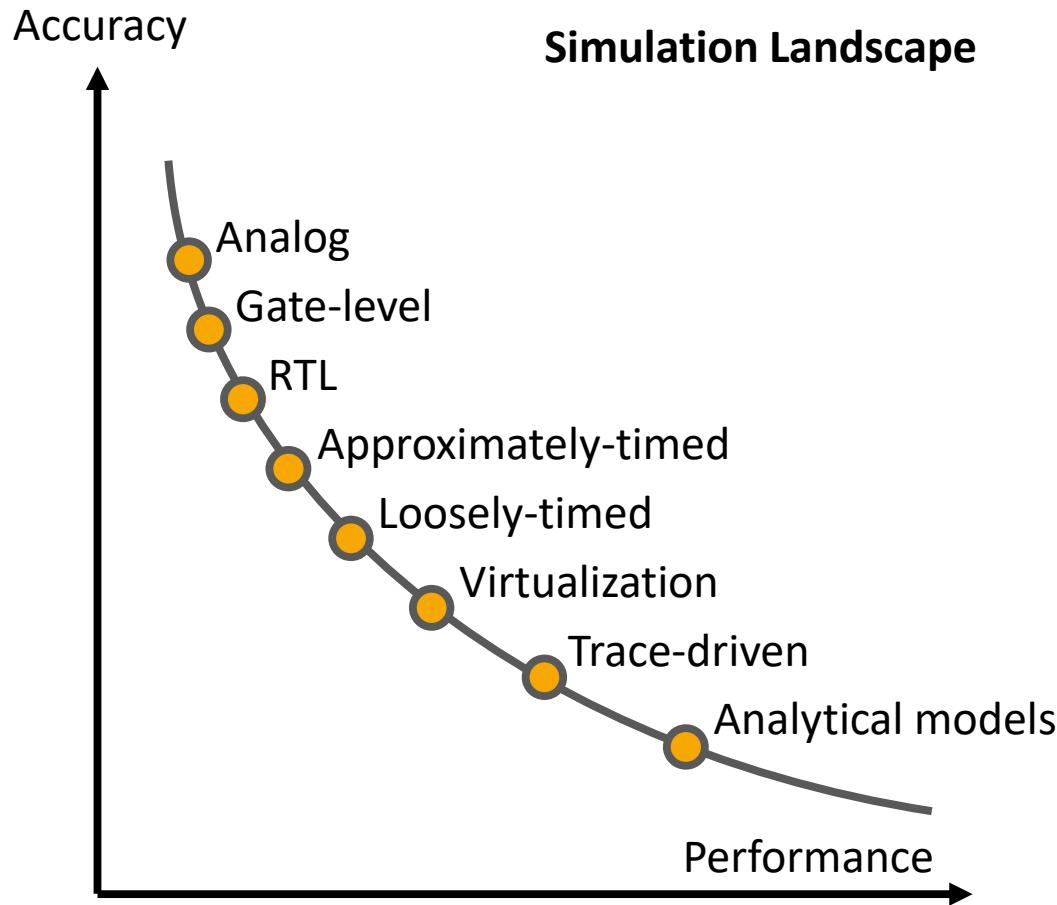




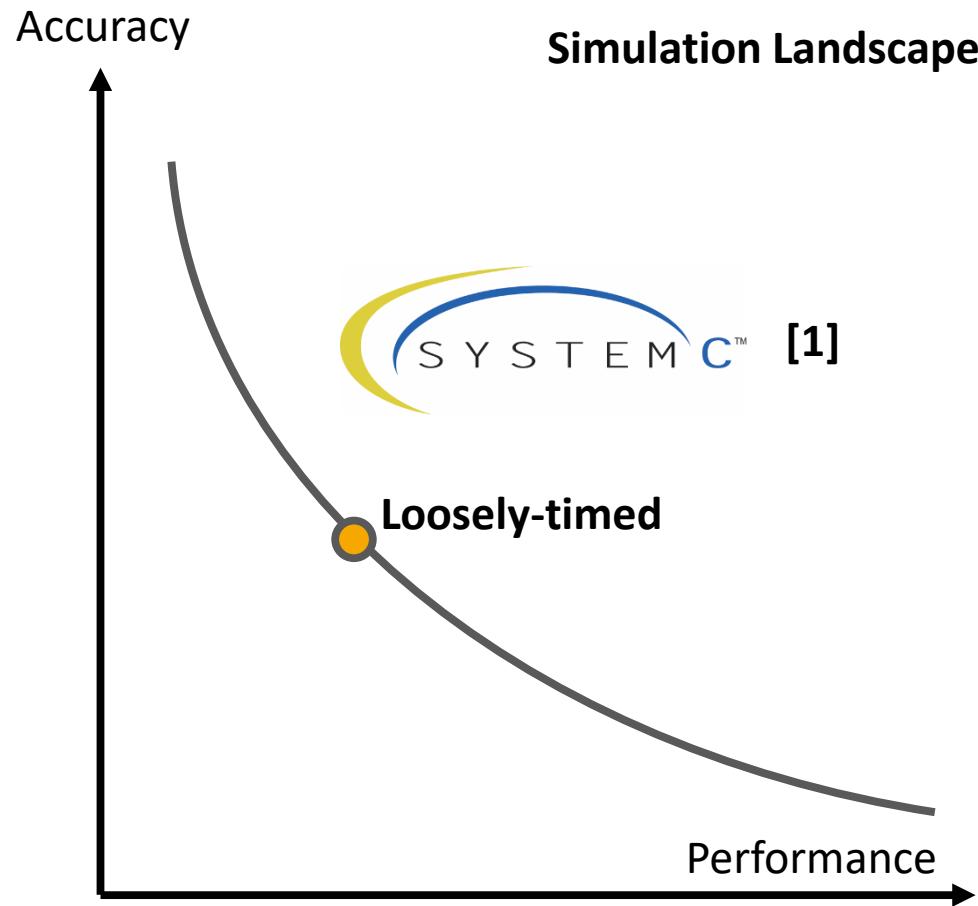
The Optimal Quantum of Temporal Decoupling

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

Background: Simulations



Background: Simulations



Performance

Near-native (1-10 GIPS)

Accuracy

Instruction-accurate
 $1 \text{ instruction} \leqq 1 \text{ cycle}$

Use for

Early SW development, SW analysis, ...

Do not use for

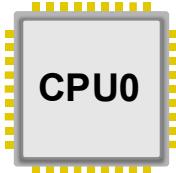
Design space exploration, power estimation, ...

3 The Optimal Quantum of Temporal Decoupling, Niko Zurstraßen

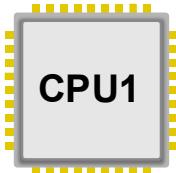
[1] "IEEE Standard for Standard SystemC Language Reference Manual," in *IEEE Std 1666-2011 (Revision of IEEE Std 1666-2005)*

Background: Traditional SystemC Simulation

Target time: 



cycle 0 | cycle 1 | cycle 2 ...



cycle 0 | cycle 1 | cycle 2 ...

Host time: 



cycle 0 | CS | cycle 0 | CS | cycle 1 | CS | cycle 1 | CS | cycle 2 | CS | cycle 2 | CS ...

Background: Traditional SystemC Simulation

Target time →



< 20 MIPS



> 1000 MIPS

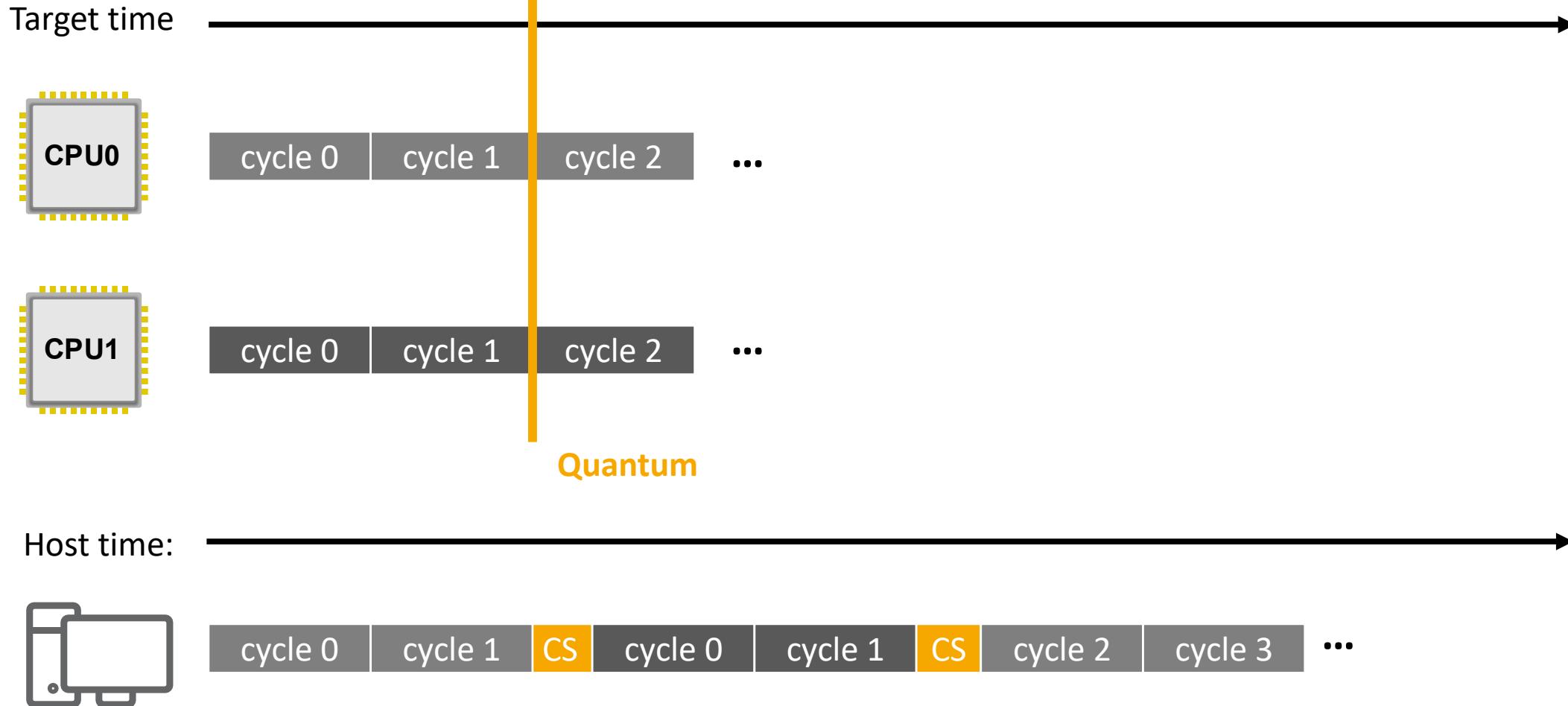
Host time: →



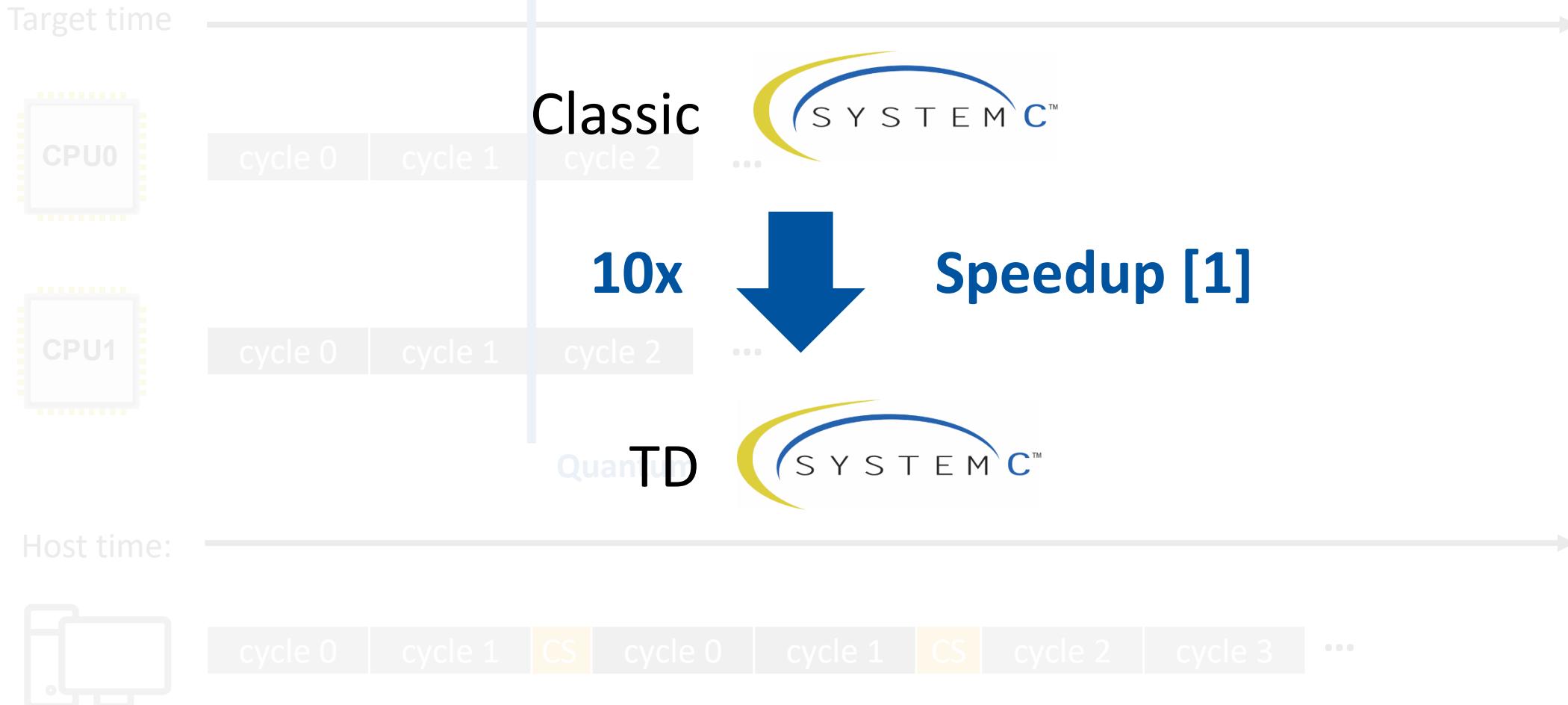
General opinion: „SystemC is slow“



Background: Temporally-Decoupled Simulations



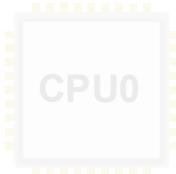
Background: Temporally-Decoupled Simulations



[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.

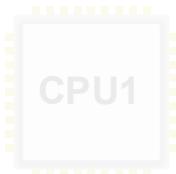
Background: Temporally-Decoupled Simulations

Target time



cycle 0 | cycle 1 | cycle 2 | ...

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

Host time:



cycle 0 | cycle 1 | cs | cycle 0 | cycle 1 | cs | cycle 2 | cycle 3 | ...

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

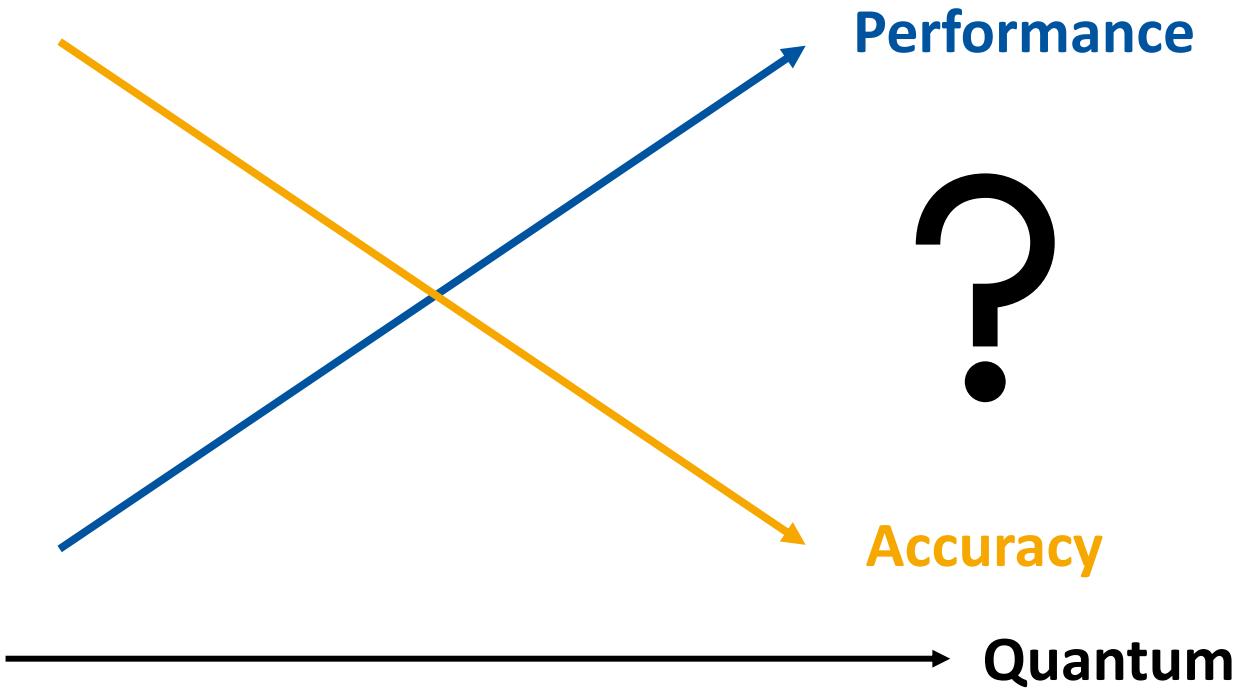
“To increase performance, **the quantum should be as large as possible** to reduce context switching.

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

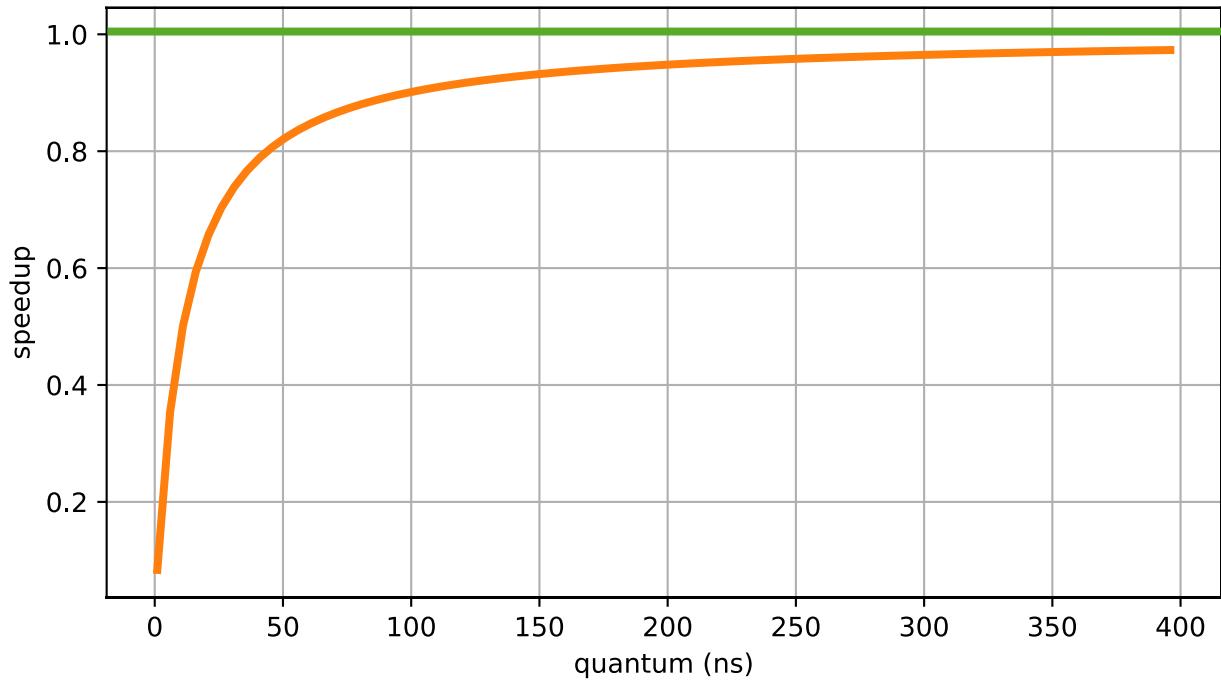


$$Speedup(t_{\Delta q}) = \frac{T_{ideal}}{T_{real}} = \dots = \frac{t_{\Delta q}}{t_{\Delta q} + O_c'}$$

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

Methods: A Performance Model

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



$$O_c' = 11\text{ns}$$

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
```

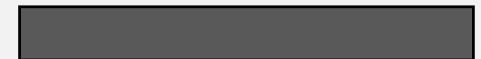
Quantitative

Example: Target time

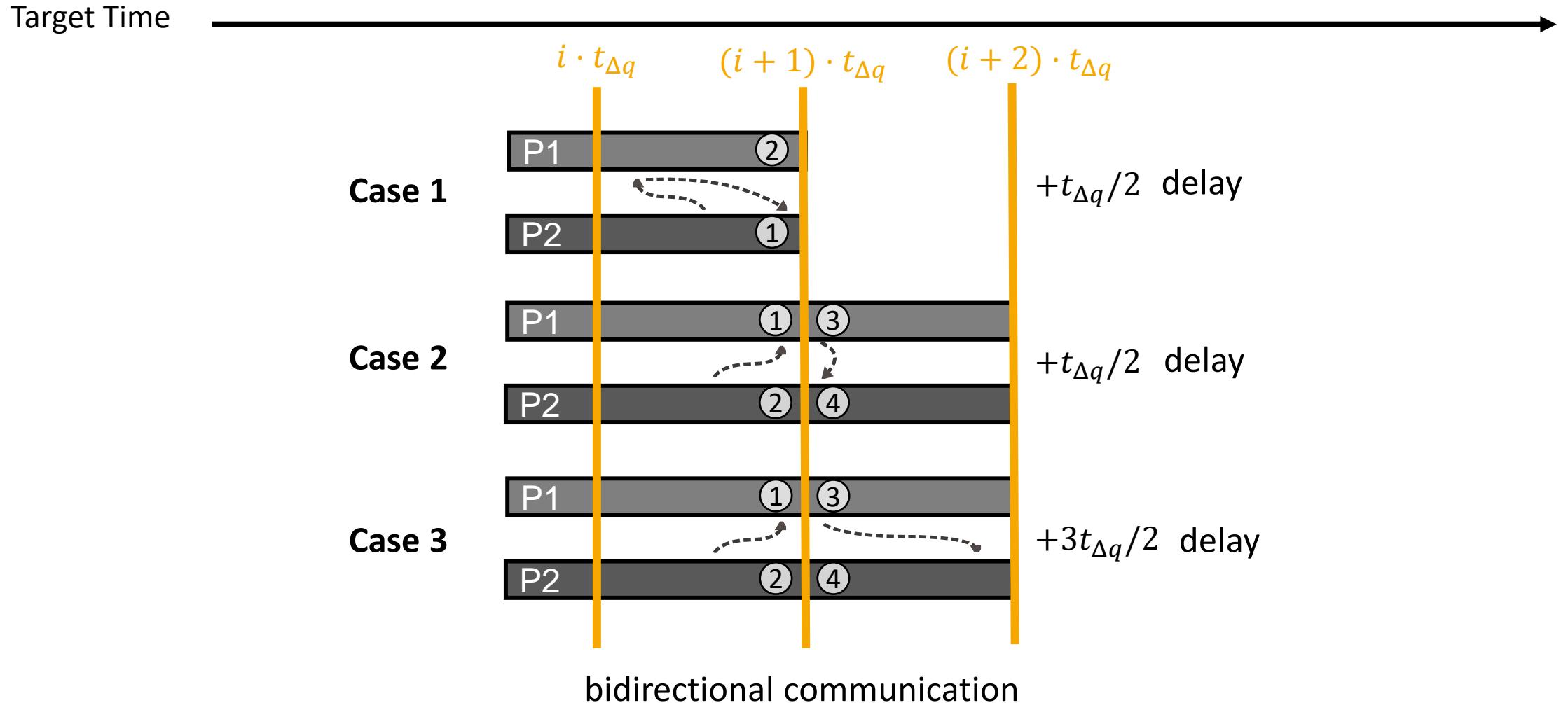
Default:

 1s

With TD:

 2s

Methods: Quantitative Accuracy



Methods: Quantitative Accuracy

$$\begin{aligned} t_d &= t_{\Delta q} - E(X|X \leq t_{\Delta q})P(X < t_{\Delta q}) - t_{\Delta q}P(X > t_{\Delta q}) \\ &= t_{\Delta q} \int_0^{t_{\Delta q}} rte^{-rt} dt - \int_{t_{\Delta q}}^{\infty} rt_{\Delta q}e^{-rt} dt \\ &= t_{\Delta q} - \frac{(1-e^{-rt_{\Delta q}})}{r} \end{aligned}$$

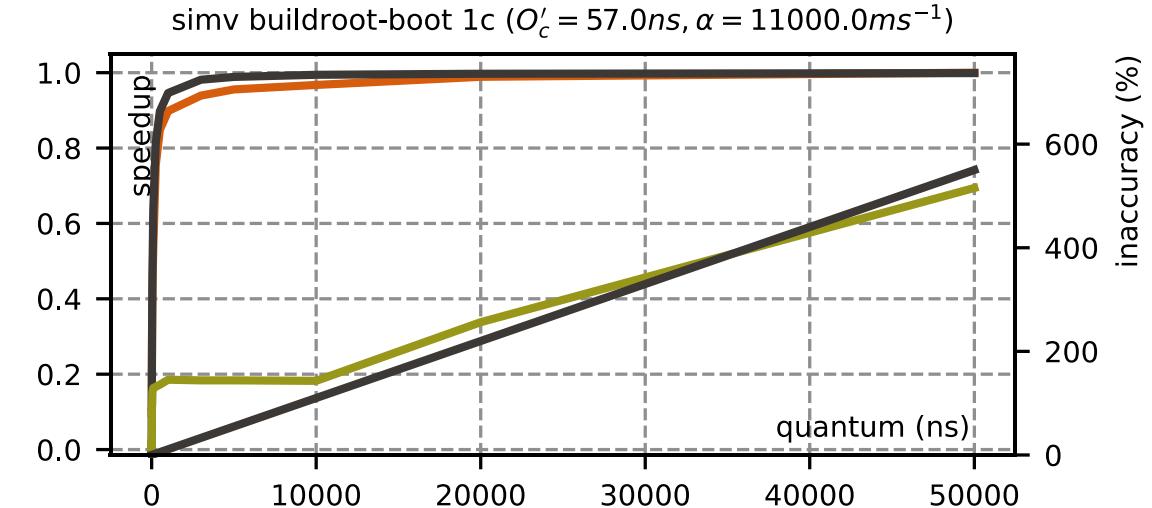
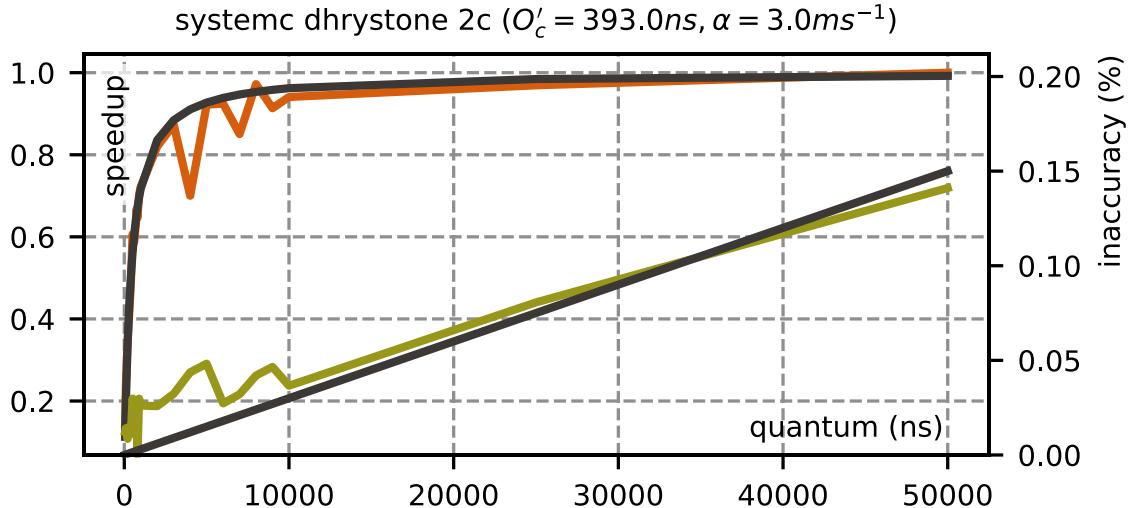
$$I = \frac{t_{t\Delta q}}{t_{t\Delta q} - t_d} - 1 = r \cdot \frac{t_{\Delta q}}{1 - e^{-rt_{\Delta q}}} - 1 \approx r \cdot t_{\Delta q}$$

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

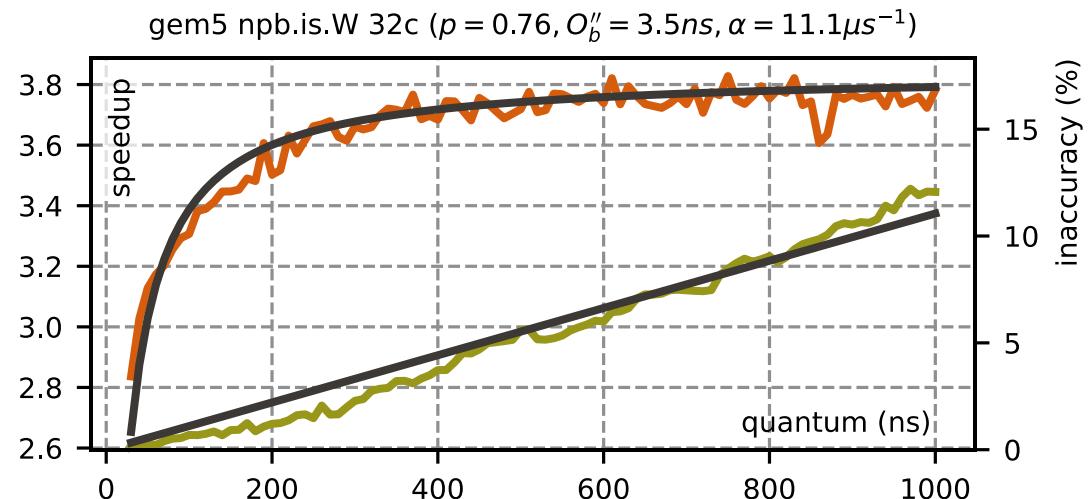
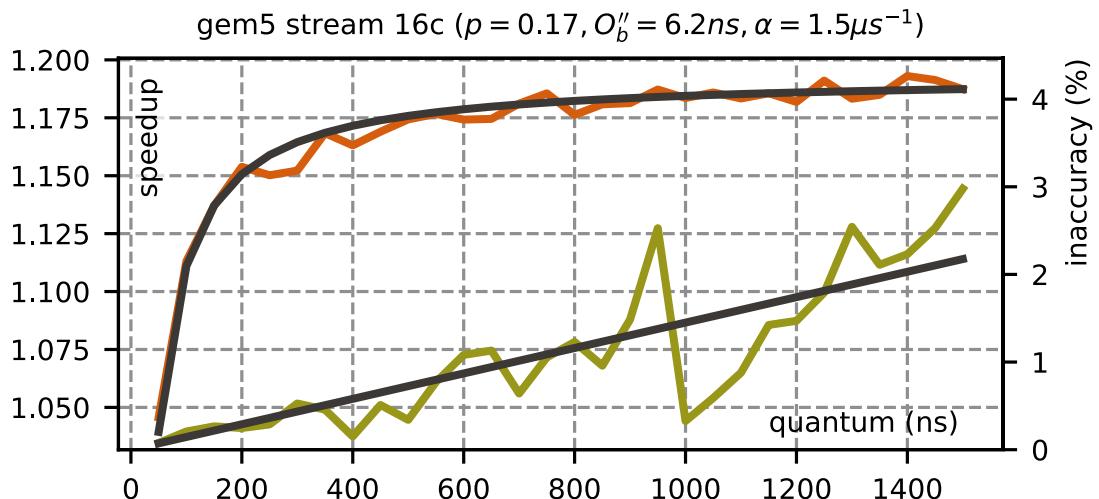
Timing inaccuracy grows linearly with the quantum ($t_{\Delta q}$)!

Results

Results: SystemC Speedup vs. Accuracy



Results: gem5 Speedup vs. Accuracy



Results: Qualitative Accuracy

Without Temporal Decoupling

```
[ 0.000075] sched_clock: 56 bits at 25MHz, [...]
[ 0.000128] Console: colour dummy device 80x25
[ 0.000134] Calibrating delay loop (skipped) preset value..
[ 0.000139] pid_max: default: 32768 minimum: 301
[ 0.000385] Mount-cache hash table entries: 32768 [...]
[ 0.000396] Mountpoint-cache hash table entries: [...]
[ 0.024140] ASID allocator initialised with 128 entries
[ 0.032140] Hierarchical SMCU implementation.
[ 0.048162] smp: Bringing up secondary CPUs ...
[ 0.080218] Detected PIPT-Icache on CPU1
```

Results: Qualitative Accuracy

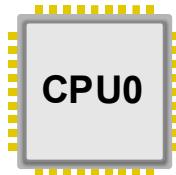
With Temporal Decoupling

```
[ 0.000075] sched_clock: 56 bits at 25MHz, [...]
[ 0.000128] Console: colour dummy device 80x25
[ 0.000134] Calibrating delay loop (skipped) preset value..
[ 0.000139] pid_max: default: 32768 minimum: 301
[ 0.000385] Mount-cache hash table entries: 32768 [...]
[ 0.000396] Mountpoint-cache hash table entries: [...]
[ 422.828066] ASID allocator initialised with 128 entries
[ 3495.801687] Hierarchical SRCU implementation.
[ 845.656091] smp: Bringing up secondary CPUs ...
[ 5877.941435] Detected PIPT-Icache on CPU1
```

Results: Qualitative Accuracy

What happened to the linux boot timestamps?

@200ns

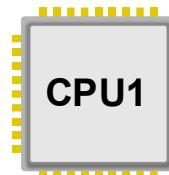


time request



Timer

@100ns

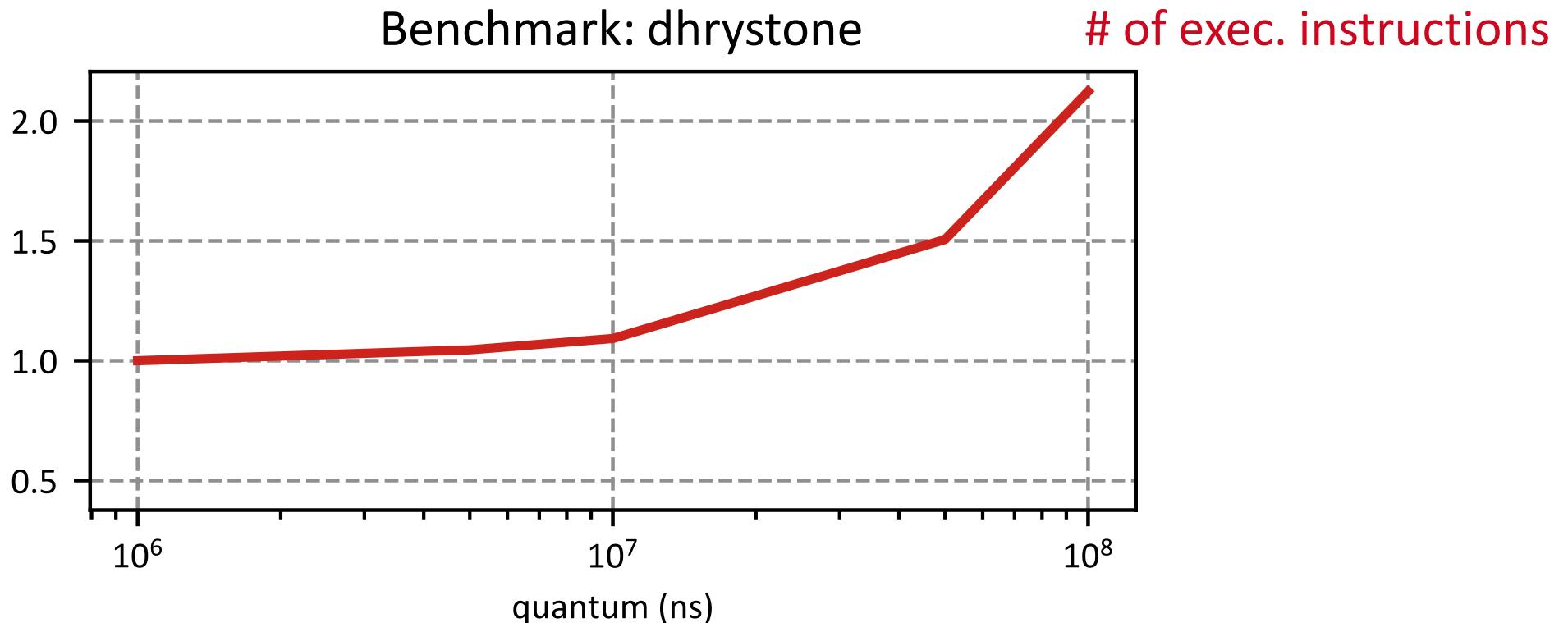


time request

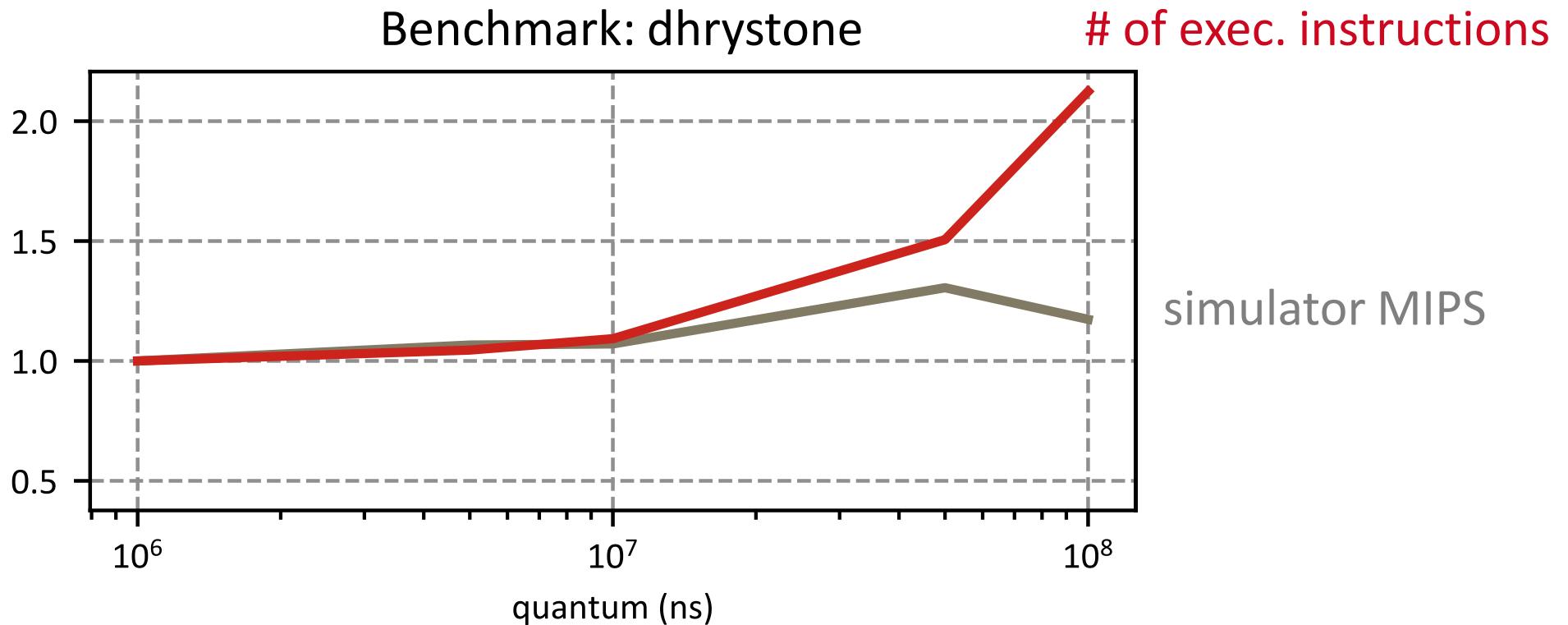
```
// From generic_timer.cc:  
uint64_t update = current_time - last_time
```

$$\text{update} = 100 - 200 = 18,446,744,073,709,551,516$$

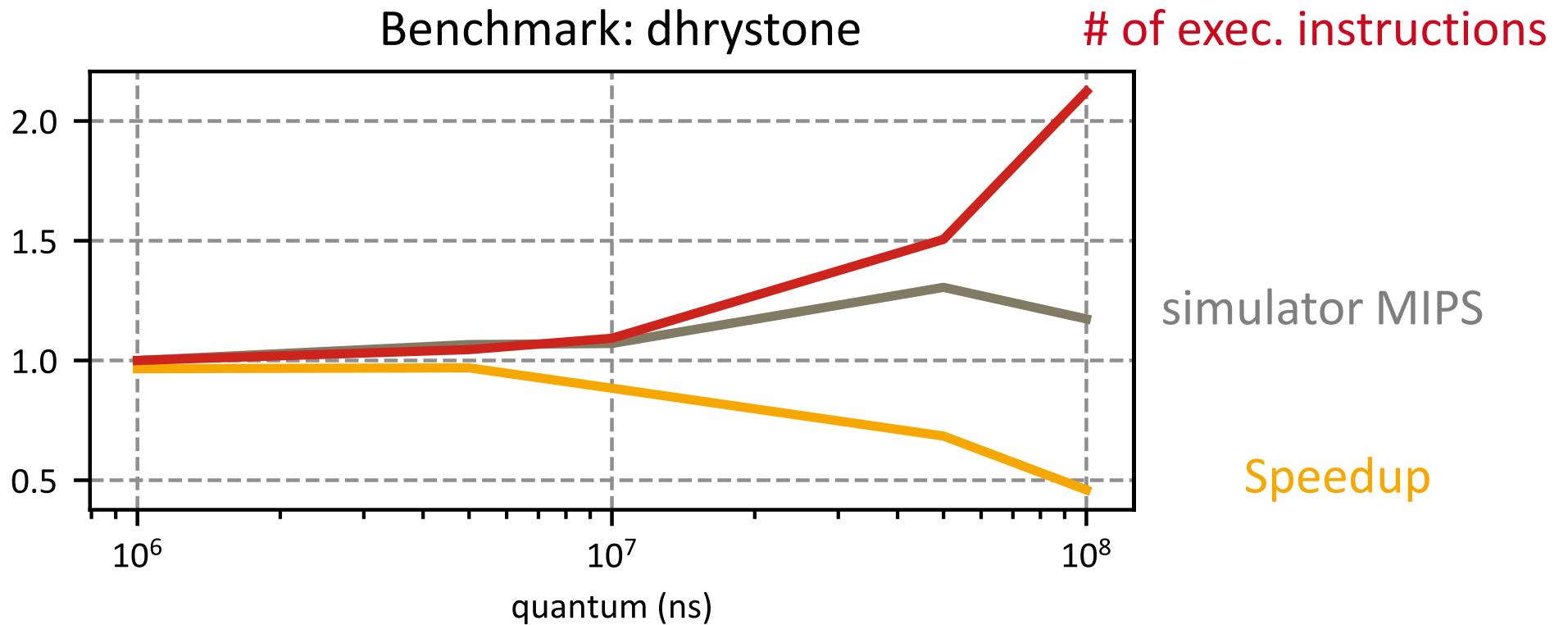
Results: Other Interesting Effects



Results: Other Interesting Effects



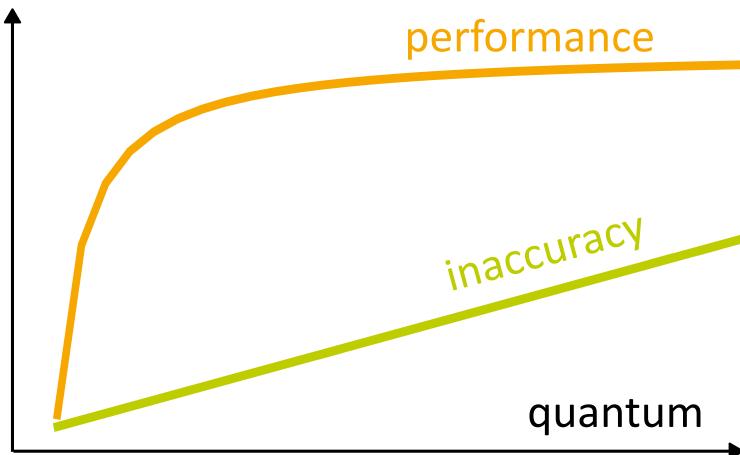
Results: Other Interesting Effects



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?

