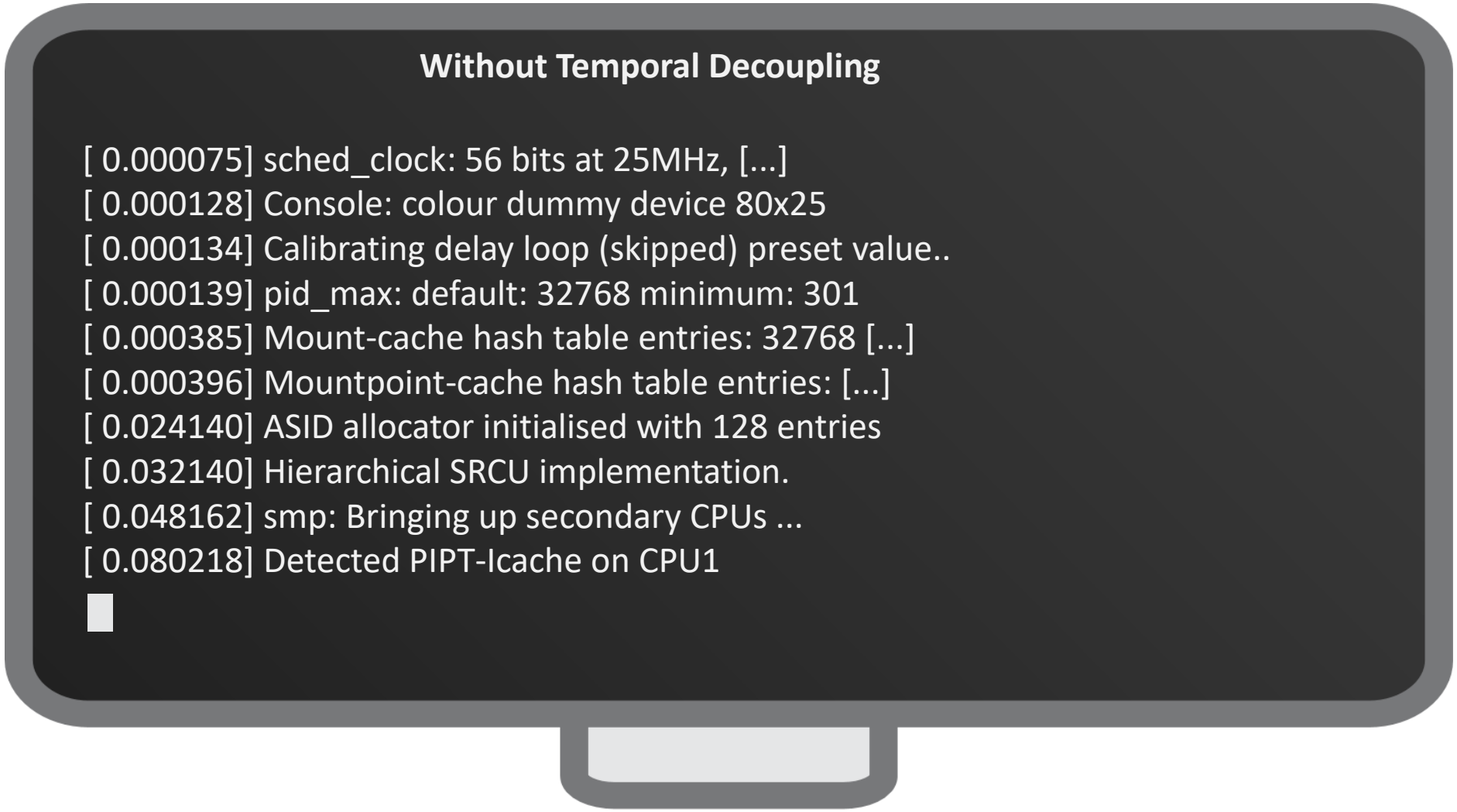


Results: gem5 Speedup vs. Accuracy

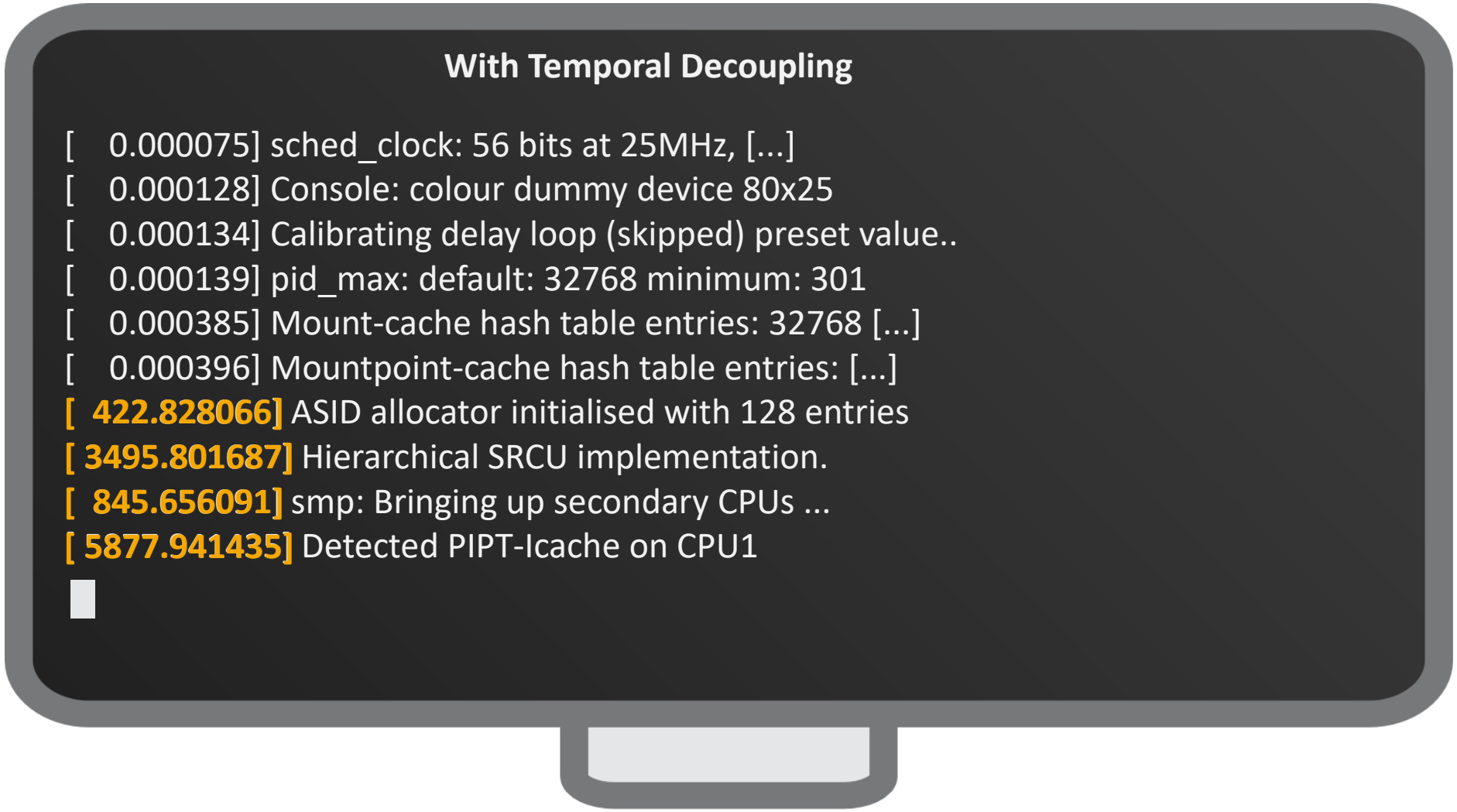




Results: Qualitative Accuracy



Results: Qualitative Accuracy



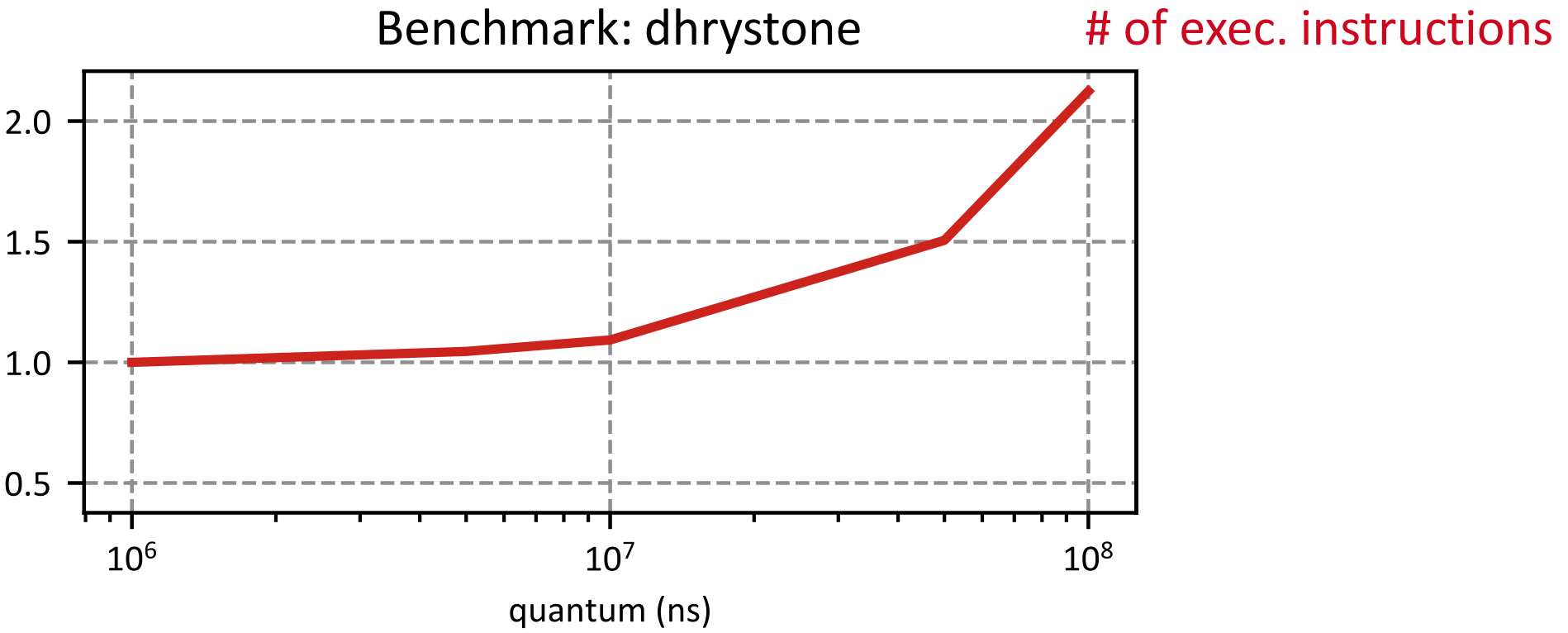
Results: Qualitative Accuracy

What happened to the linux boot timestamps?

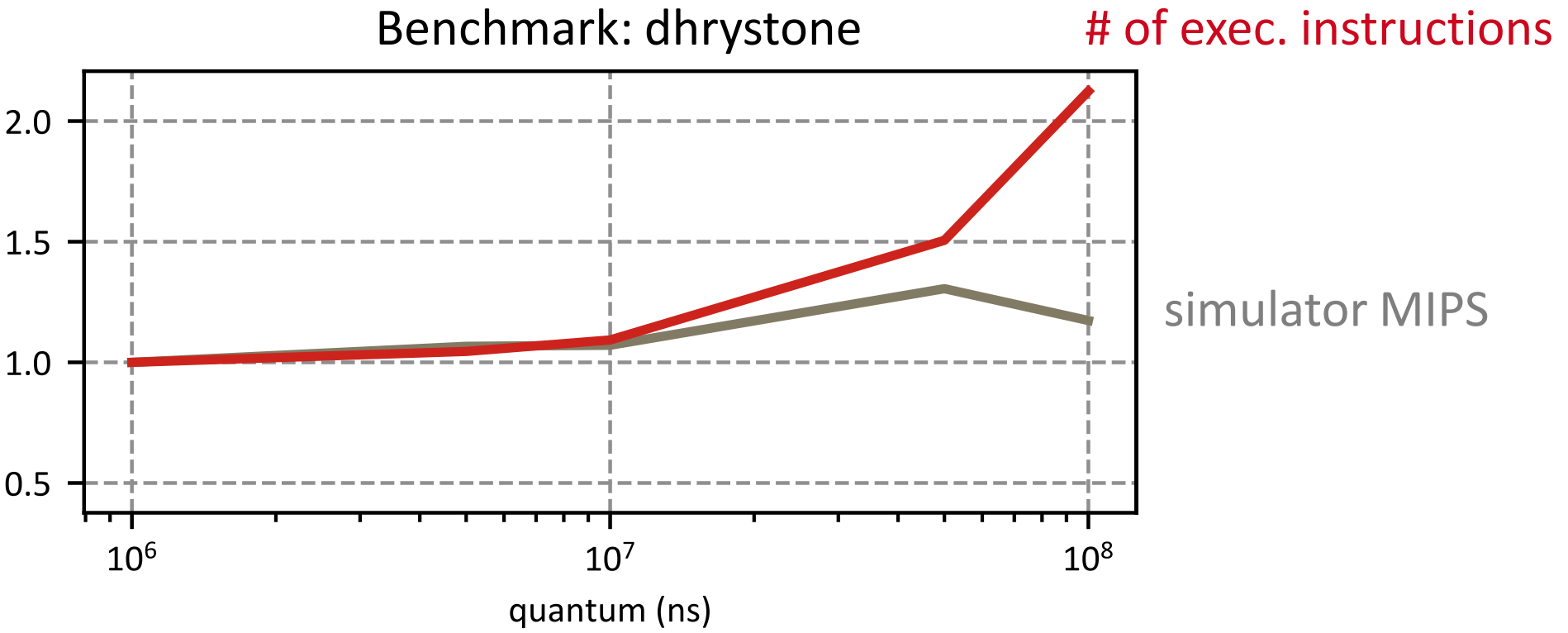

```
// From generic_timer.cc:
uint64_t update = current_time – last_time
```

update = 100 – 200 = 18,446,744,073,709,551,516

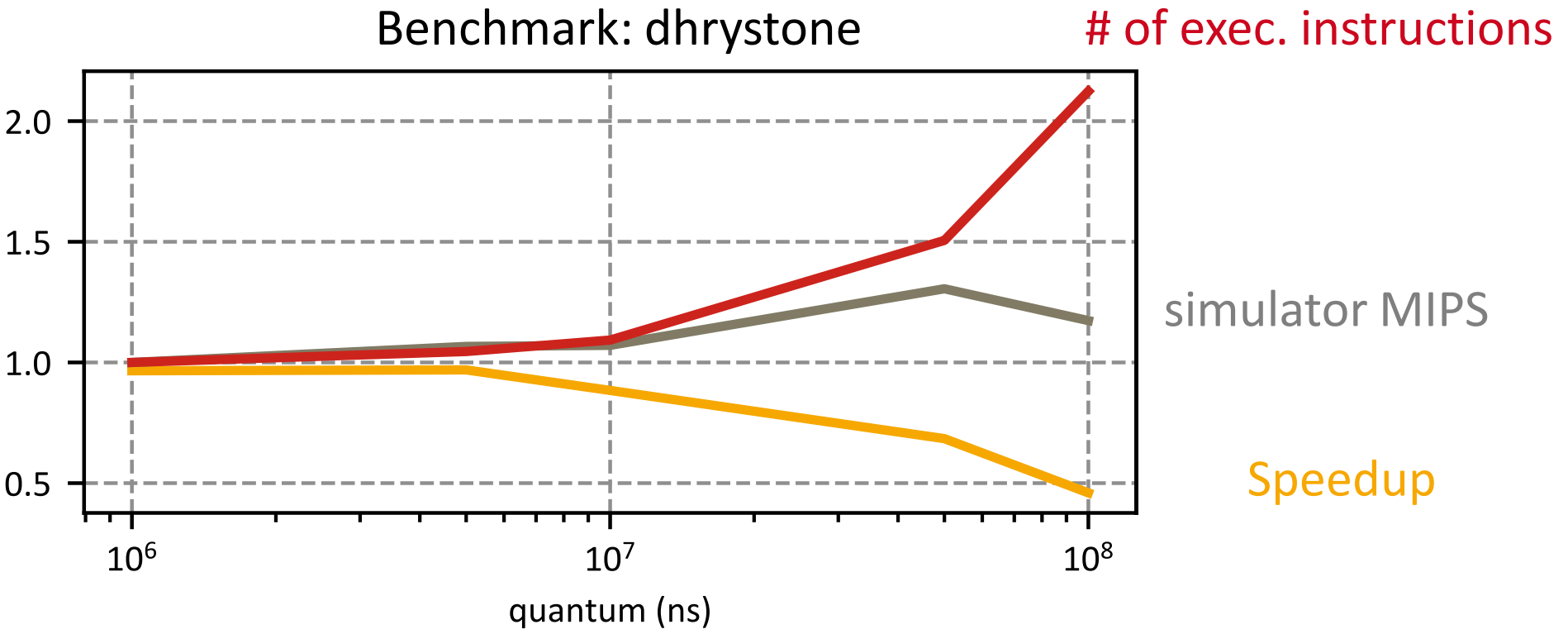
Results: Other Interesting Effects



Results: Other Interesting Effects



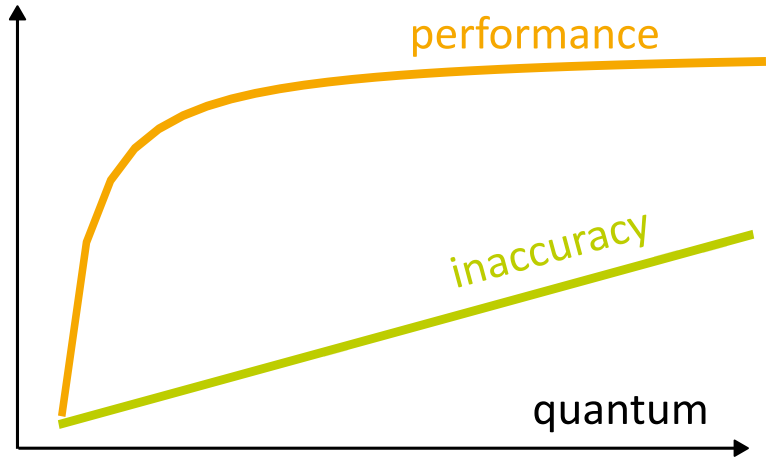
Results: Other Interesting Effects



Conclusion

Conclusion

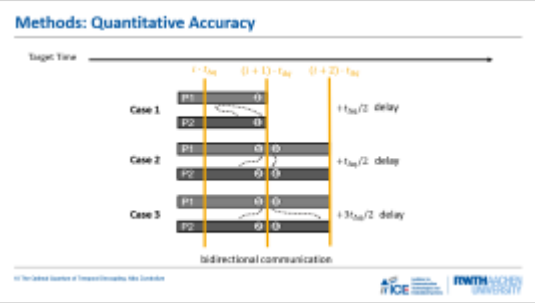
Contribution: An analytical model of temporal decoupling

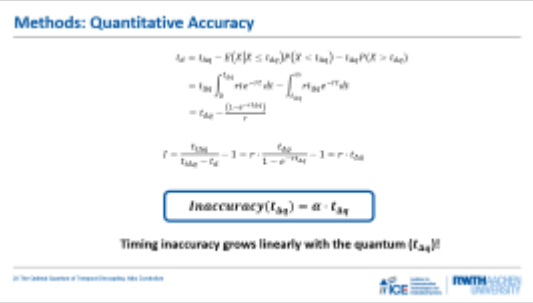


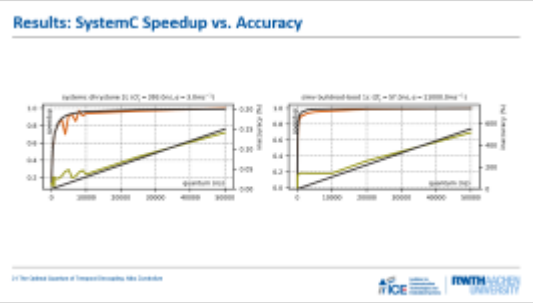
“All models are wrong, but some are useful.”
- George Box, 1976

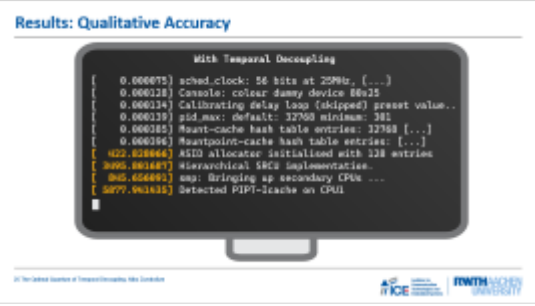
Conclusion

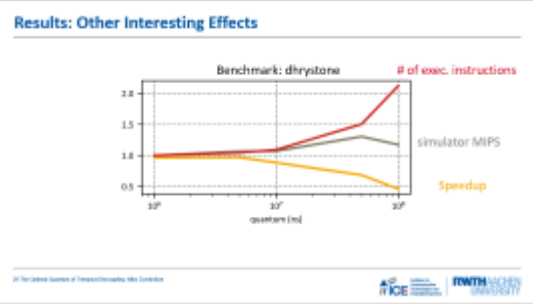
Thanks for listening!
Questions?

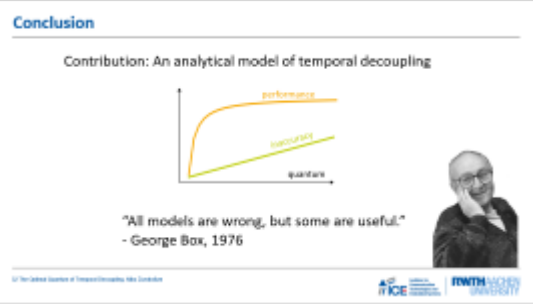












Institute for Communication Technologies and Embedded Systems | RWTH AACHEN UNIVERSITY