TIMING VERIFICATION OF REAL-TIME AUTOMOTIVE ETHERNET NETWORKS: WHAT CAN WE EXPECT FROM SIMULATION?

Nicolas NAVET, University of Luxembourg
Jan R. SEYLER, Daimler A.G, Mercedes Cars
Jörn MIGGE, RealTime-at-Work (RTaW)

Use-cases for Ethernet in vehicles

Infotainment

- Synchronous traffic
- Mixed audio and video data
- MOST like

Cameras

- High data rates
- Continuous streaming
- LVDS like

Diag. & flashing

- Interfacing to external tools
- High throughput needed

Control functions ADAS

- Time-sensitive communication
- Small and large data payload
- Cover CAN / Flexray use cases and more

TWISTED-PAIR

Empirical study

Early stage verification techniques

- ✓ Simulation
- ✓ Analysis
- ✓ Lower bounds
- ✓ Performance metrics

Simulation Methodology

- ✓ **Q1:** is a single run enough ?
- ✓ **Q2:** can we run simulation in parallel and aggregate results ?
- ✓ **Q3:** simulation length ?

What to expect from simulation and analysis?

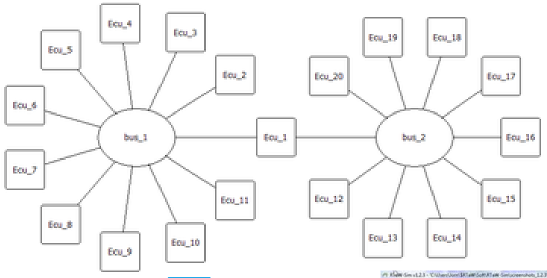
- ✓ **Q4:** is worst-case analysis accurate?
- ✓ **Q5:** simulation to derive worst-case latencies?
- ✓ **Q6:** the case of a synchronous startup

Schedulability analysis *"mathematic model of the worst-case possible situation"* VS **Simulation** *"program that reproduces the behavior of a system"*

$$K_i^k(t) \overset{\text{def}}{=} \underbrace{\left\lfloor \frac{J_i^k + \varphi_i^k(\phi^i)}{T_i^k} \right\rfloor}_{\text{max number of instances that can accumulate at critical instants}} + \underbrace{\left\lfloor \frac{t - \varphi_i^k(\phi^i)}{T_i^k} \right\rfloor + 1}_{\text{max number of instances arriving after critical instants}}$$

max number of instances that can accumulate at critical instants

max number of instances arriving after critical instants



$$S_{n+1} = F(S_n)$$

☺ Upper bounds on the perf. metrics → safe if model is correct and assumptions met

☹ Might be a gap between models and real systems → unpredictably unsafe then

☺ Models close to real systems

☺ Fine grained information

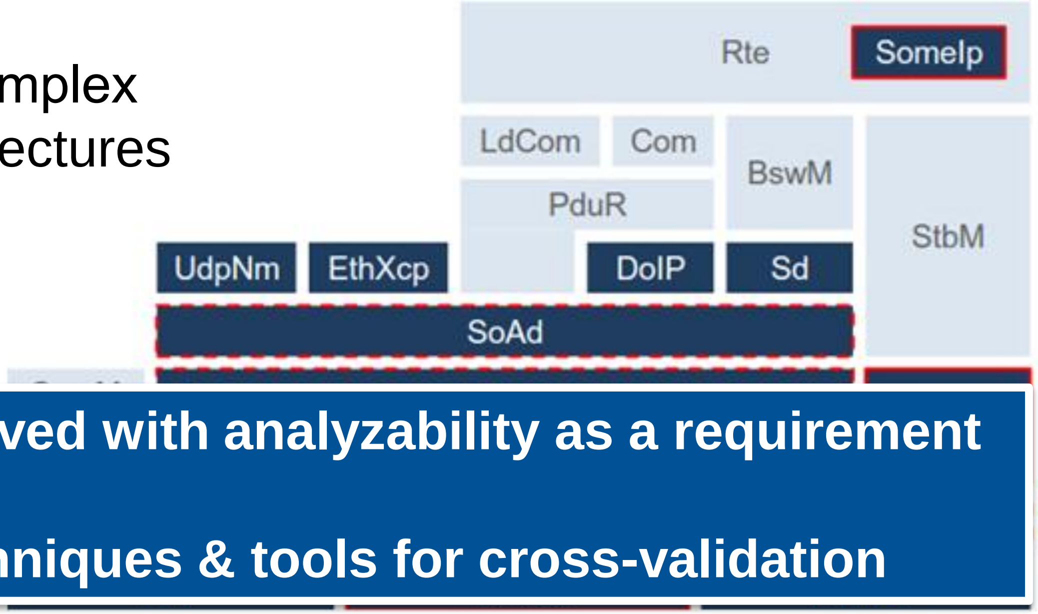
☹ Worst-case response times are out of reach - occasional deadline misses must be acceptable

Is schedulability analysis alone is sufficient ?

1. Pessimism due to conservative and coarse-grained models → over-dimensioning of the resources

2. Complexity that makes analytic models error prone and hard to validate: black-box software, unproven and published analyses, small user-base, no qualification process, no public benchmarks, …, main issue: do system meets analysis' assumptions?

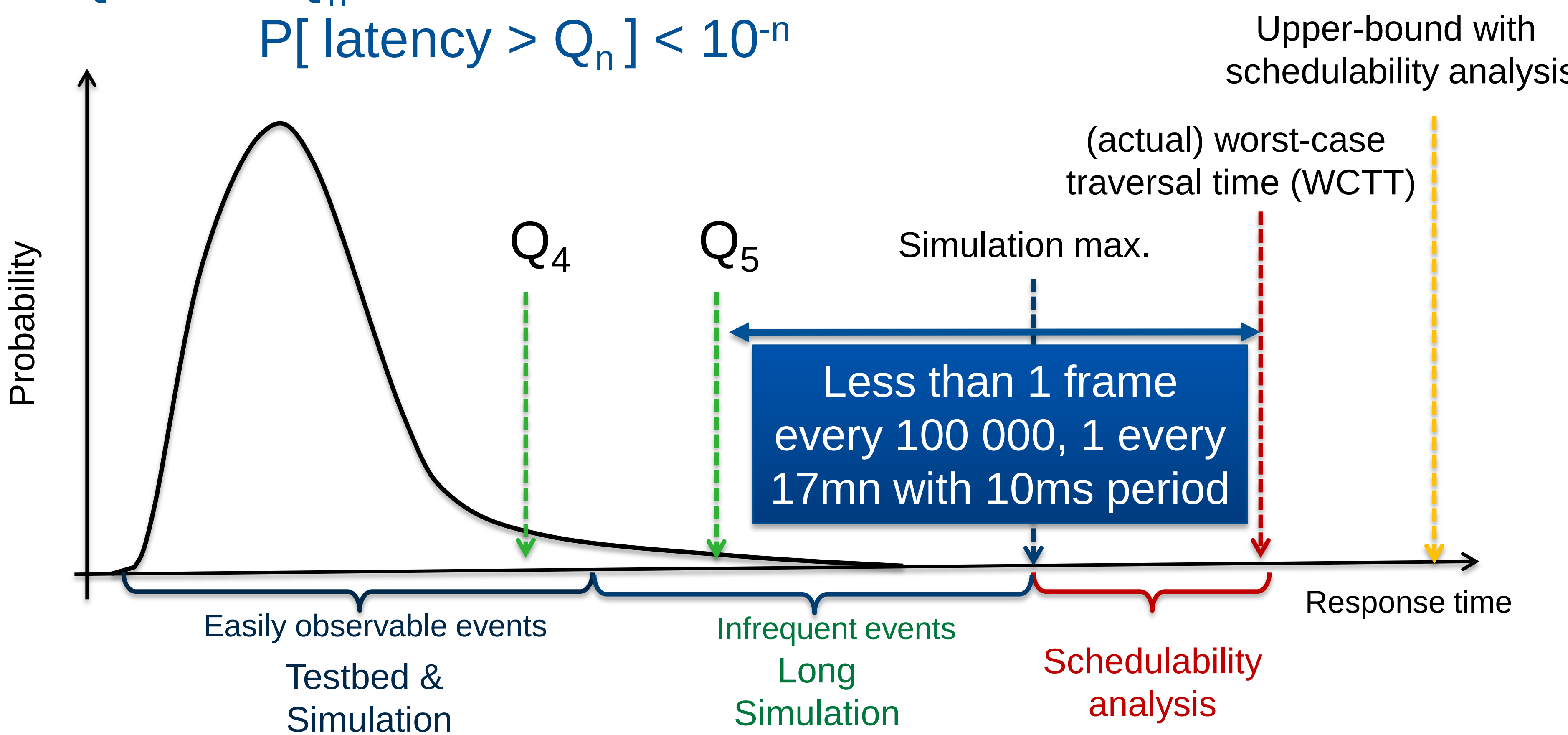
3. Inability to capture today's complex software and hardware architectures
   → e.g., Socket Adaptor



- ➢ **No, except if system conceived with analyzability as a requirement**
- ➢ **Good practice - several techniques & tools for cross-validation**

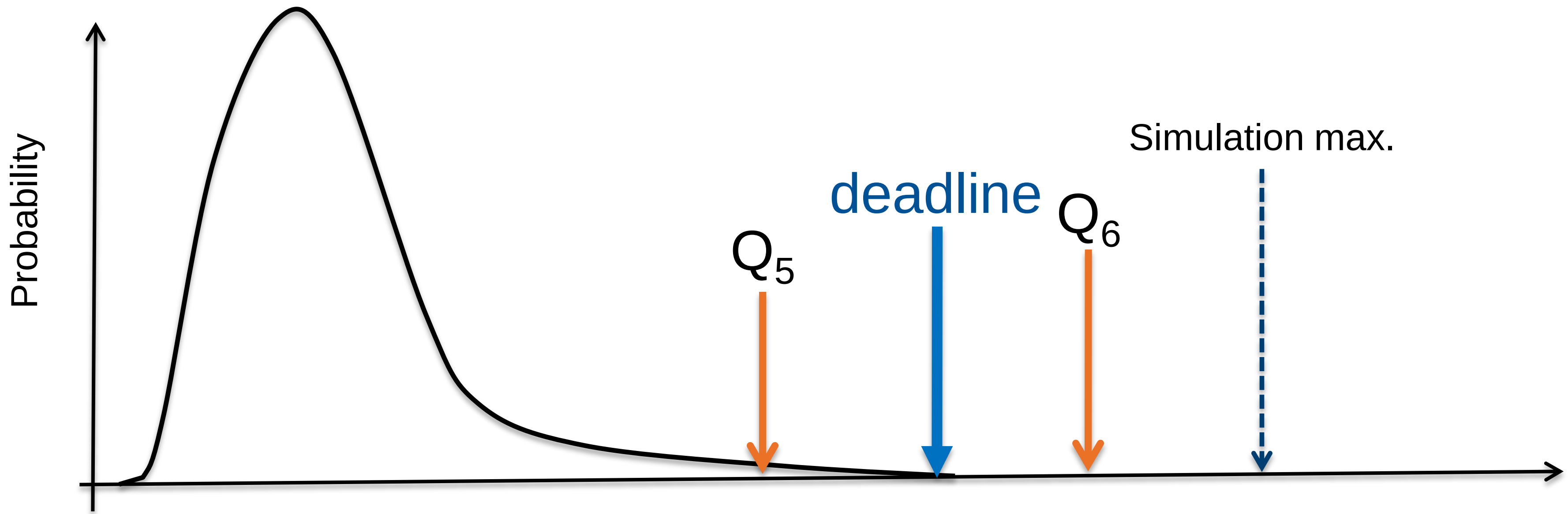
Performance metrics for frame latencies – or buffer usage

Quantile Q_n: smallest value such that
$P[\text{ latency} > Q_n] < 10^{-n}$



Using simulation means accepting a quantified risk system must be robust to that

Working with quantiles in practice – see [5]



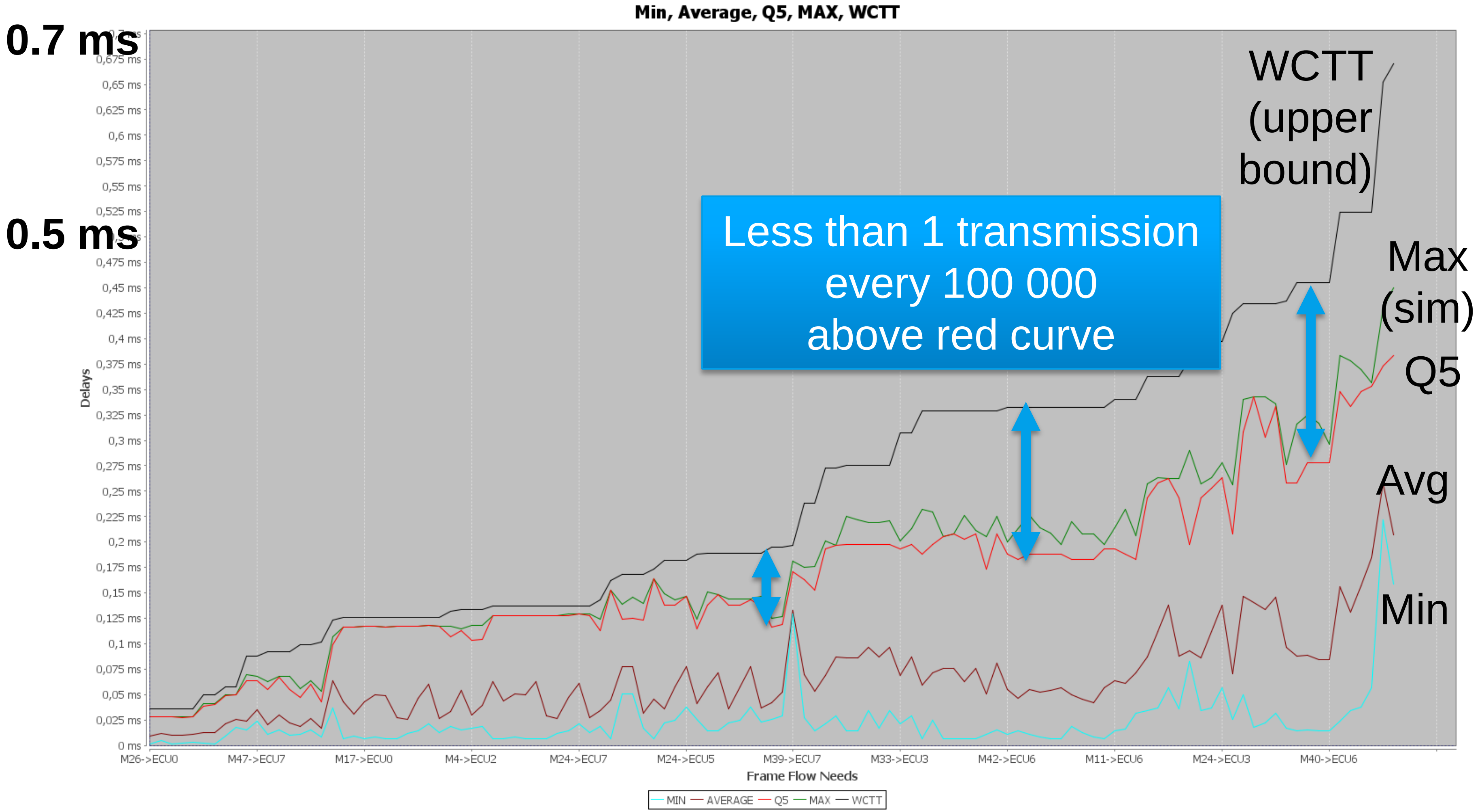
1. Identify frame deadline
2. Decide the tolerable risk → target quantile
3. Simulate "sufficiently" long
4. If target quantile value is below deadline, performance objective is met

Quantiles vs average time between deadline misses

Quantile	One frame every …	Mean time to failure Frame period = 10ms	Mean time to failure Frame period = 500ms
Q3	1 000	10 s	8mn 20s
Q4	10 000	1mn 40s	≈ 1h 23mn
Q5	100 000	≈ 17mn	≈ 13h 53mn
Q6	1000 000	≈ 2h 46mn	≈ 5d 19h
…	…	…	

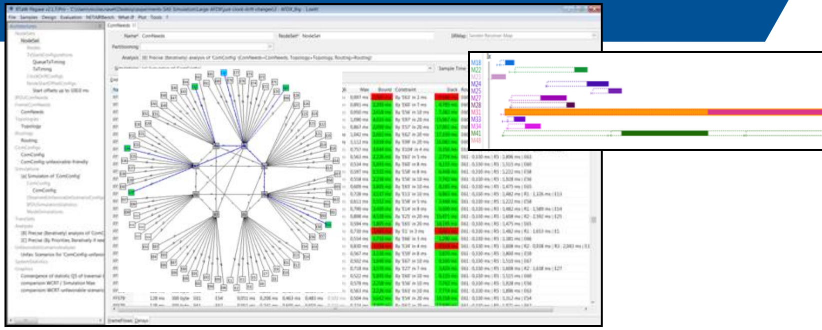
Warning : successive failures in some cases might be temporally correlated, this can be assessed.

Performance metrics: illustration on a Daimler prototype network (ADAS, control functions)



Case-study #1: flows sorted by increasing WCTT

Software Toolset and performance evaluation techniques



✓***RTaW-Pegase*** – modeling and analysis of switched Ethernet (industrial, automotive, avionics) + CAN (FD) and ARINC

✓Higher-level protocols (e.g. Some IP) and functional behavior can be programmed in CPAL® language [4]

✓Developed since 2009 in partnership with Onera

✓Ethernet users include Daimler Cars, Airbus Helicopters and ABB

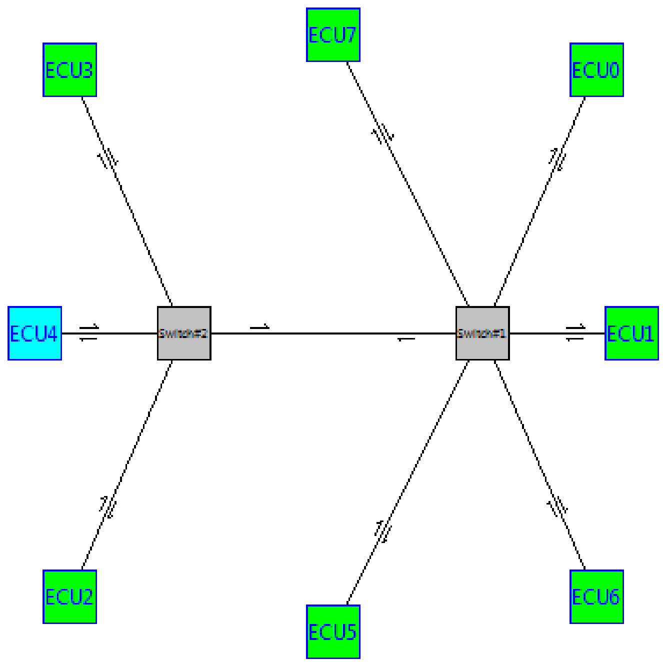
Performance evaluation techniques

✓Worst-case Traversal Time (WCTT) analysis - based on state-of-the-art Network-Calculus, all algorithms are published, core proven correct [2]

✓Timing-accurate Simulation – *ps* resolution, $\approx 4 \cdot 10^6$ events/sec on a single core (I7 - 3.4Ghz), suited up to $(1\text{-}10^6)$ quantiles

✓Lower-bounds on the WCTT - “unfavorable scenario” [3]

CASE-STUDY #1 - Mercedes prototype Ethernet network

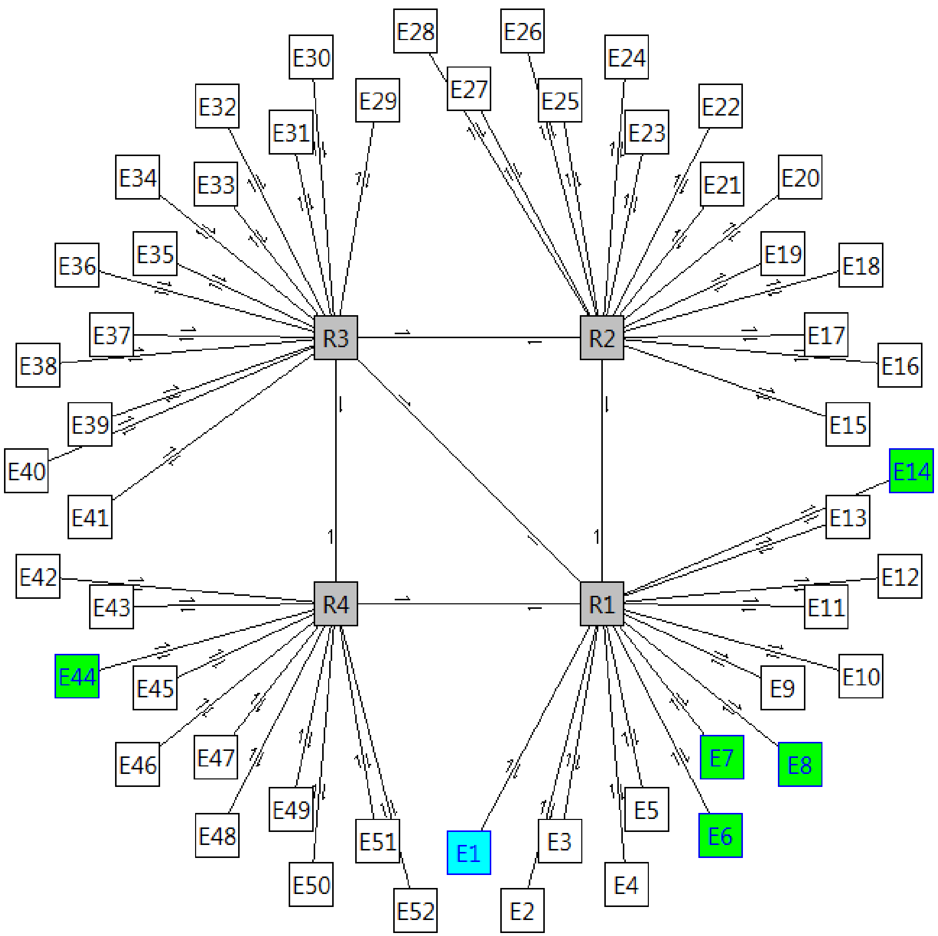


[RTaW-Pegase screenshot]

Topology of case-study #1 with a broadcast stream sent by ECU4

#Nodes	8
#Switches	2
#Maximum switching delay	6us
#streams	**58**
#priority levels	2
Cumulated workload	0,33Gbit/s
Link data rates	100Mbit/s and 1Gbit/s (2 links)
Latency constraints	confidential
Number of receivers	1 to 7 (avg: 2.1)
Packet period	0.1 to 320ms
Frame size	51 to 1450bytes

CASE-STUDY #2 – medium AFDX network

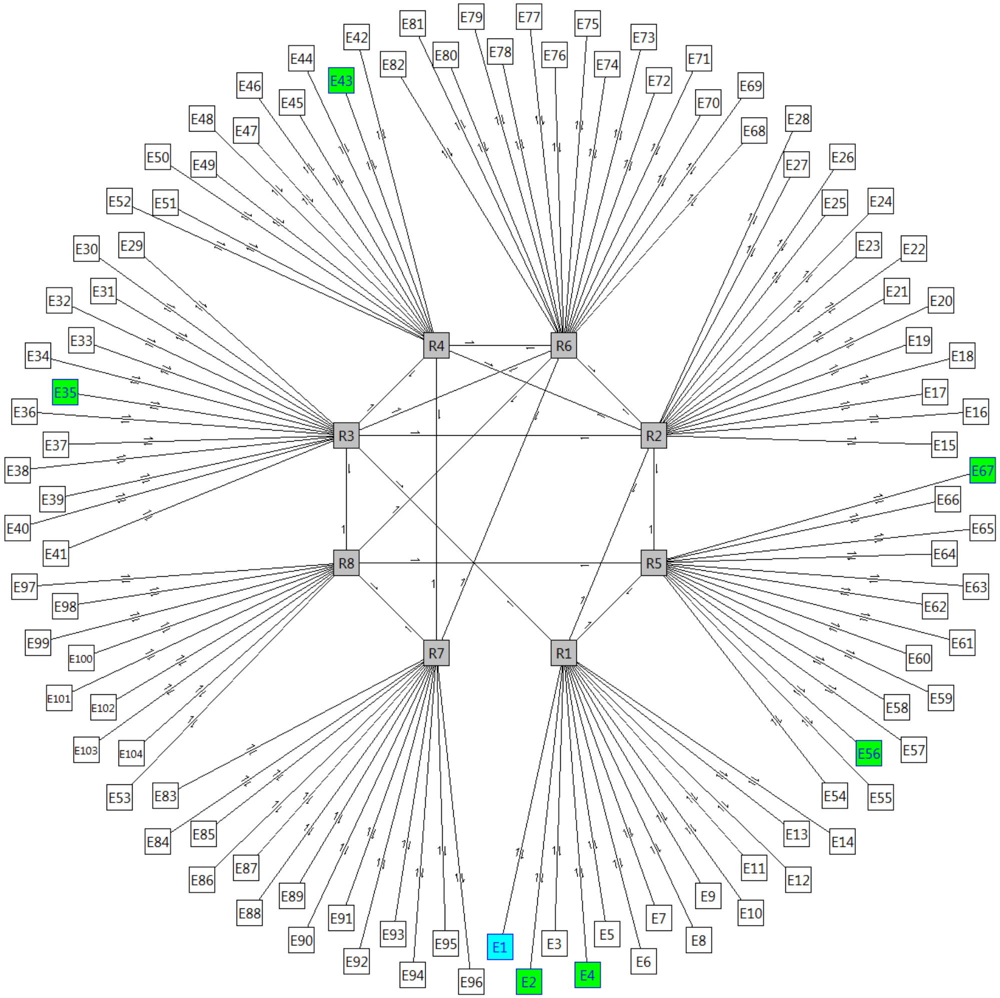


[RTaW-Pegase screenshot]

Topology of case-study #2 with a multi-cast stream sent by node E1

#Nodes	52
#Switches	4
#Maximum switching delay	7us
#streams	**3214**
#priority levels	none
Cumulated workload	0.49Gbit/s
Link data rates	100Mbit/s
Latency constraints	2 to 30ms
Number of receivers	1 to 42 (avg: 7.1)
Packet period	2 to 128ms
Frame size	100 to 1500bytes

CASE-STUDY #3 – large AFDX network, as used in civil airplanes



[RTaW-Pegase screenshot]

Topology of case-study #3 with a multi-cast stream sent by node E1

#Nodes	104
#Switches	8
#Maximum switching delay	7us
#streams	**5701**
#priority levels	5
Cumulated workload	0.97Gbit/s
Link data rates	100Mbit/s
Latency constraints	1 to 30ms
Number of receivers	1 to 83 (avg: 6.2)
Packet period	2 to 128ms
Frame size	100 to 1500bytes

System model and experimental setup

- ✓ Simulation and analysis models are in line in terms of what they model

- ✓ Assumptions:
  - Streams are strictly periodic and successive packets of a stream are all of the same size
  - Nodes are not synchronized on startup, they start to send within 100ms (same results with larger values)
  - Communication stack reduced to a queue: FIFO or priority queue
  - Store-and-forward communication switches with a sub-10us max. switching delays
  - No transmission errors, no packet losses in the switches

- ✓ Simulation’s specific setup:
  - Nodes’ clock drifts: 200ppm (same results with 400ppm)
  - Each experiment repeated 10 times (with random offsets and clock drifts)
  - Long simulation means at least 2 days of functioning time (samples large enough for Q5 for sub-100ms flows)

Simulation methodology

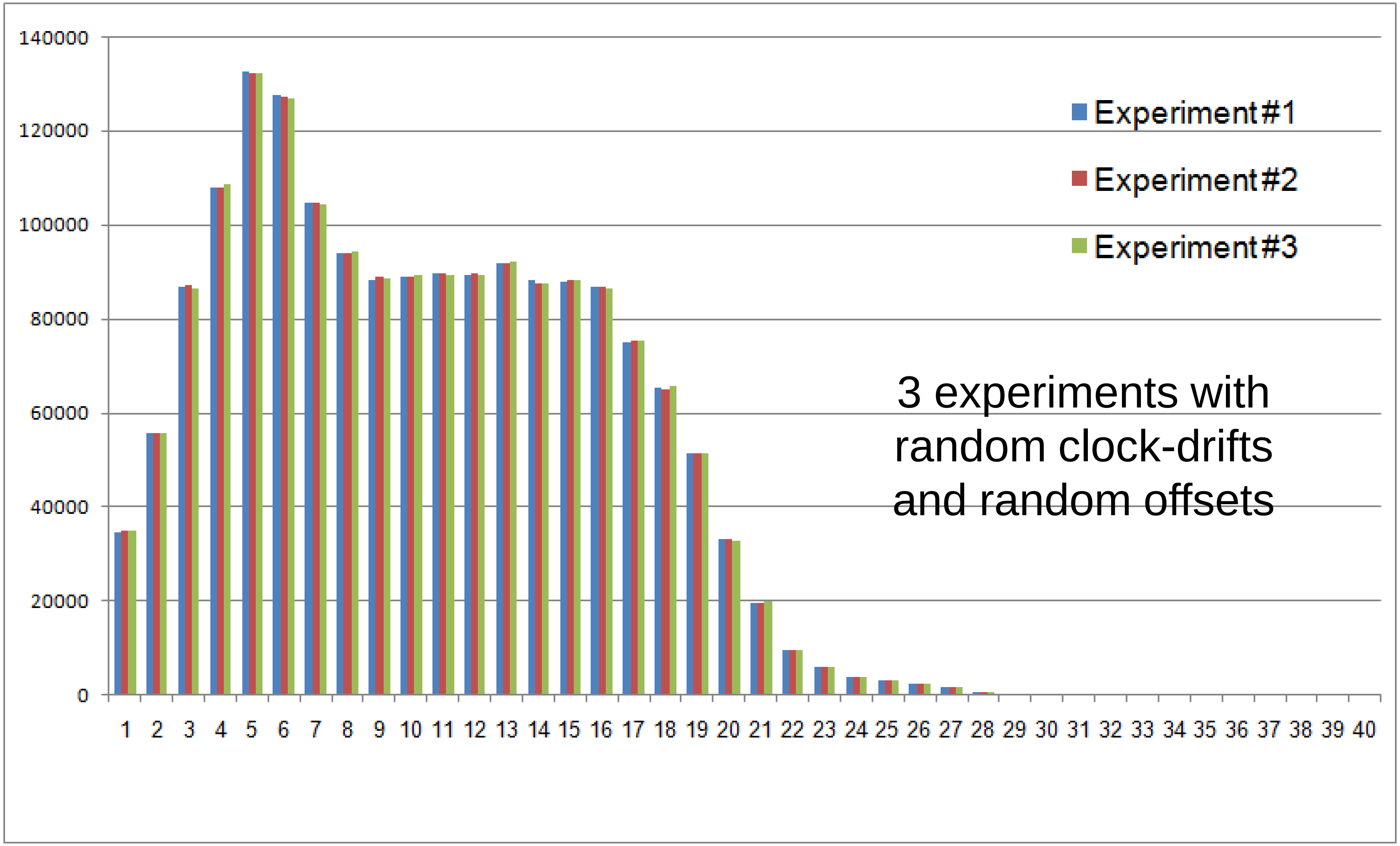
Ergodicity of the simulated system

- ✓ Intuitively, *“a dynamic system is said to be ergodic if, after a certain time, every trajectory of the system leads the same distribution of the state of the system, called the equilibrium state”*

- ✓ Consequences:
  - Q1: a single simulation run enough, initial conditions do not matter
  - Q2: results from simulation run in parallel can be aggregated – how long is the transient state that occurs at the start ?

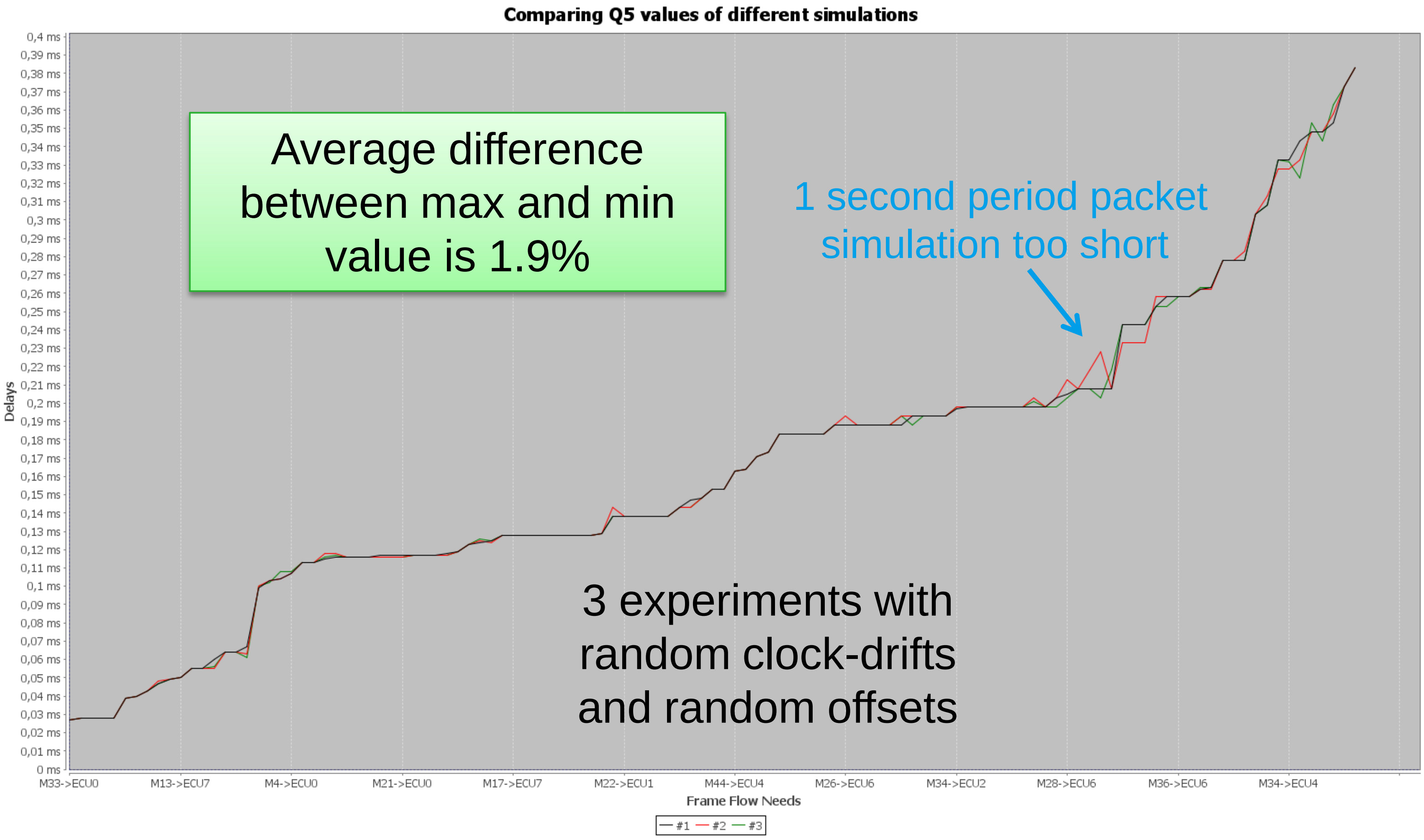
- ✓ Empirical approach: test if the distributions converge though the Q5 quantiles:
  - Random offsets and random clock drifts
  - Random offsets and fixed clock drifts
  - Fixed offsets and random clock drifts

Q5 quantile: visual verification for a number of frames



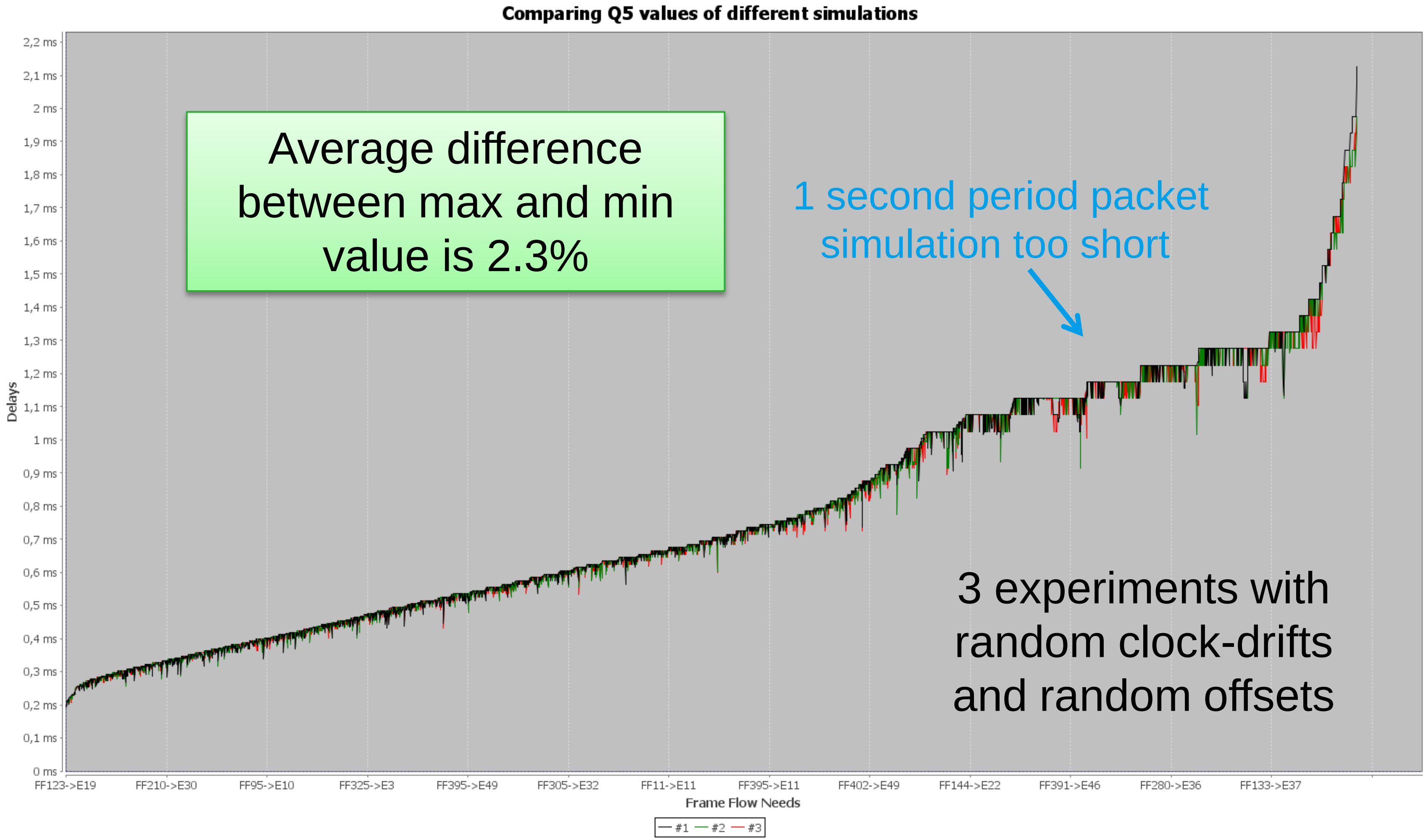
Case-study #1: flows sorted by increasing WCTT

Q5 : Case-study #1 – convergence of the Q5 quantiles



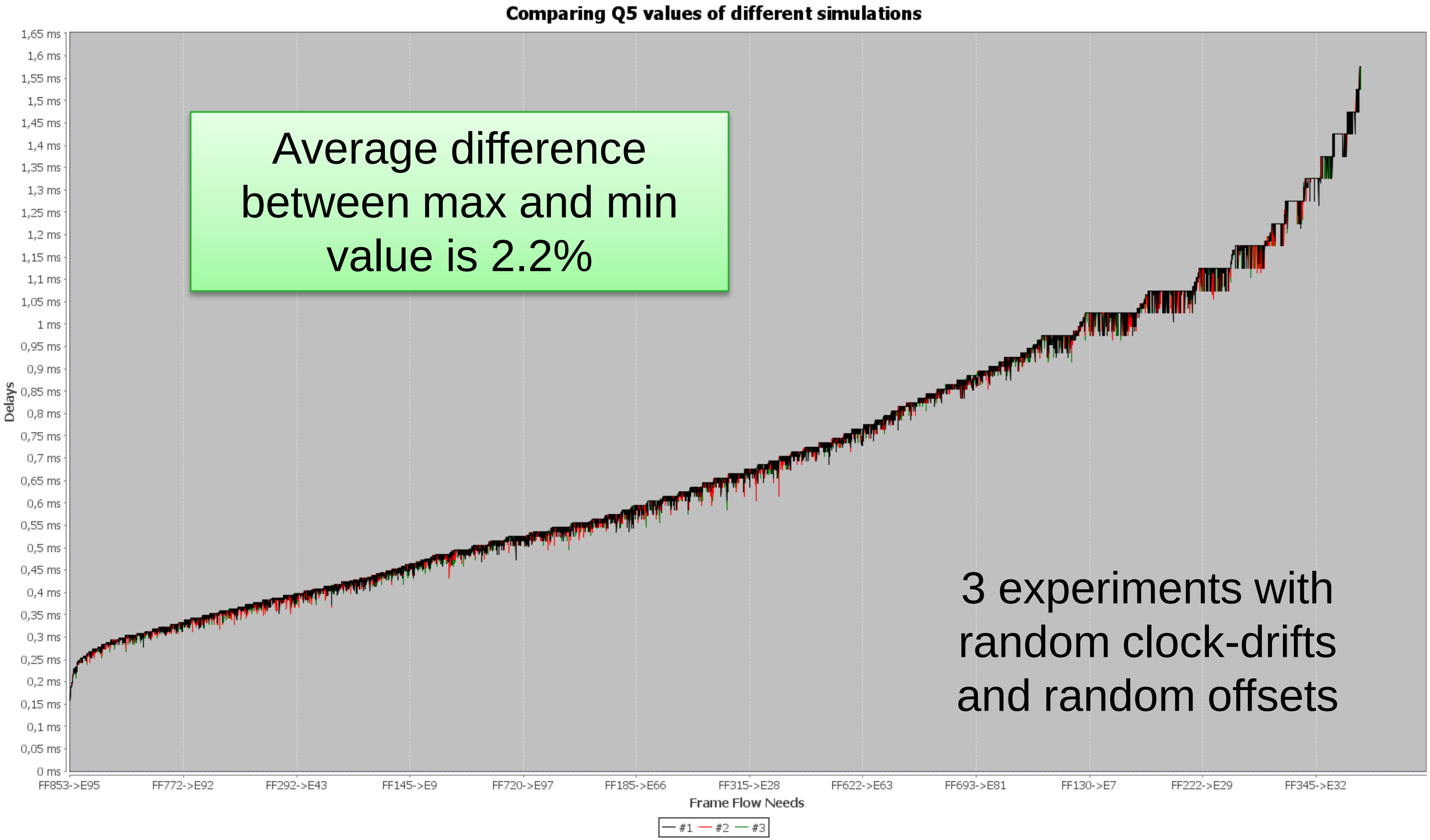
Case-study #1: flows sorted by increasing WCTT

Q5 : Case-study #2 – convergence of the Q5 quantiles



Case-study #2: flows sorted by increasing WCTT

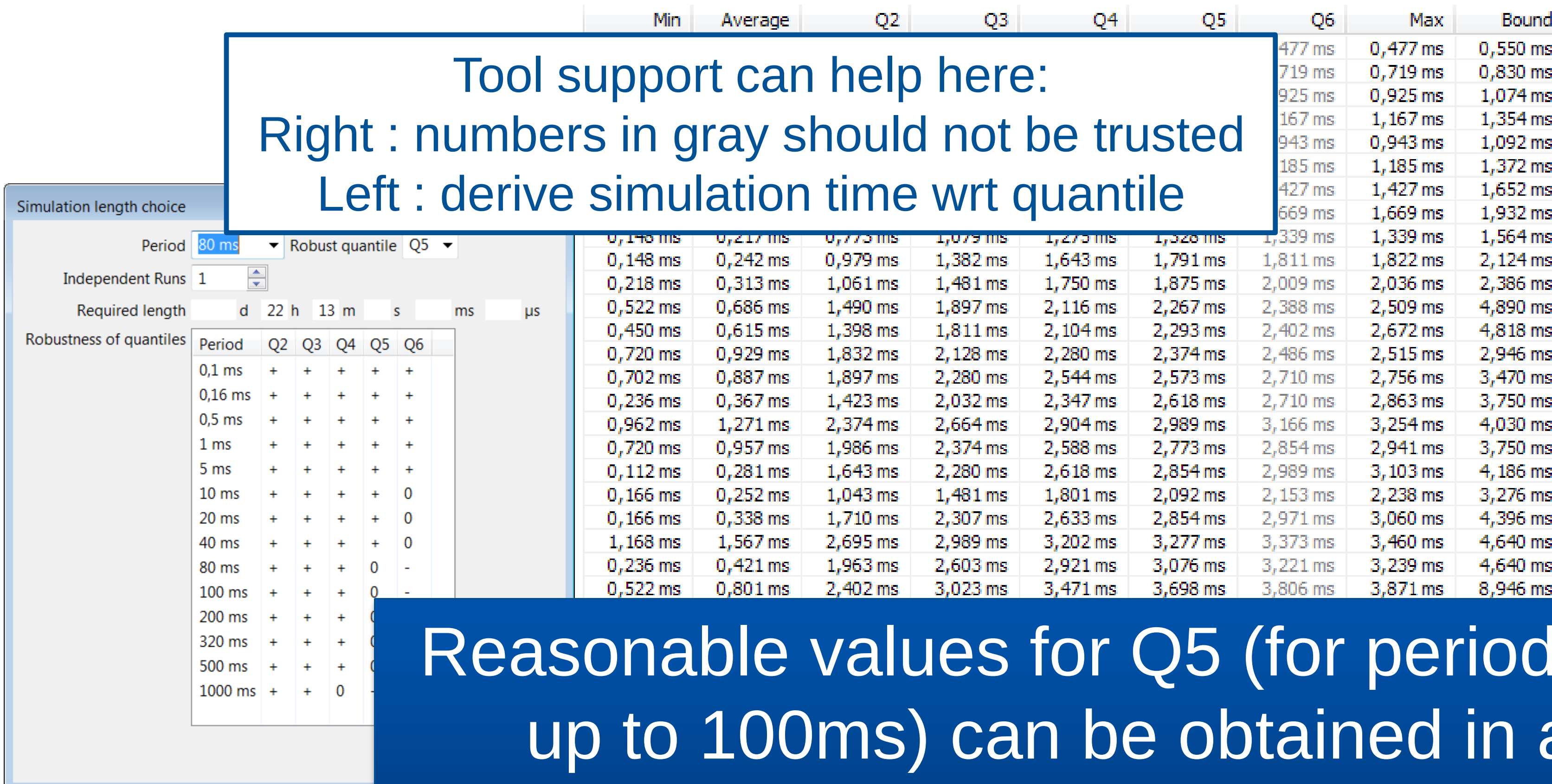
Q5 : Case-study #3 – convergence of the Q5 quantiles



Case-study #1: flows sorted by increasing WCTT

Determine the minimum simulation length

- ✓ time needed for convergence
- ✓ reasonable # of values: a few tens…



Tool support can help here:
Right : numbers in gray should not be trusted
Left : derive simulation time wrt quantile

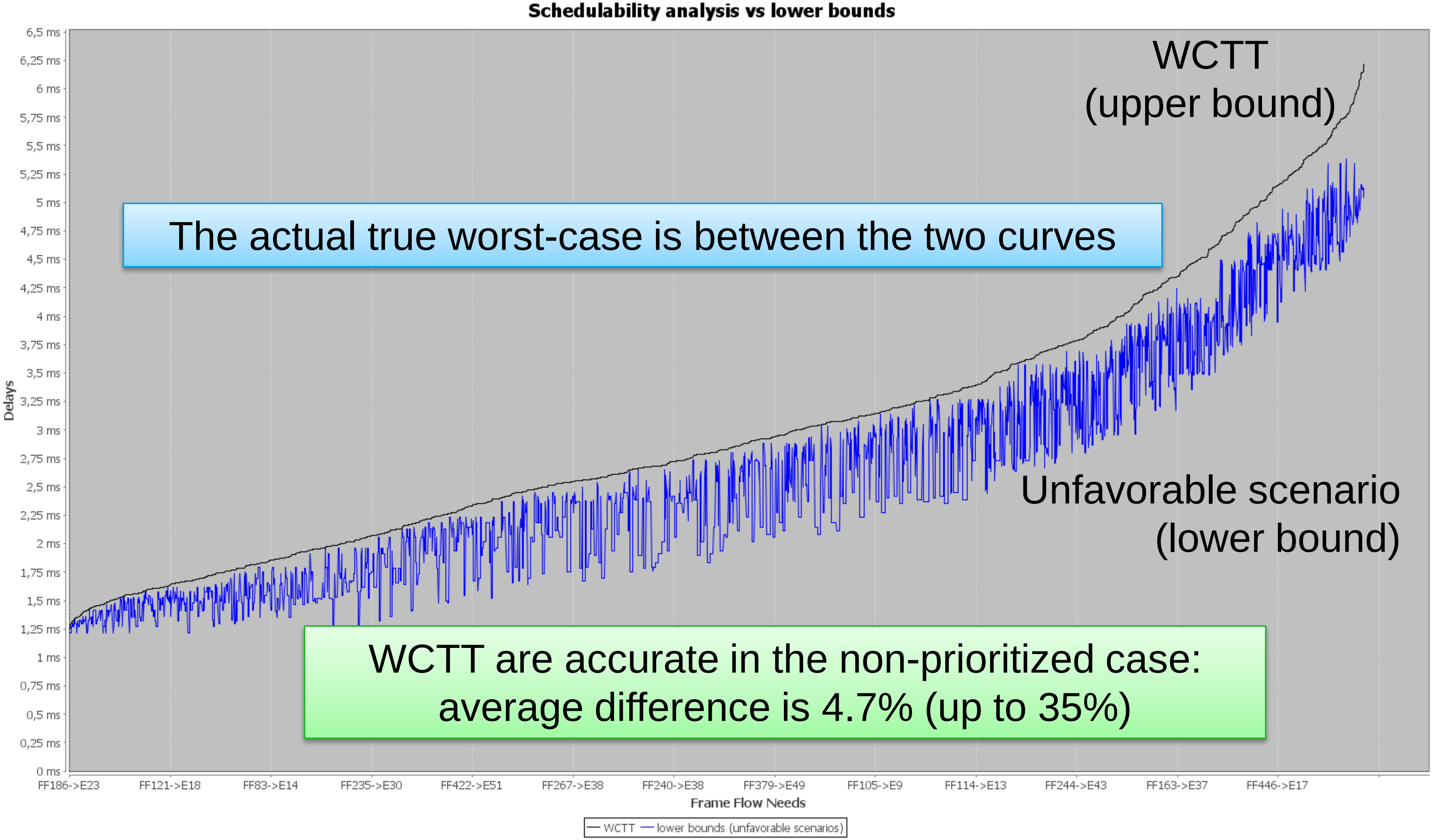
[RTaW-pegase screenshot]

Reasonable values for Q5 (for periods up to 100ms) can be obtained in a few hours of simulation

What to expect from simulation and analysis ?

Analysis (Network-Calculus)
VS
Lower-bound (unfavorable scenario)
VS
Timing-Accurate Simulation

Q4: Are Worst-Case Traversal Times (WCTT) computed with Network Calculus accurate?

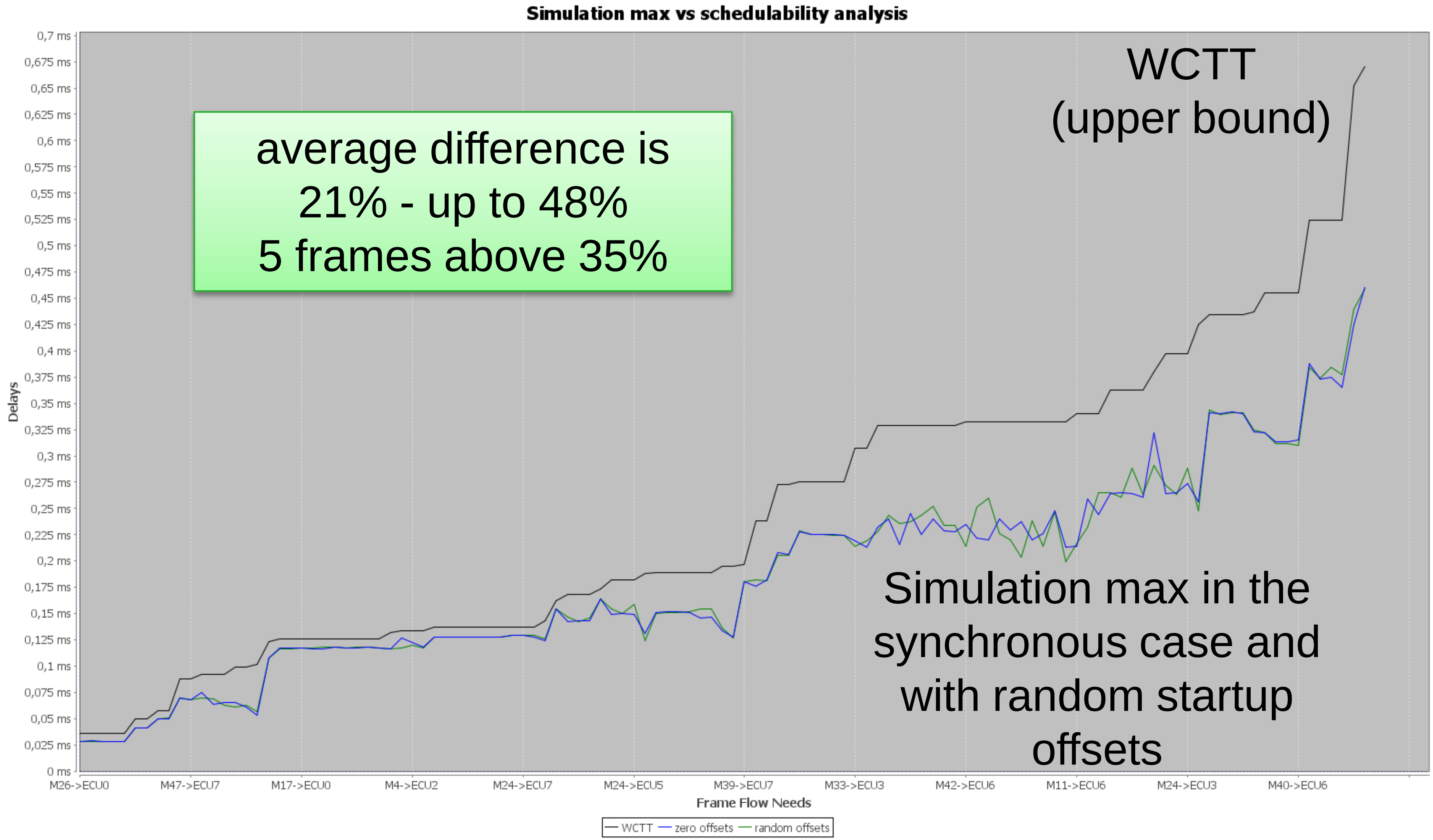


The actual true worst-case is between the two curves

WCTT are accurate in the non-prioritized case: average difference is 4.7% (up to 35%)

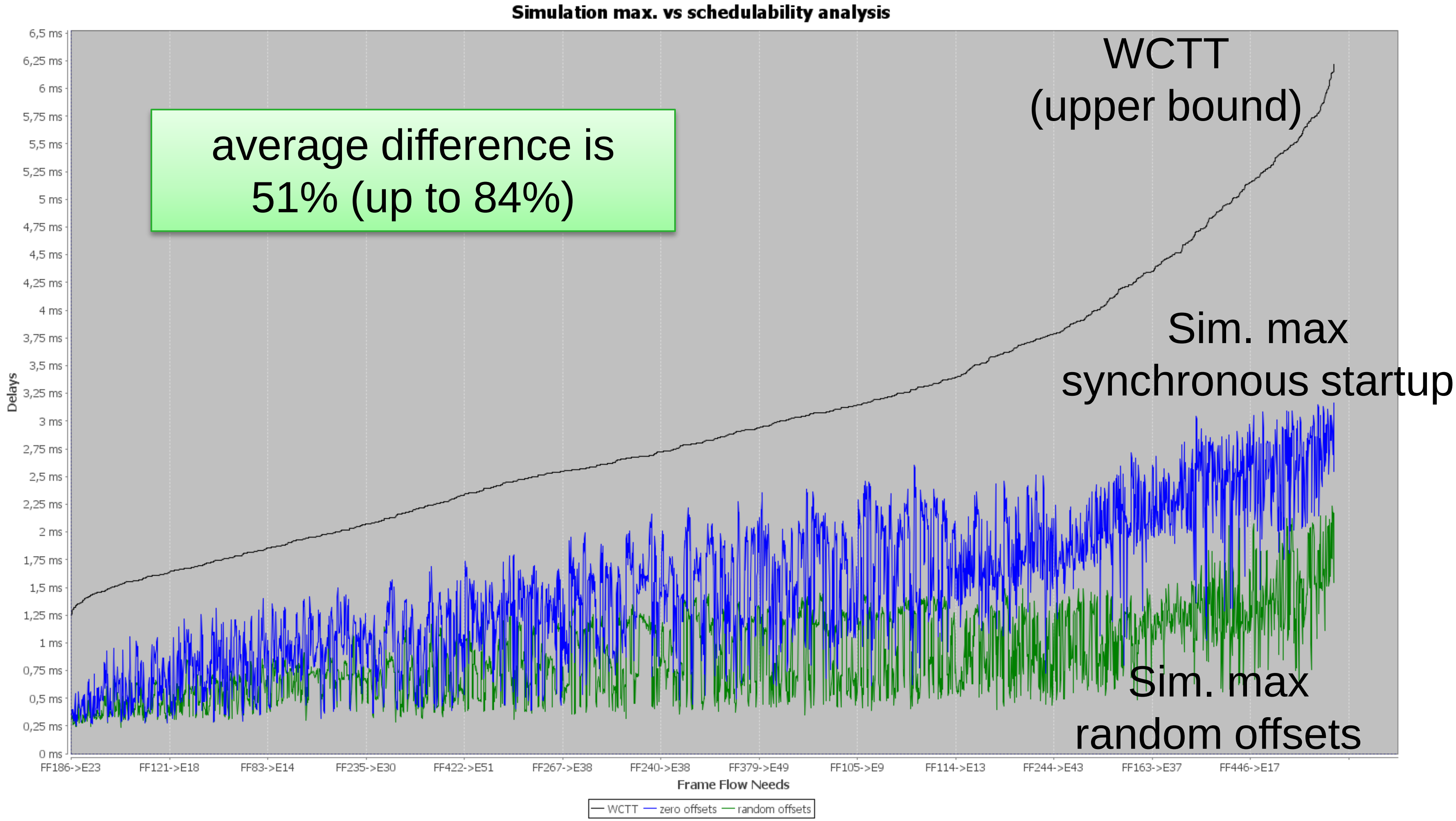
Case-study #2 : flows sorted by increasing WCTT

Q5 : Case-study #1 – difference between analysis upper bounds and simulation maxima



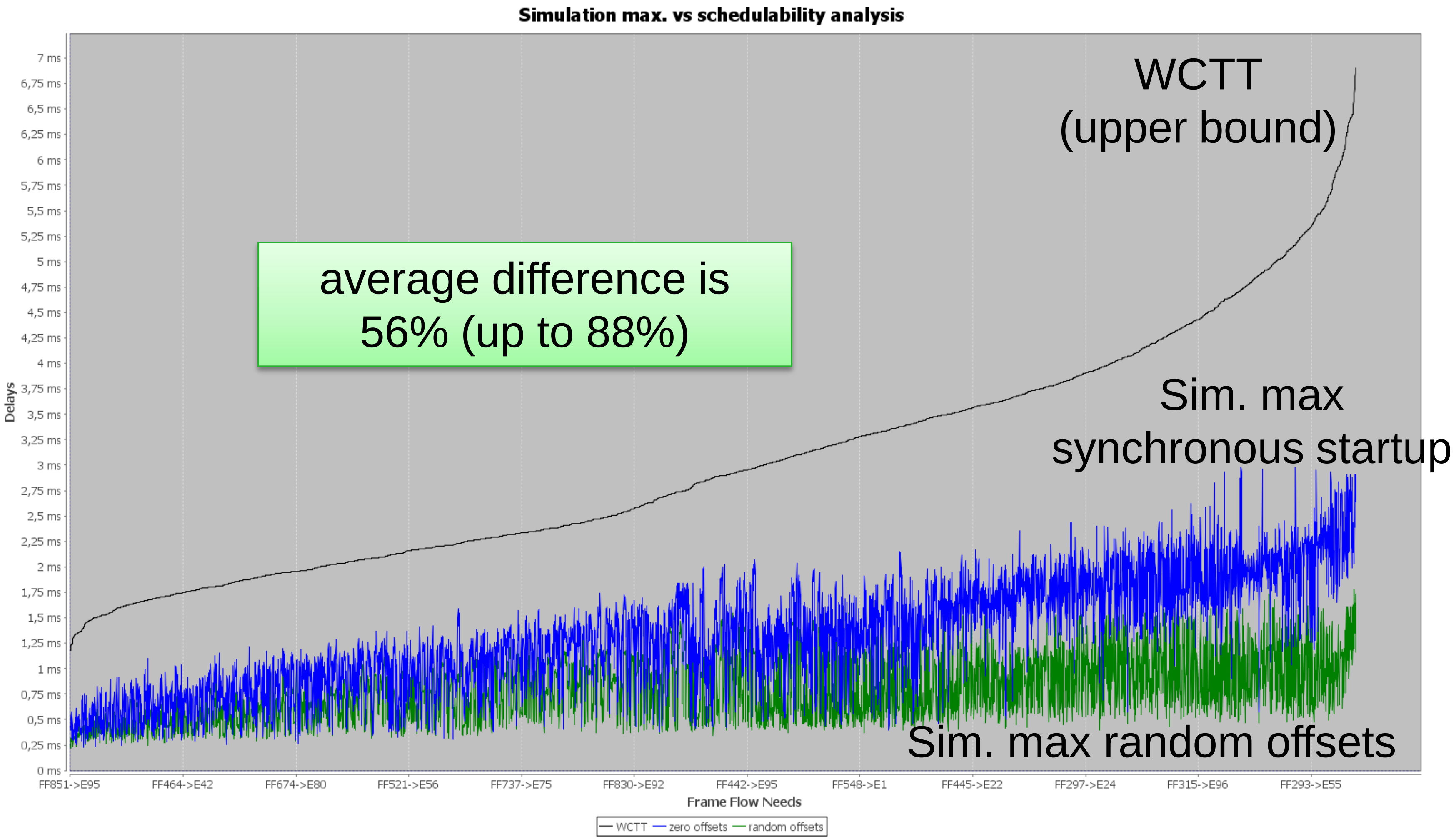
Case-study #1: flows sorted by increasing WCTT

Q5 : Case-study #2 – difference between analysis upper bounds and simulation maxima



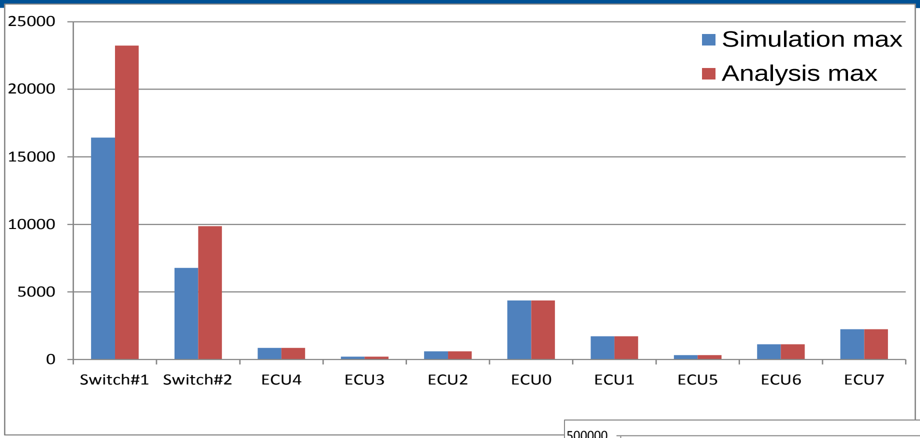
Case-study #2 : flows sorted by increasing WCTT

Q5 : Case-study #3 – difference between analysis upper bounds and simulation maxima

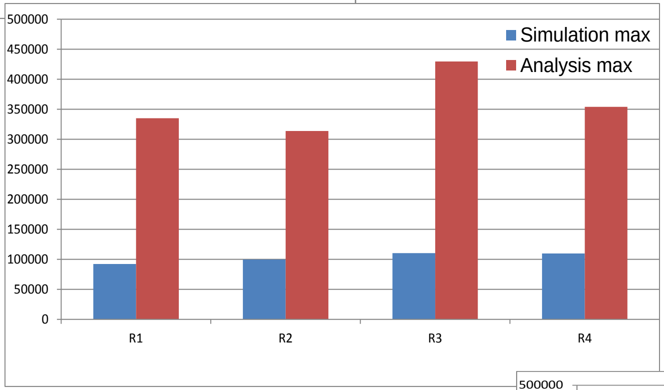


Case-study #3 : flows sorted by increasing WCTT

Q5 : Memory usage in the switches: difference between analysis upper bounds and simulation maxima

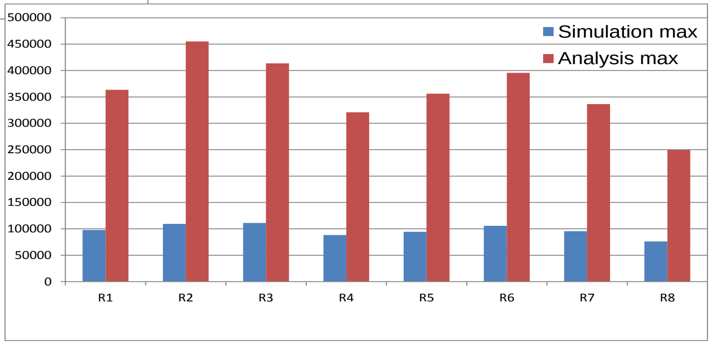


Case-study #1:
max. difference 31%



Case-study #2:
max. difference 74%

Ongoing work to reduce the pessimism of the memory usage analysis



Case-study #3:
max. difference 76%

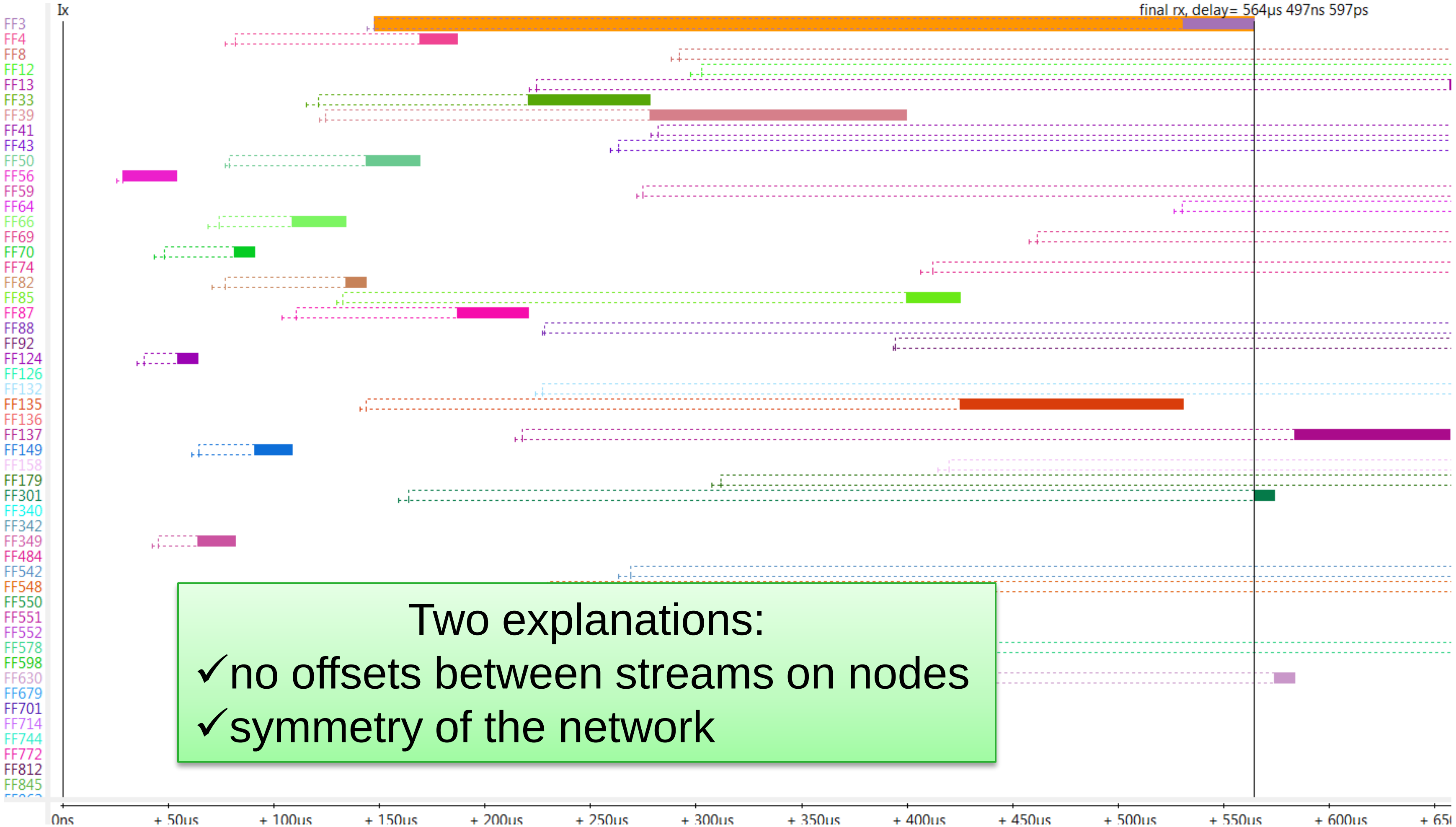
Performance evaluation techniques - Key takeaways

- ✓ State-of-the-start Network-Calculus is an accurate and fast technique for switched Ethernet - can be coupled with other types schedulability analysis for CAN (FD), gateways, ECUs.
- ✓ Deriving lower-bounds with unfavorable scenarios approaches is key to validate correctness and accuracy → more research still needed here
- ✓ Simulation suited to assess – with high confidence - the performances in a typical functioning mode → worst-case latencies/buffer usage are out of reach - except in small systems

Worst-case latencies are extremely rare events (less than once every 10^6 transmissions) - if network can be made robust to these cases, then designing with simulation is more effective in terms of resource usage

Q6 : synchronous startup of the node leads to very unfavorable trajectories

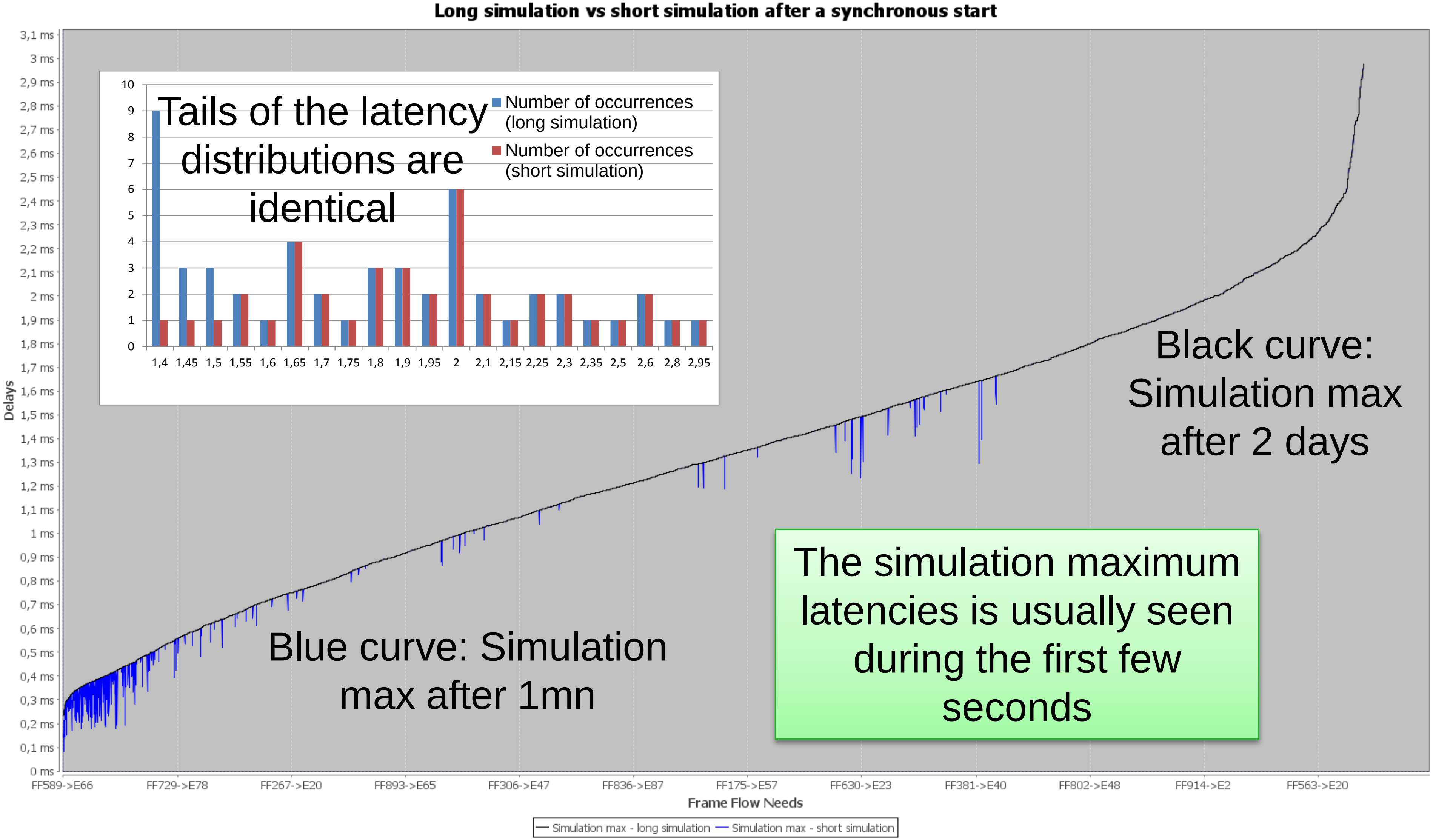
Synchronous startup of the system : many large latencies observed shortly in after startup - statistics are biased wrt typical functioning mode



Two explanations:
- ✓no offsets between streams on nodes
- ✓symmetry of the network

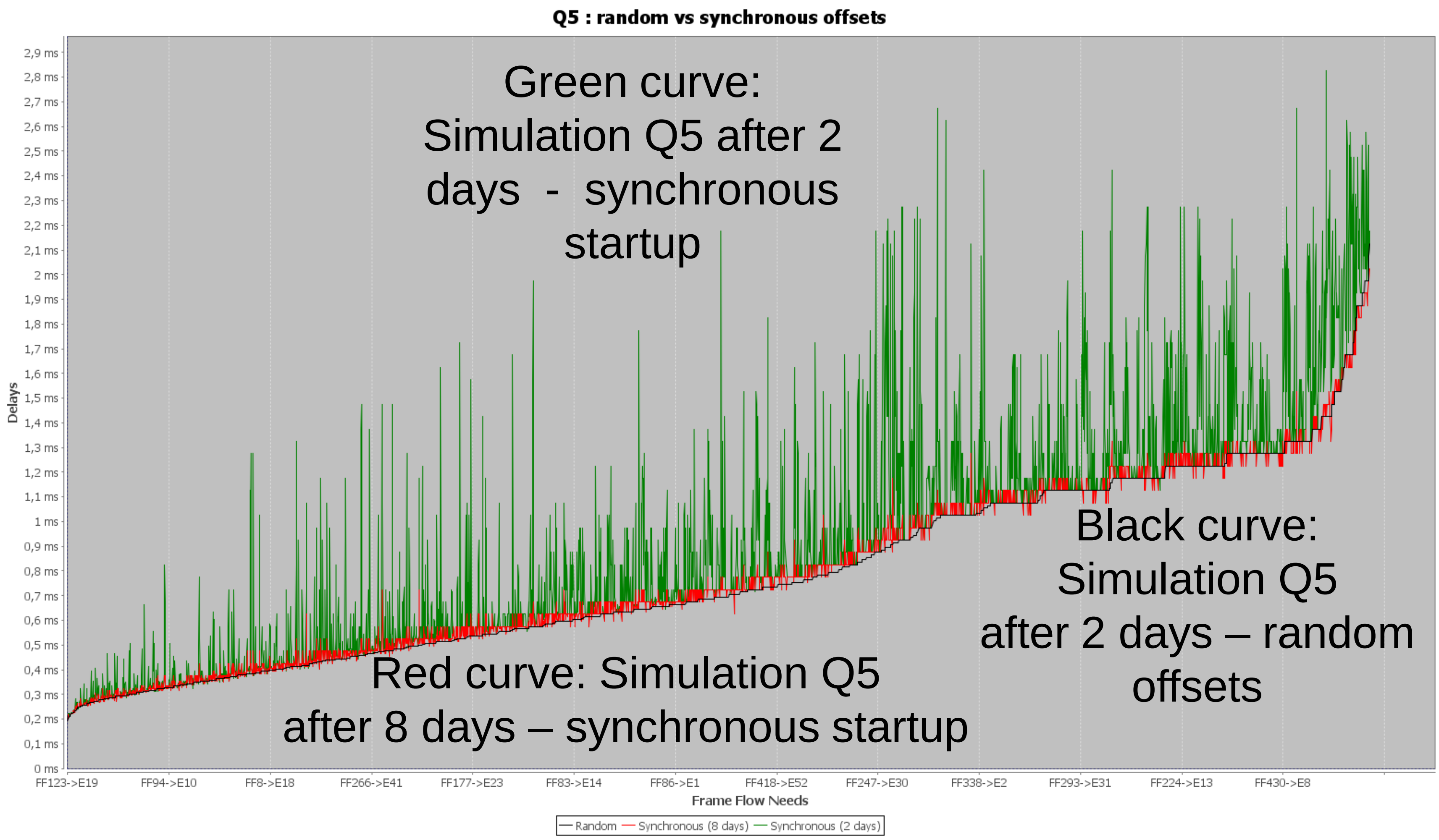
Case-study #3 - maximum latencies observed in simulation in last switch for flow FF3 (top) occurring immediately after a synchronous startup

Synchronous startup of the system – short simulation are enough for maxima



Case-study #3 : flows sorted by increasing simulation maximum (2 days)

Synchronous startup of the system – all other statistics eventually converge, but transient state takes time to be amortized



Case-study #3 : flows sorted by increasing simulation maximum

Concluding remarks

- ✓ Timing verification techniques & tools should not be trusted blindly → body of good practices should be developed
- ✓ AUTOSAR communication stacks support the numerous automotive communication requirements at the expense of complexity → schedulability analyses cannot capture everything
- ✓ Simulation is well suited to automotive systems that can tolerate deadline misses with a *controlled* risk
- ✓ Today: timing accurate simulation of complete heterogeneous automotive communication architectures
- ✓ Tomorrow: system-level simulation with models of the *functional* behavior
- ✓ Ergodicity, evidenced here empirically for Ethernet, must be studied theoretically at a the scope of the system

Thank you

Interested in this talk? Please consult the technical report available next week from www.realtimeatwork.com

References

Interested in this talk? Please consult the technical report available from www.realtimeatwork.com

[1] N. Navet, J. Seyler, J. Migge, “Timing verification of real-time automotive Ethernet networks: what can we expect from simulation?”, Technical report, 2015.

[2] E. Mabille, M. Boyer, L. Fejoz, and S. Merz, “Certifying Network Calculus in a Proof Assistant”, 5th European Conference for Aeronautics and Space Sciences (EUCASS), Munich, Germany, 2013.

[3] H. Bauer, J.-L. Scharbarg, C. Fraboul, “Improving the Worst-Case Delay Analysis of an AFDX Network Using an Optimized Trajectory Approach“, IEEE Transactions on Industrial informatics, Vol 6, No. 4, November 2010.

[4] CPAL – the Cyber-Physical Action Language, freely available from http://www.designcps.com, 2015.

[5] N. Navet, S. Louvart, J. Villanueva, S. Campoy-Martinez, J. Migge, “Timing verification of automotive communication architectures using quantile estimation“, Embedded Real-Time Software and Systems (ERTS 2014), Toulouse, France, February 5-7, 2014.