Buffer Space Allocation for Real-Time Priority-Aware Networks

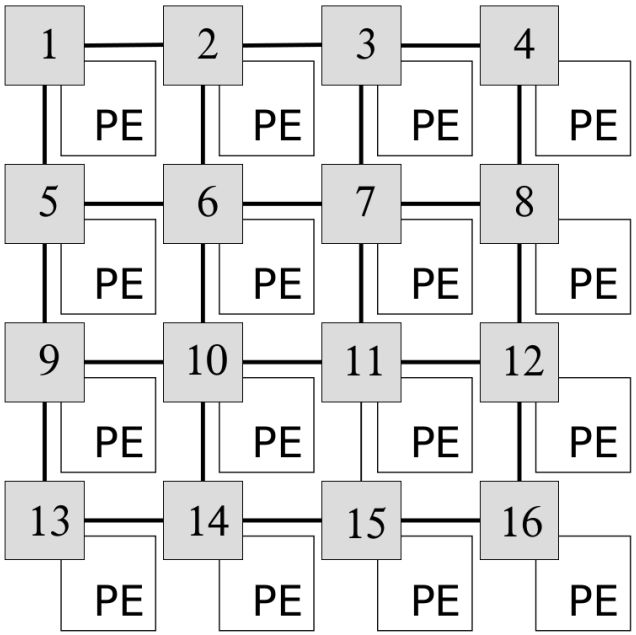
Hany Kashif and Hiren Patel

Electrical and Computer Engineering

UNIVERSITY OF WATERLOO

Motivation

- Multiple processing units require predictable communication
- Increasing data communication demands
  - Design, analyze and implement novel interconnects for real-time
- Goal:
  - Introduce buffer constraints in SLA of PAR NoC
  - Buffer allocation

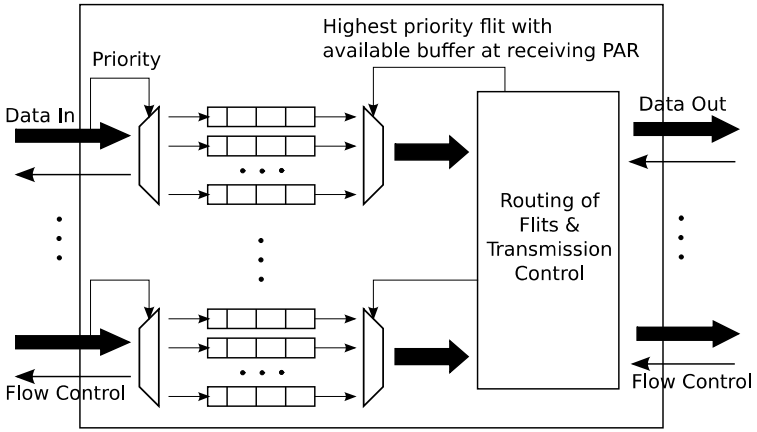


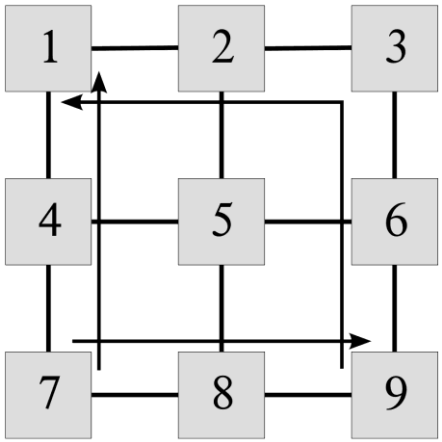
H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Motivation

- Router model proposed by Shi and Burns [RTSS08]
  - Communication has distinct priorities
  - Fixed-priority pre-emption at flit level
  - Response-time analysis to derive WC bounds (FLA)
    - Works for any buffer size



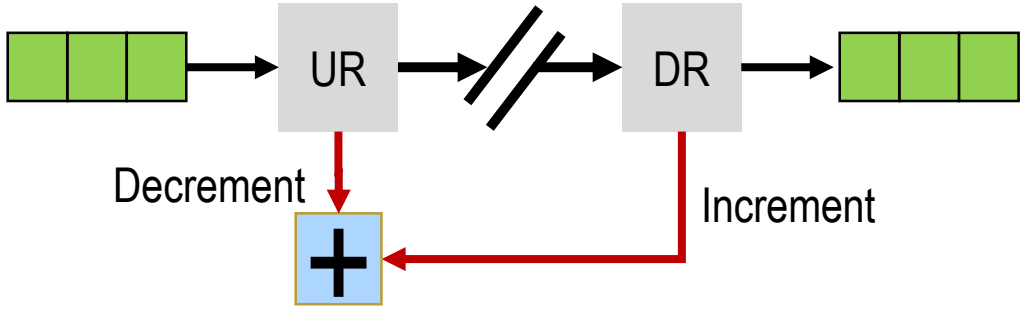


- SLA [TC15] assumption: **infinite buffers** in VCs

Problem Statement

- Extend SLA
  - Worst-case response times with buffer constraints using credit feedback policy

- Buffer space allocation
  - Given finite number of buffers per router, allocate buffers to VCs



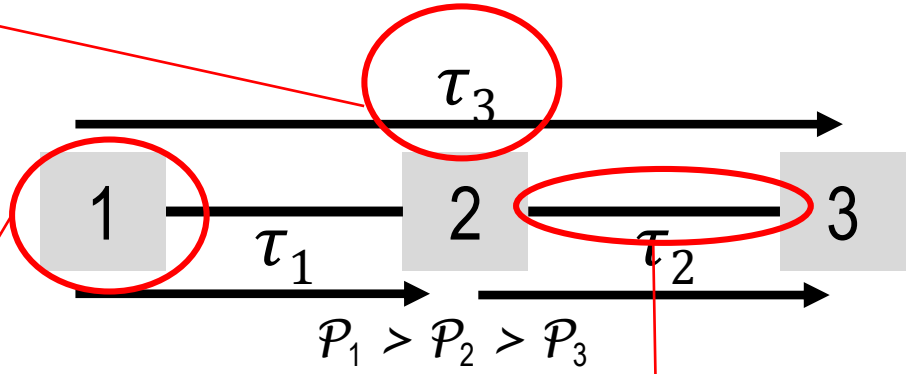
H. Kashif and H. Patel (RTAS 2016)

Outline

- Motivation
- Problem statement
- Model
- Stage-level analysis
  - Buffer constraints with credit feedback flow control
- Buffer space allocation algorithm
- Experimental evaluation
- Conclusions

Model

- Communication task
  - Priority
  - Period
  - Deadline
  - Basic link latency
    - # flits in a packet
  - Release jitter

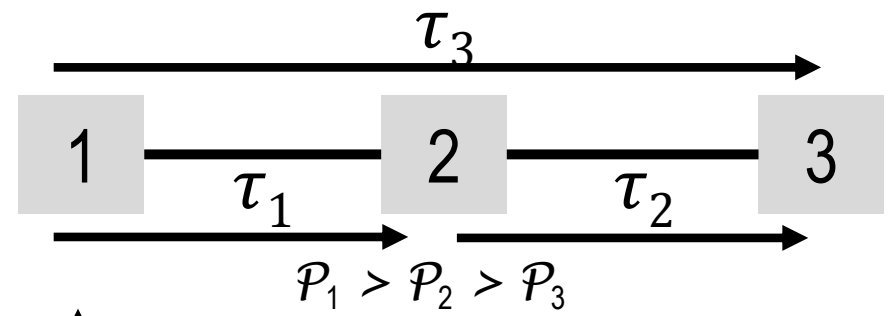


Processing element with router

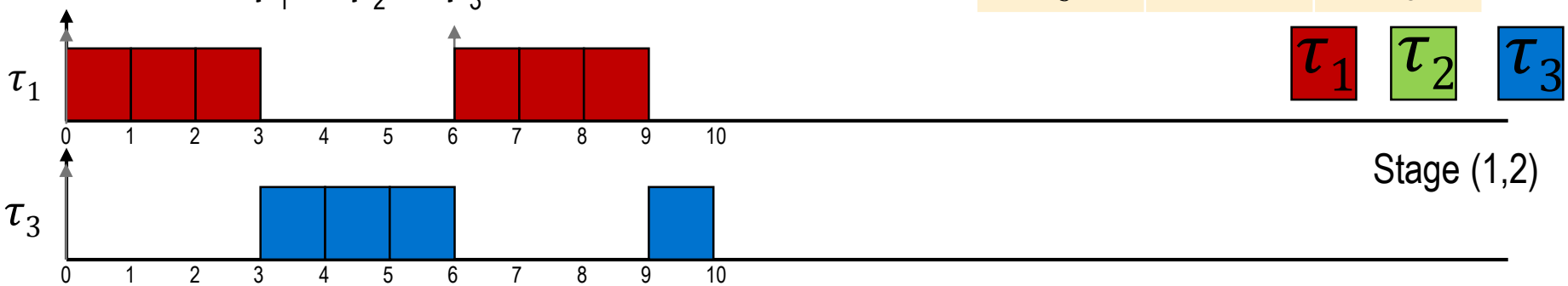
Stage

- Assume
  - Given mapping of communication tasks onto NoC
  - Given priorities

Stage-level Analysis (SLA)



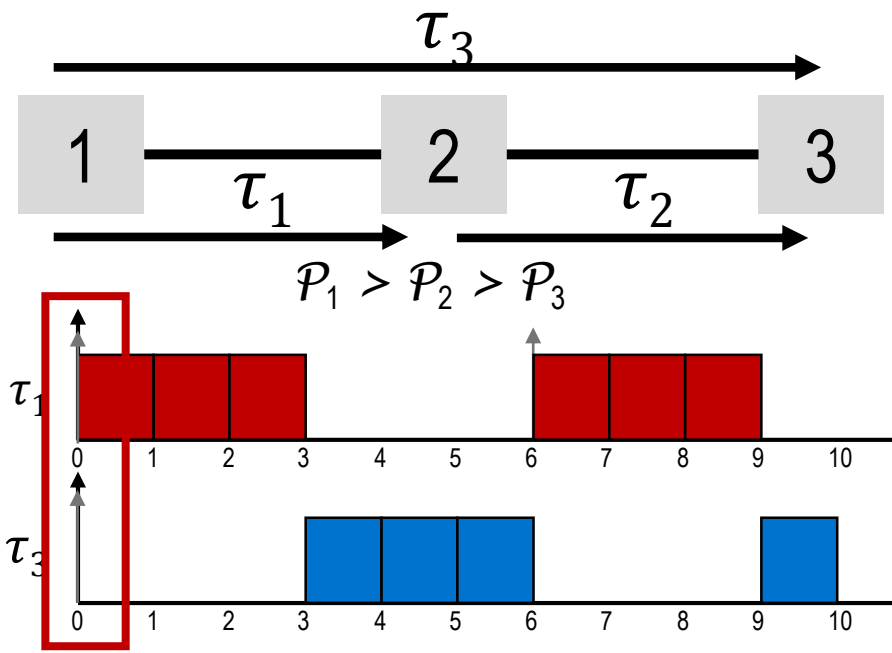
Task	# of flits	Period
τ_1	3	6
τ_2	2	6
τ_3	4	Large



H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Stage-level Analysis (SLA)



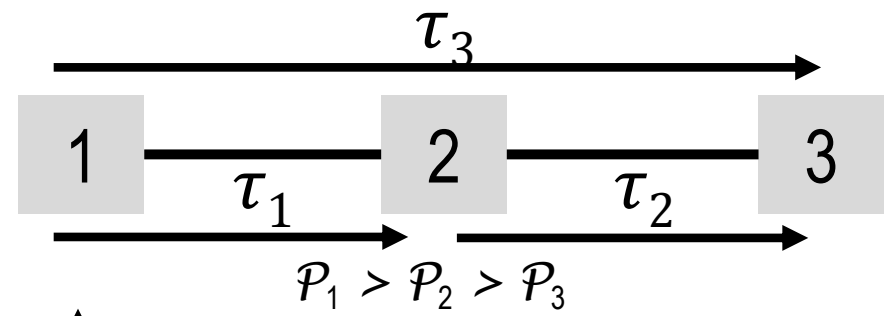
Task	# of flits	Period
τ_1	3	6
τ_2	2	6
τ_3	4	Large

Stage (1,2)

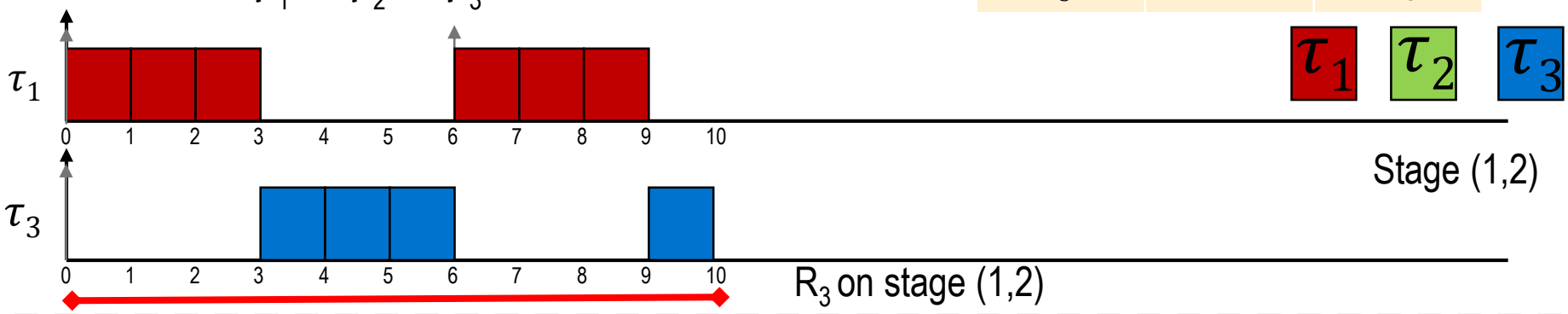
Critical instant: higher priority communication tasks released with task under analysis

H. Kashif and H. Patel (RTAS 2016)

Stage-level Analysis (SLA)



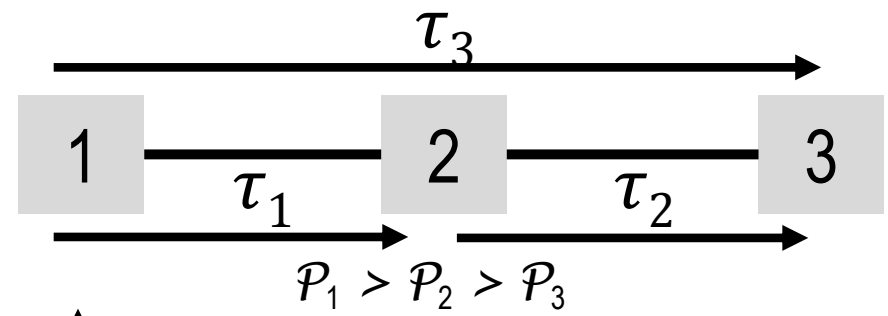
Task	# of flits	Period
τ_1	3	6
τ_2	2	6
τ_3	4	Large



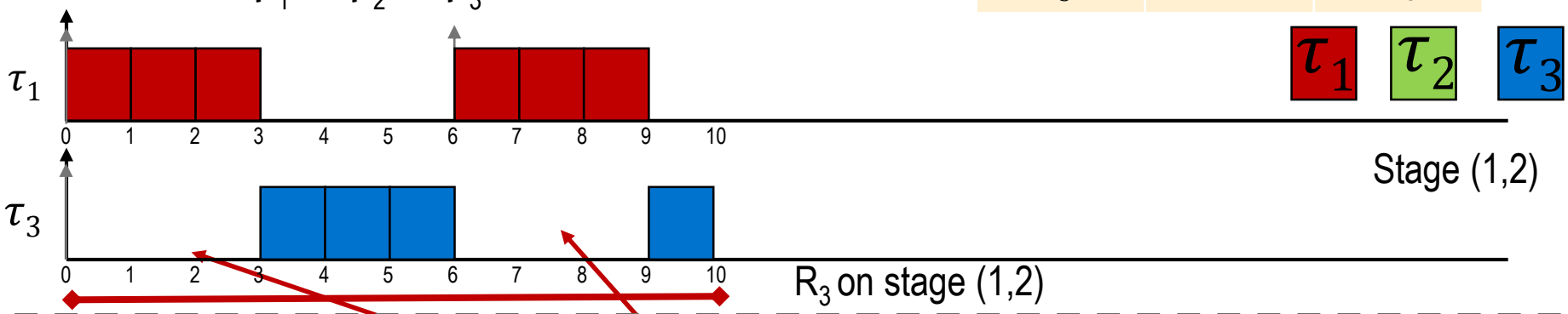
H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Stage-level Analysis (SLA)



Task	# of flits	Period
τ_1	3	6
τ_2	2	6
τ_3	4	Large

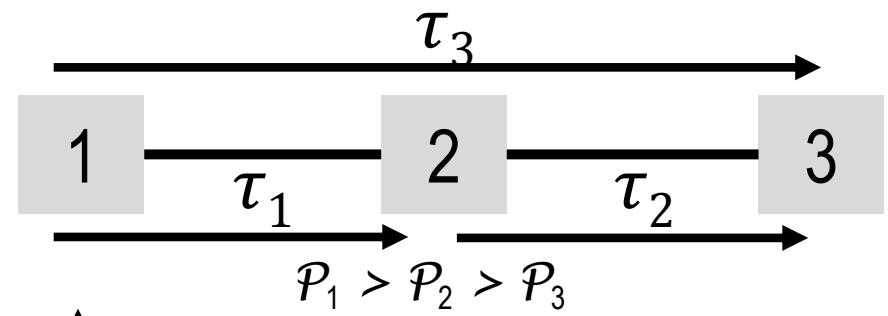


τ_1 introduces interferences (gaps) in τ_3's transmission

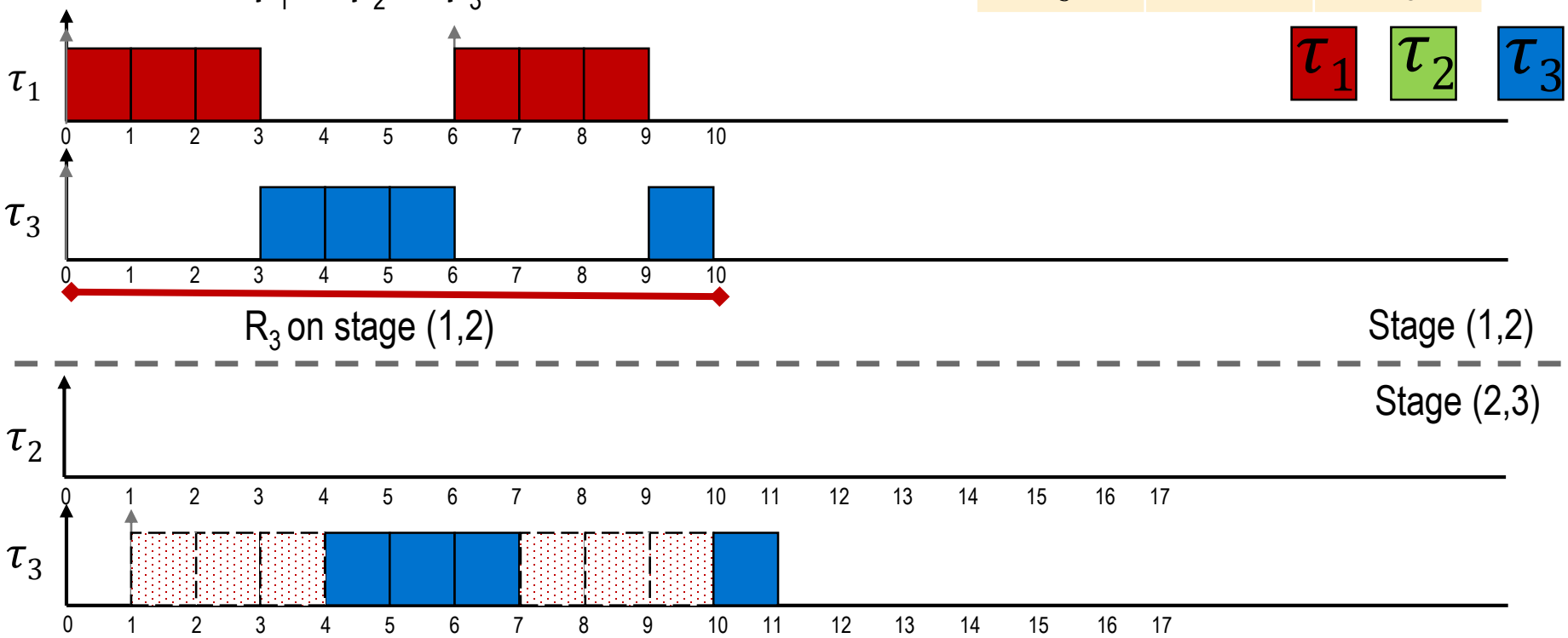
H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Stage-level Analysis (SLA)



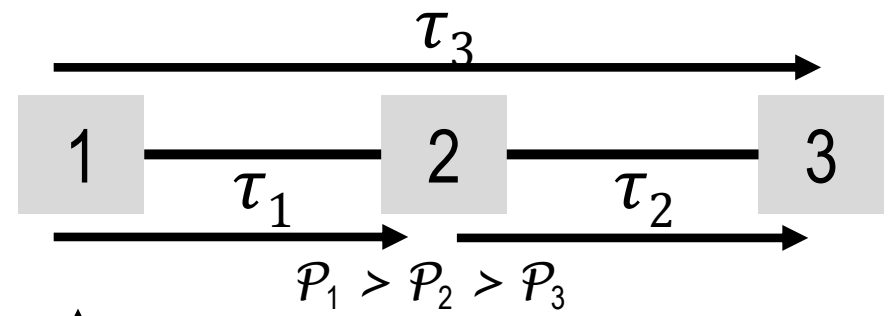
Task	# of flits	Period
τ_1	3	6
τ_2	2	6
τ_3	4	Large



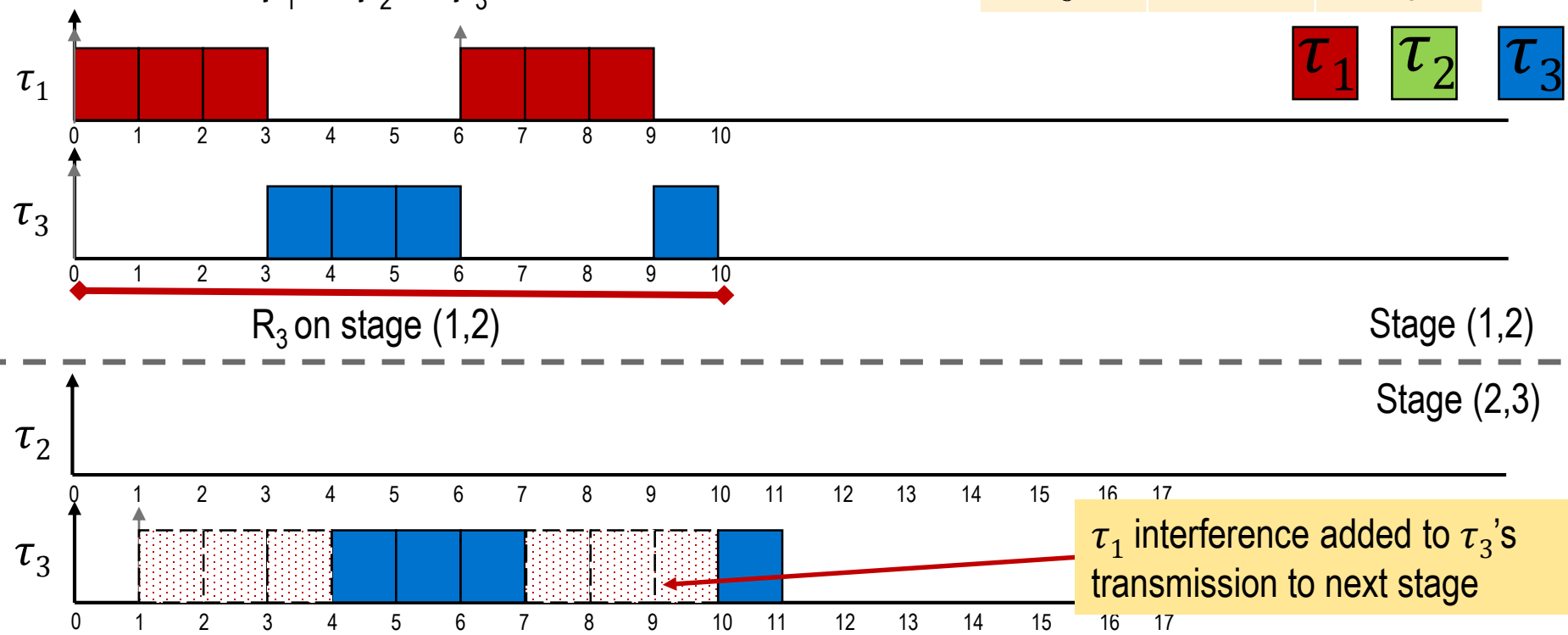
H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Stage-level Analysis (SLA)



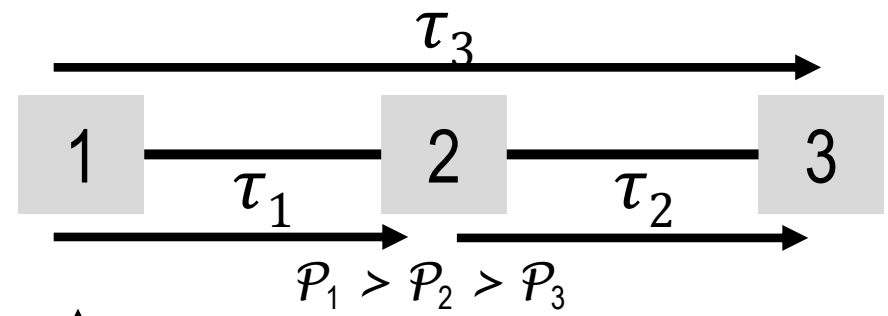
Task	# of flits	Period
τ_1	3	6
τ_2	2	6
τ_3	4	Large



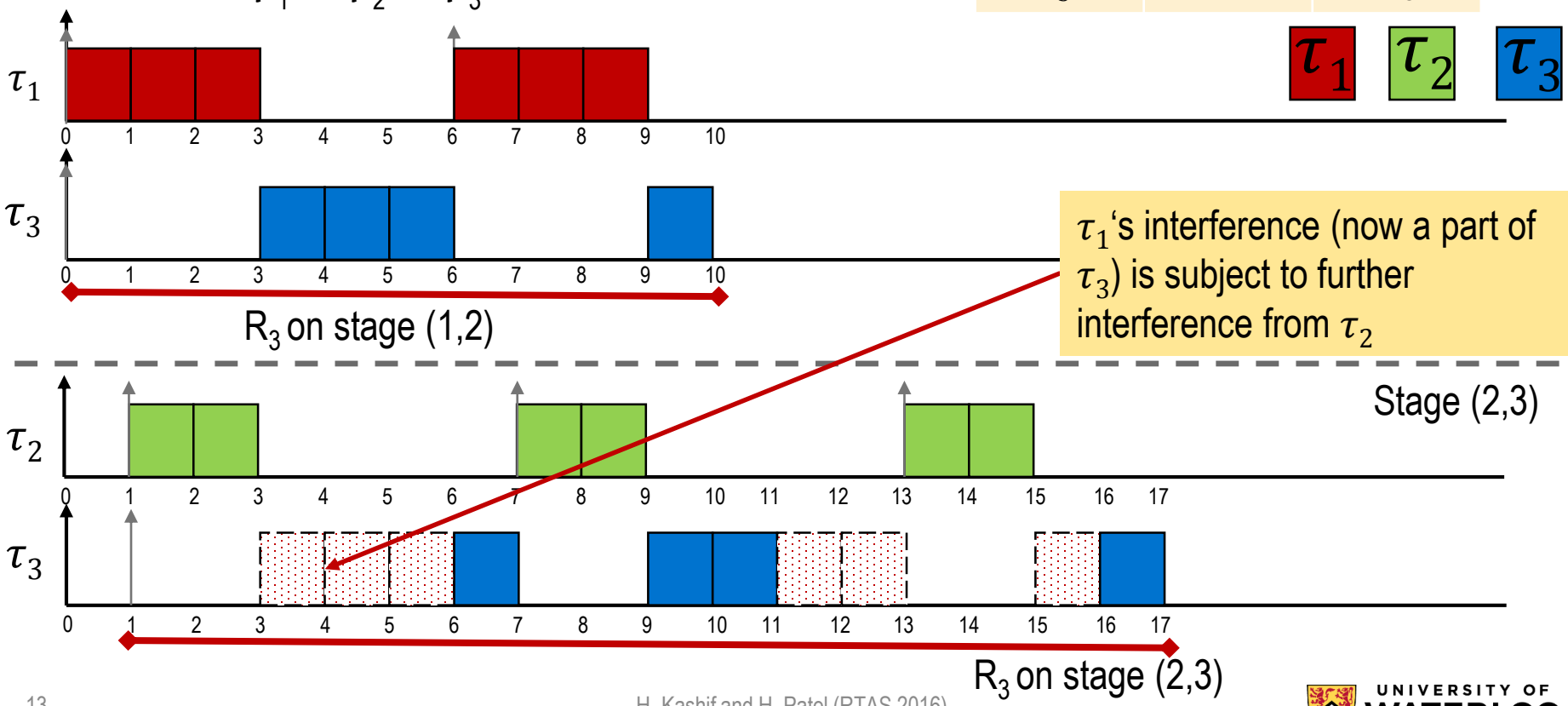
H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Stage-level Analysis (SLA)



Task	# of flits	Period
τ_1	3	6
τ_2	2	6
τ_3	4	Large

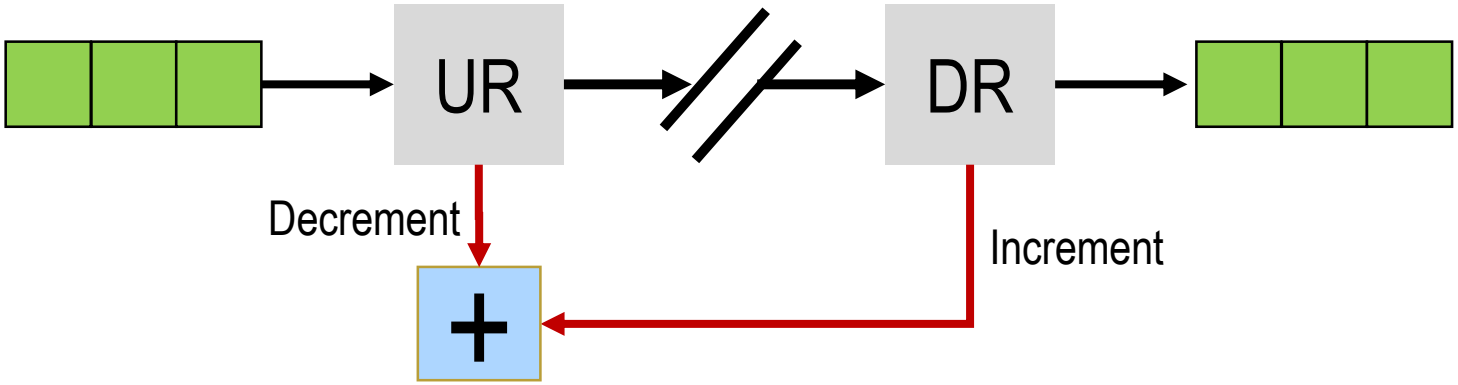


H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

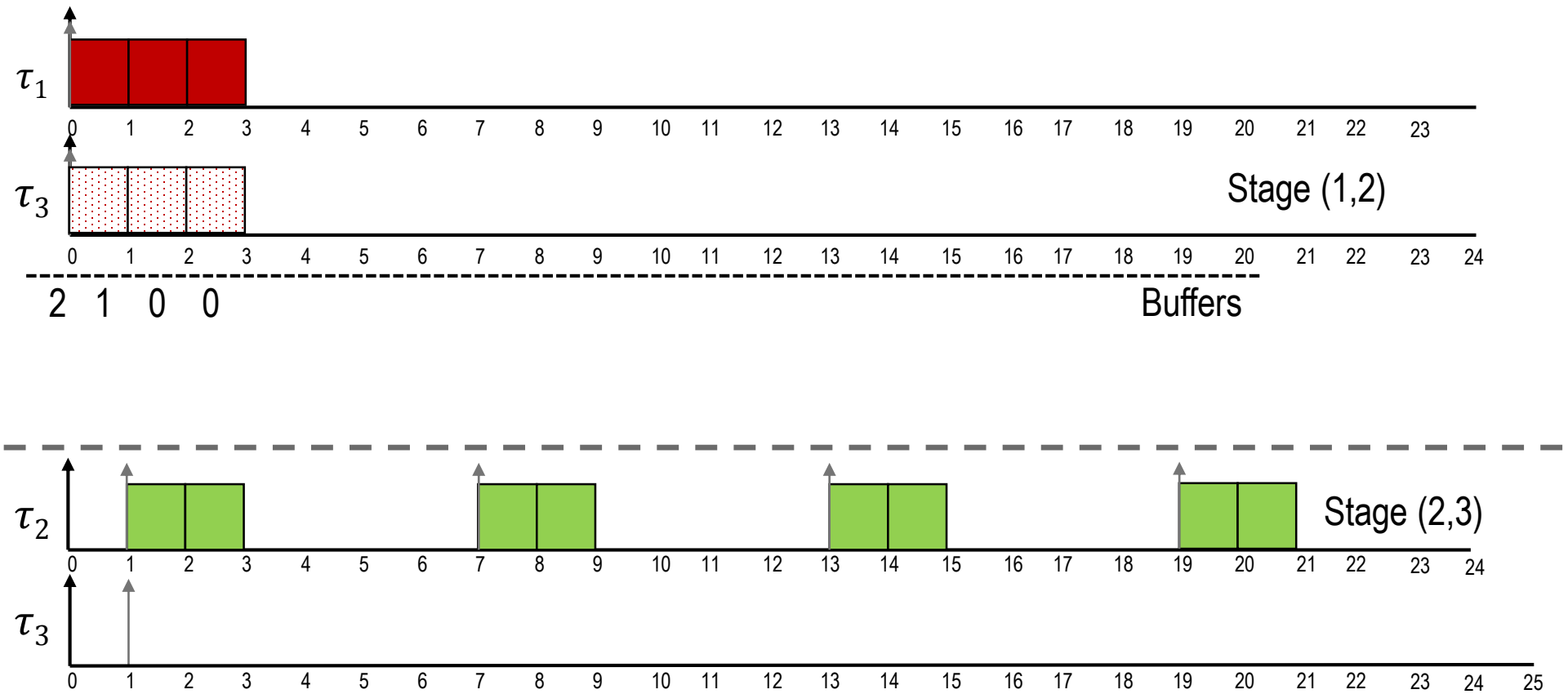
Buffers with Credit Feedback

- Finite buffers require a flow control policy
  - DR notifies UR that credits must be replenished
- We use credit feedback mechanism
  - Notification from DR to UR has a credit feedback delay



H. Kashif and H. Patel (RTAS 2016)

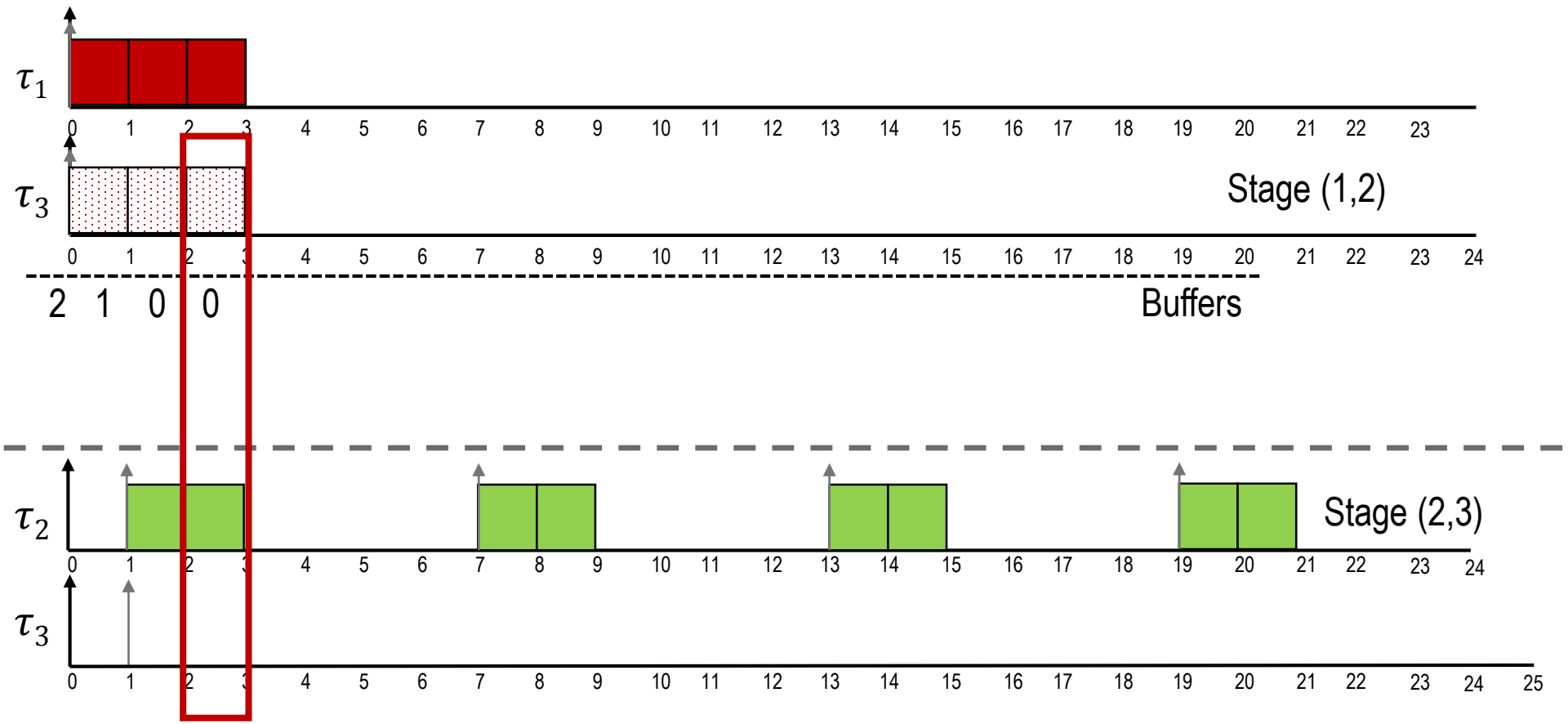
SLA with Buffer Constraints & Credit Feedback Delay



H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

SLA with Buffer Constraints & Credit Feedback Delay

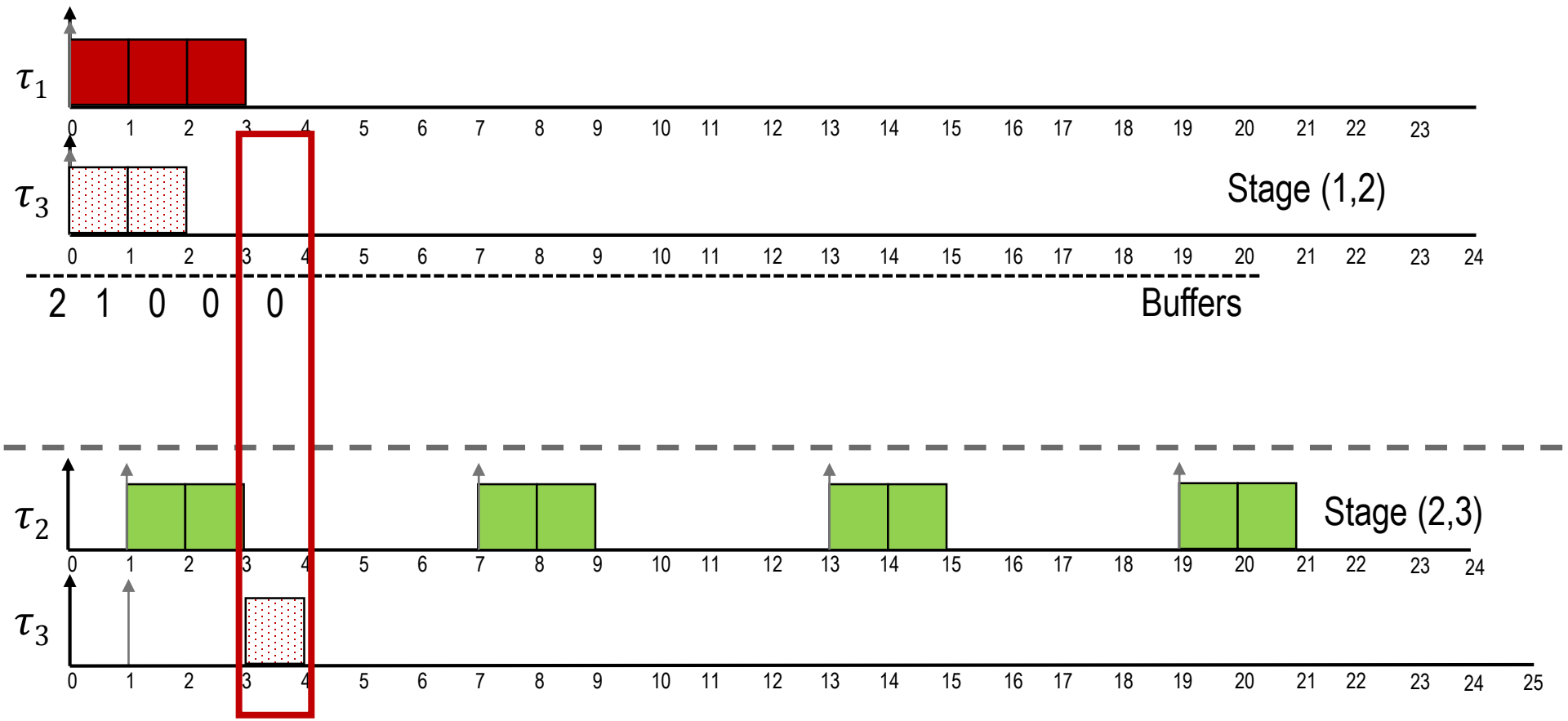


Blockage experienced on (1,2) by τ_3

H. Kashif and H. Patel (RTAS 2016)

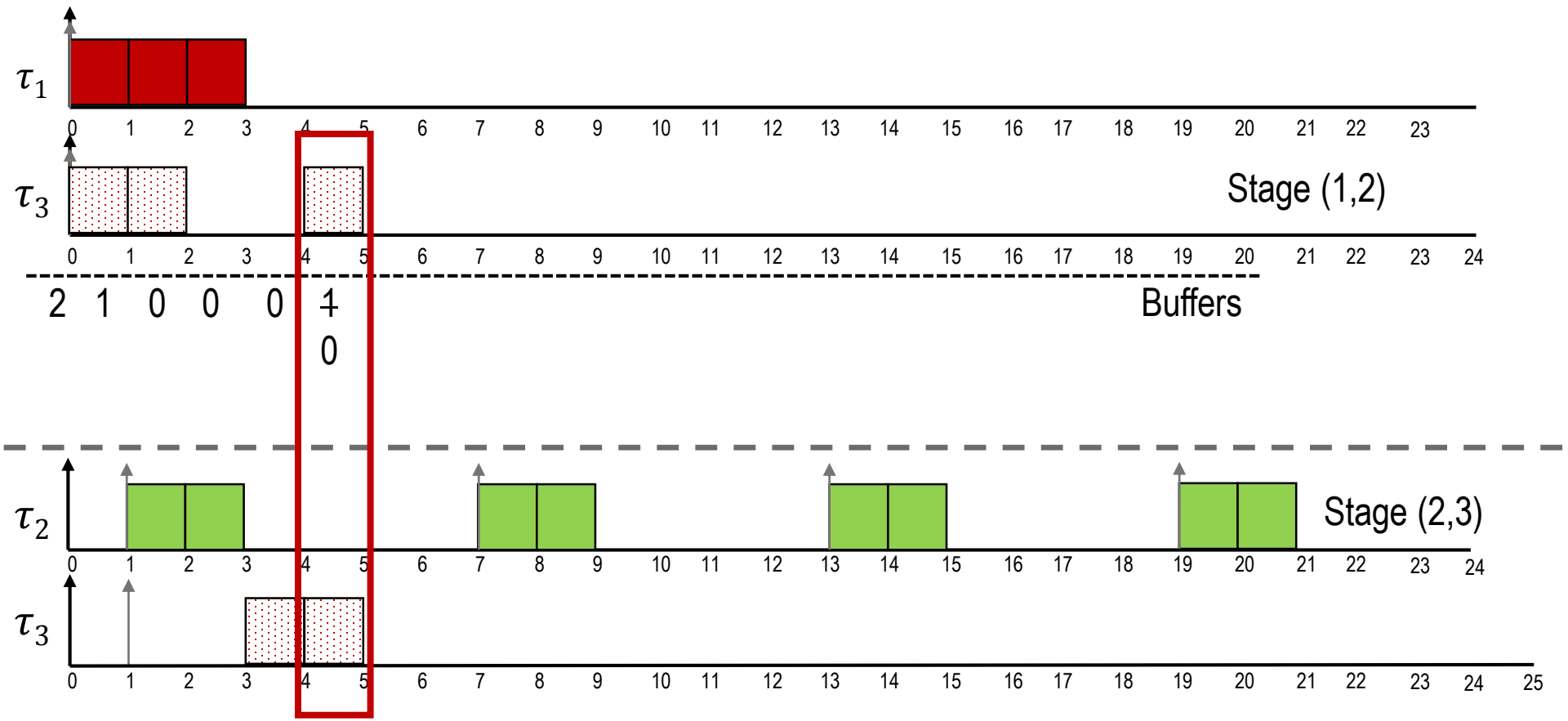
UNIVERSITY OF WATERLOO

SLA with Buffer Constraints & Credit Feedback Delay



Credit available but feedback delay of 1

SLA with Buffer Constraints & Credit Feedback Delay



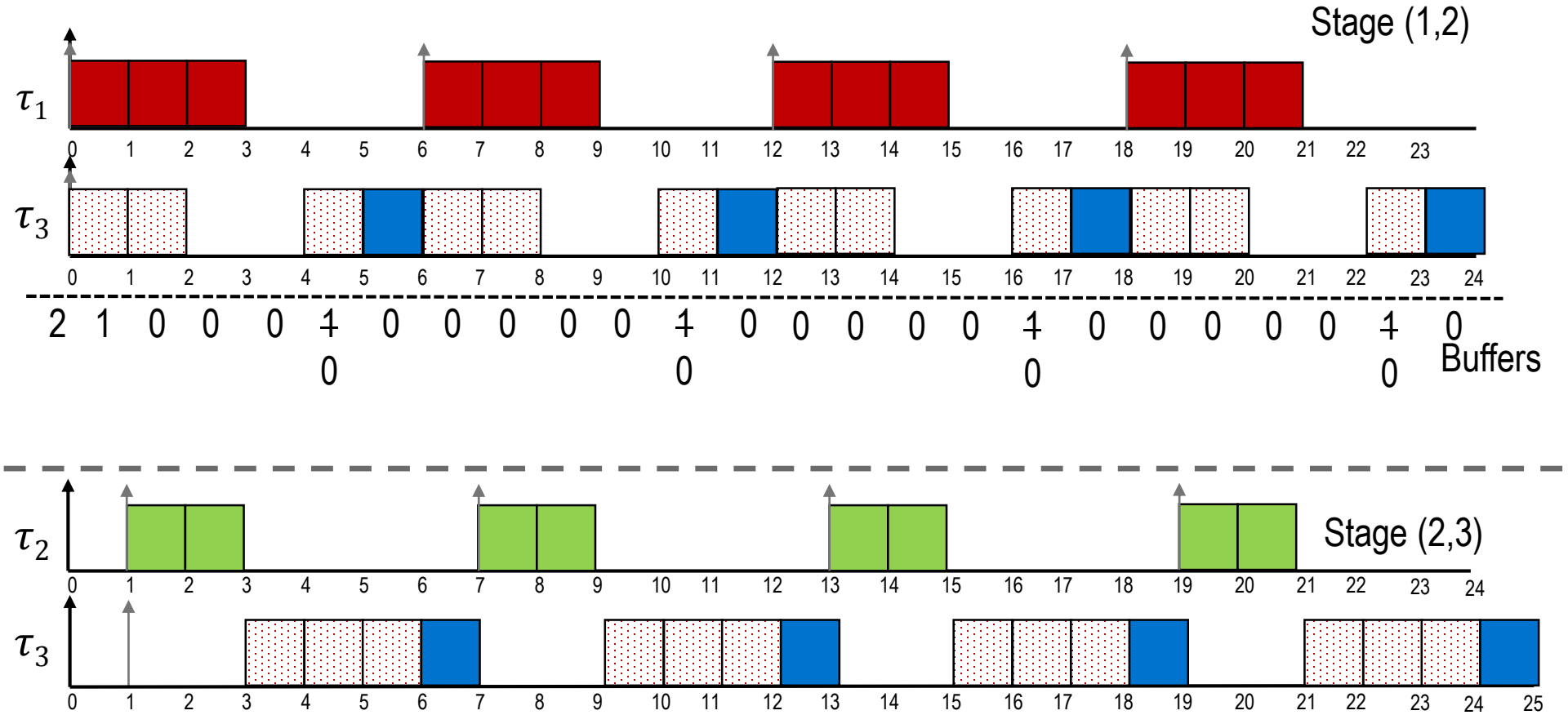
Credit acknowledged and next flit sent

Next credit sent back to stage (1,2)

H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

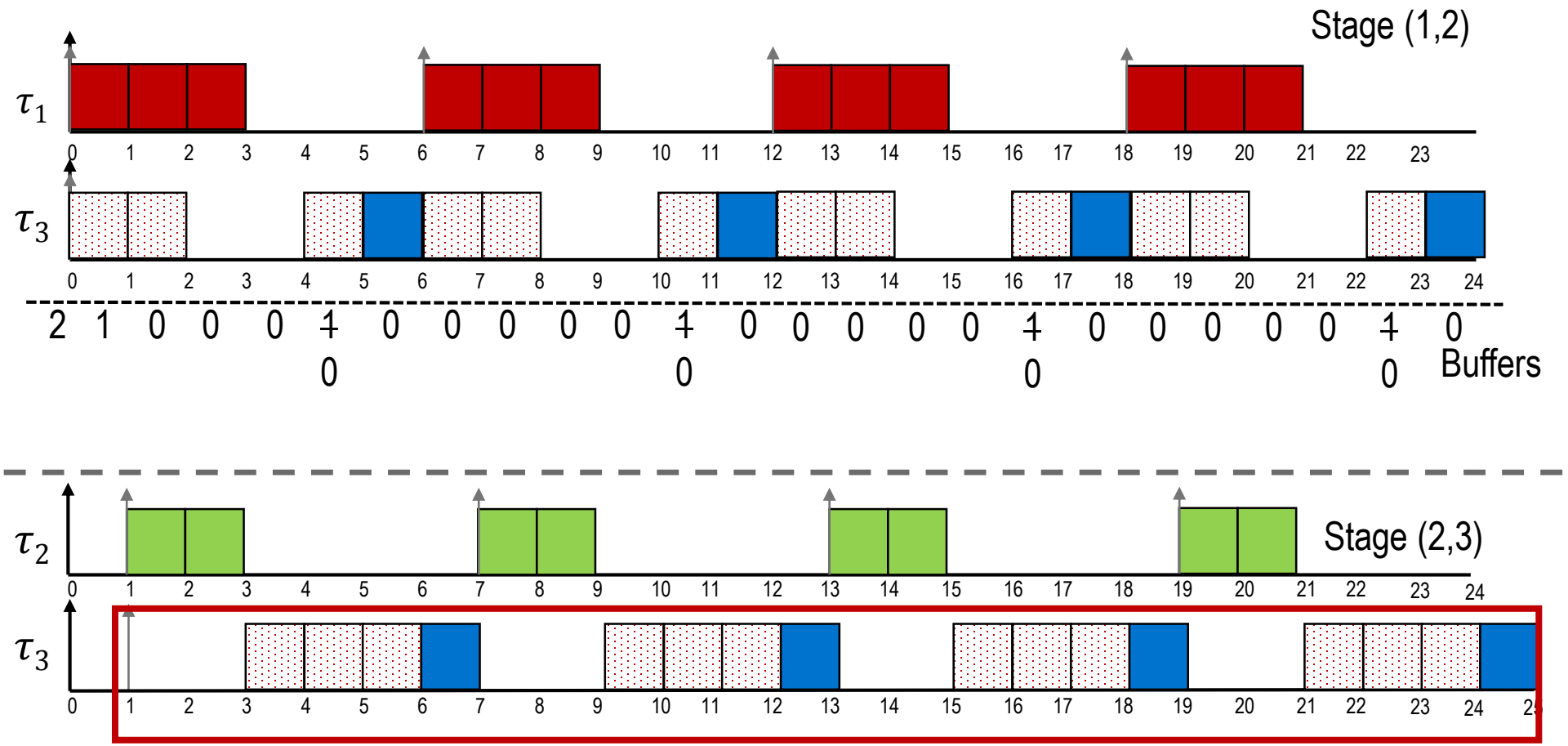
SLA with Buffer Constraints & Credit Feedback Delay



H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

SLA with Buffer Constraints & Credit Feedback Delay

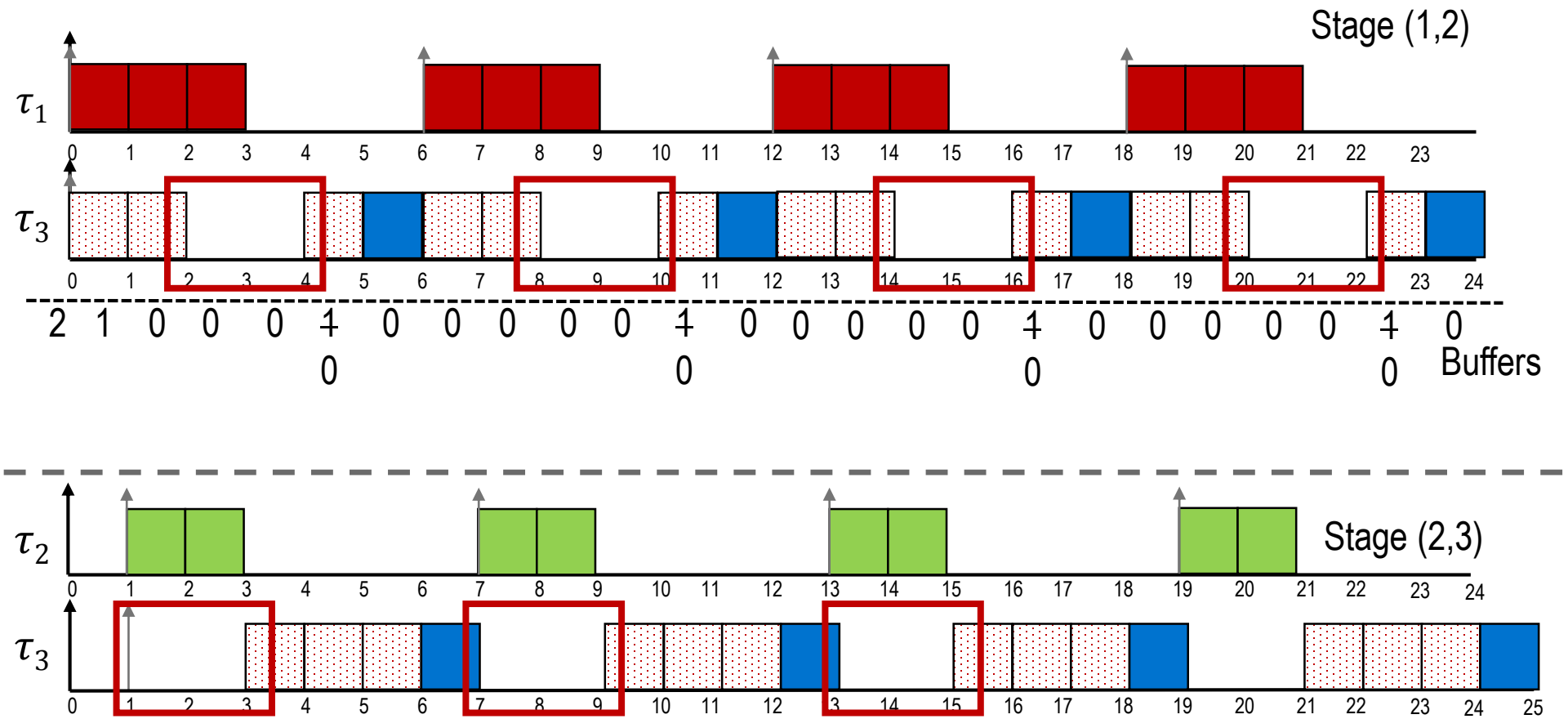


Observation 1: Since stage (2,3) is the last stage, τ_3 does not experience further backlog as flits are consumed

H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

SLA with Buffer Constraints & Credit Feedback Delay

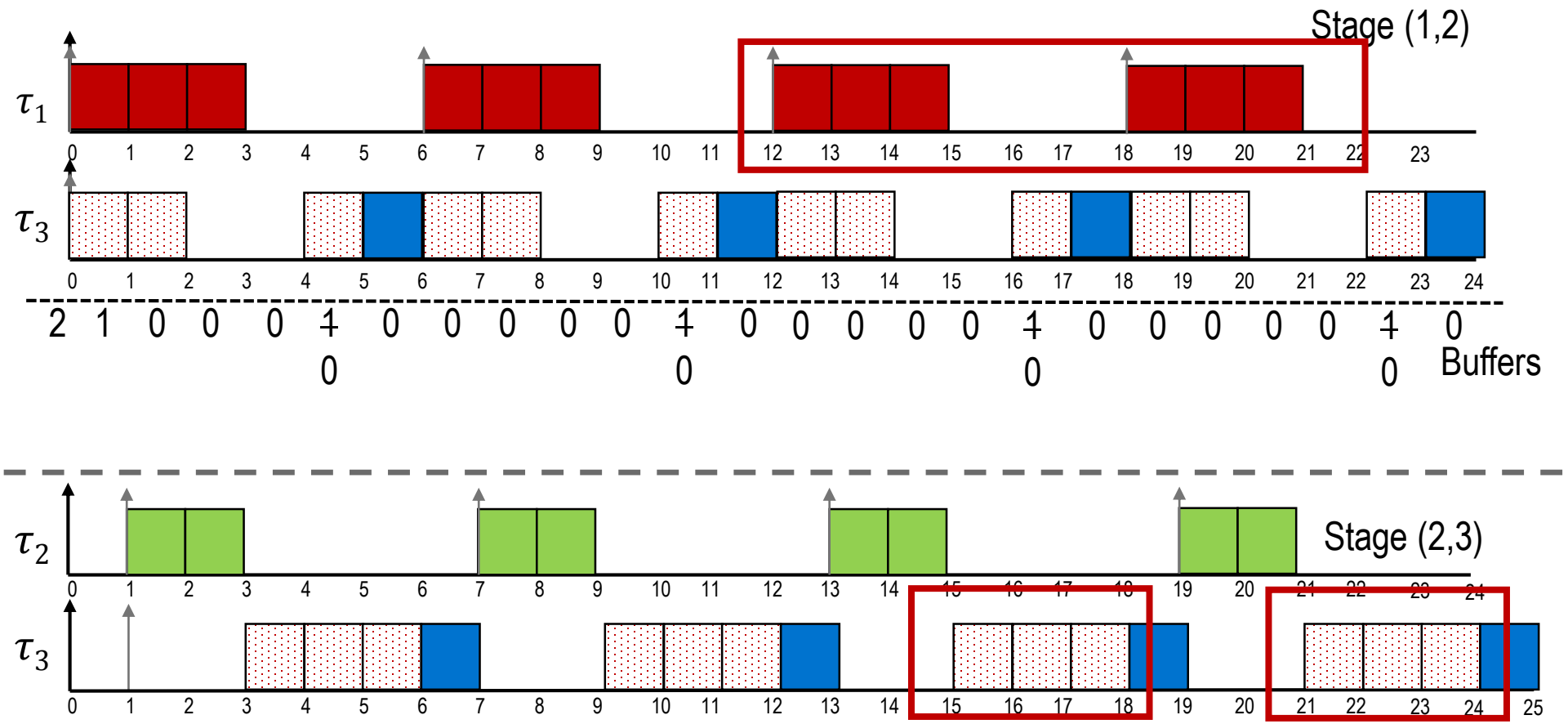


Observation 2: Amount of new interference on τ_3 on stage (2,3) contributes to blockage on (1,2): $\left\lceil \frac{R}{T} \right\rceil * L - VC$

H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

SLA with Buffer Constraints & Credit Feedback Delay

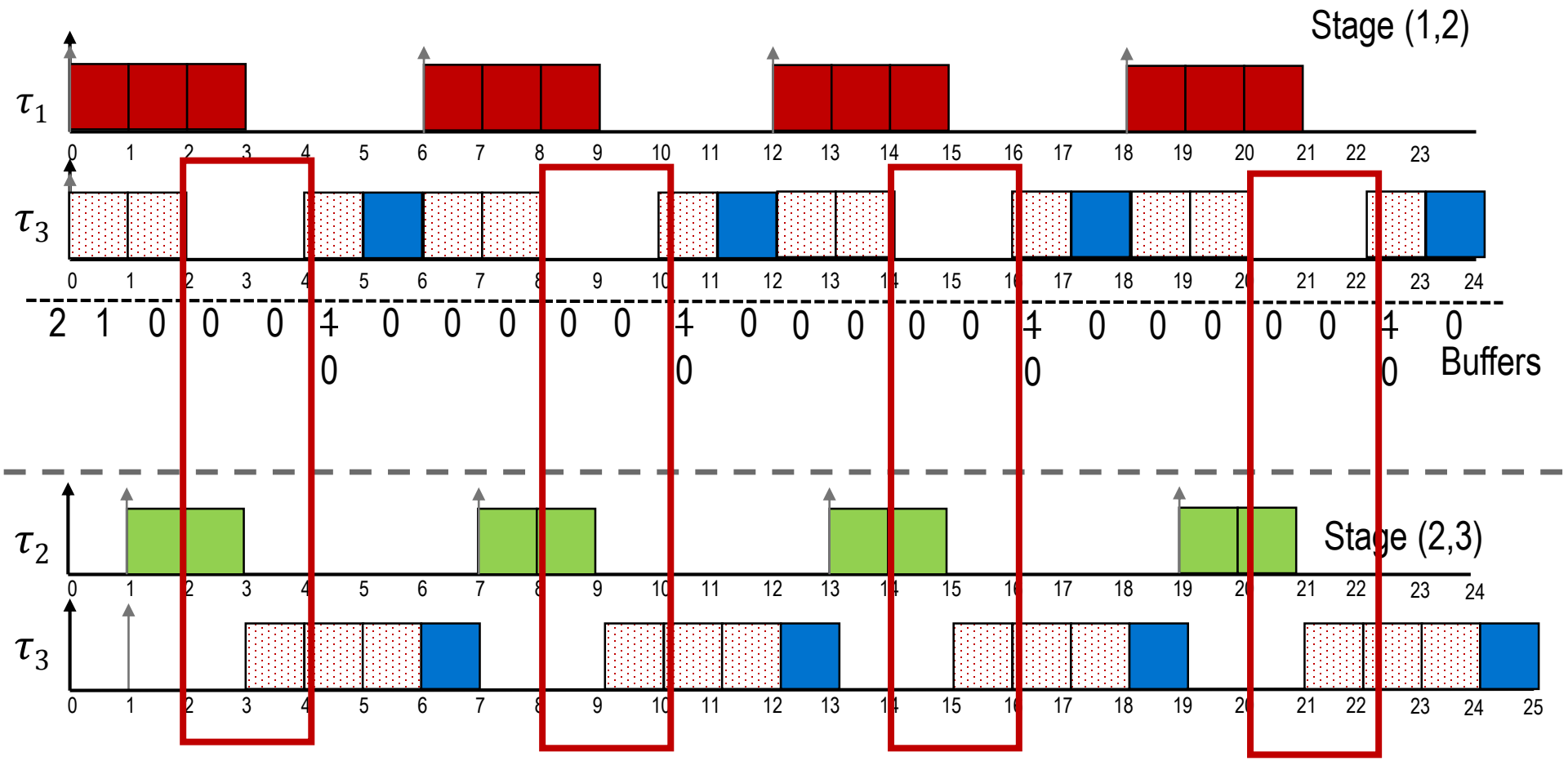


Observation 3: Blockage introduces additional interference. Carry forward this interference.

H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

SLA with Buffer Constraints & Credit Feedback Delay



Observation 4: A particular flit receives its credit CF + 1 units after transmission.

Blockage = MAX(0, IB_{S+} + $\left\lceil \frac{R}{T} \right\rceil * L - VC + CF + 1$)

H. Kashif and H. Patel (RTAS 2016)

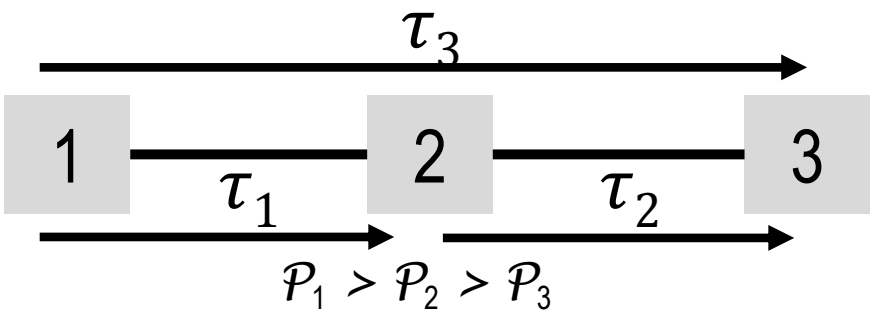
UNIVERSITY OF WATERLOO

Buffer Space Allocation

- Assume a fixed number of buffers per router
  - Allocate buffers to VCs in each router at start-up
- Objectives of allocation
  - Make communication tasks schedulable
  - Use minimal buffer space
- Key idea
  - Allocate more buffer space to VC of router on stage that suffers most interference

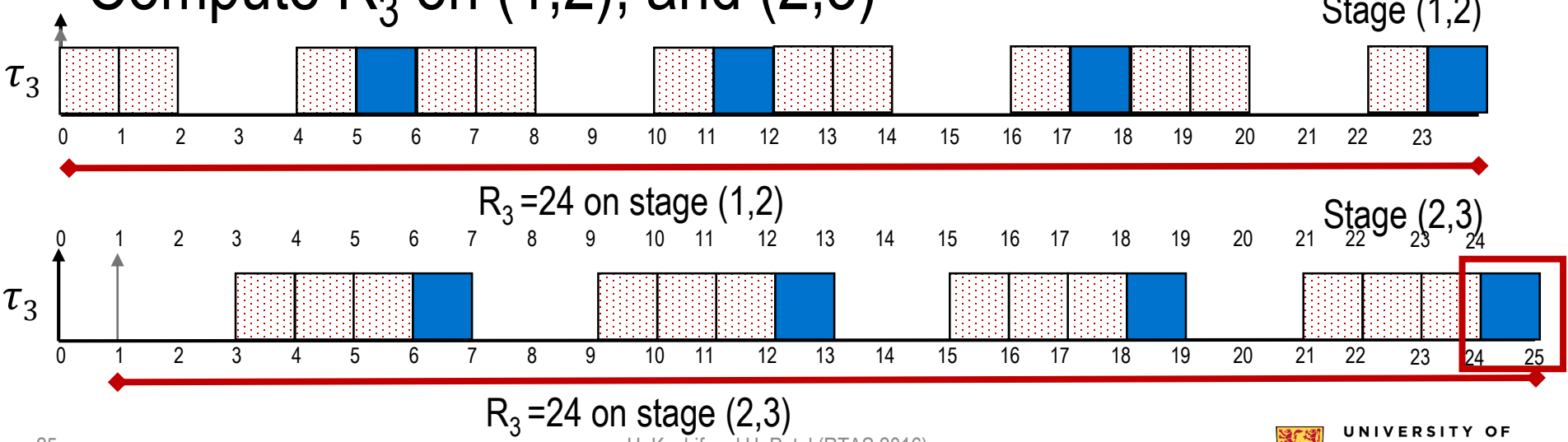
H. Kashif and H. Patel (RTAS 2016)

Buffer Space Allocation



Task	# of flits	Period	Deadline
τ_1	3	6	--
τ_2	2	6	--
τ_3	4	Large	23

- Initialize VCs to CF + 1 (Assuming CF = 1)
- Compute R_3 on (1,2), and (2,3)

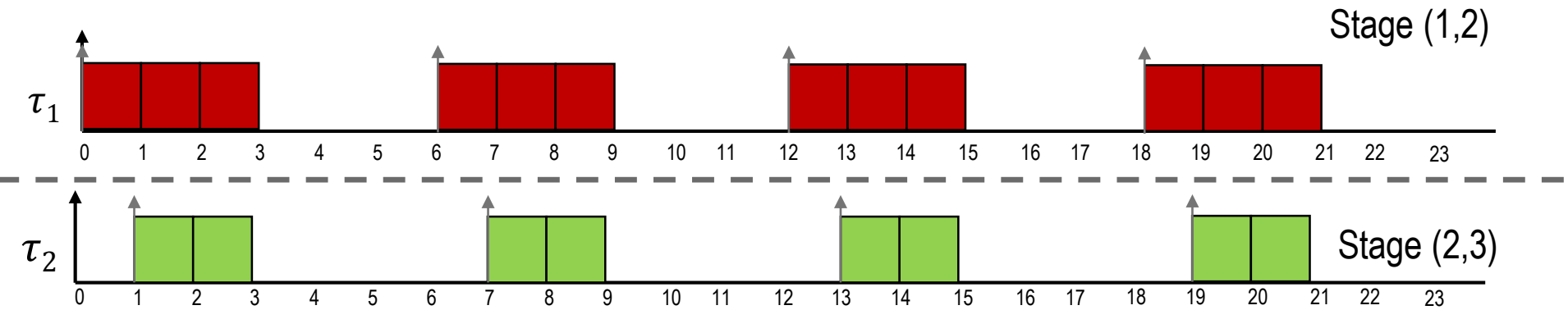


H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Buffer Space Allocation

- Compute total R_3 over path, $R_3 = 25$
- $R_3 > 23$
  - Must allocate more space to a router's VC
- Select router to increment buffer space: (1,2) or (2,3)
  - Must have available buffer space
  - Must contribute most to interference (no blockage)

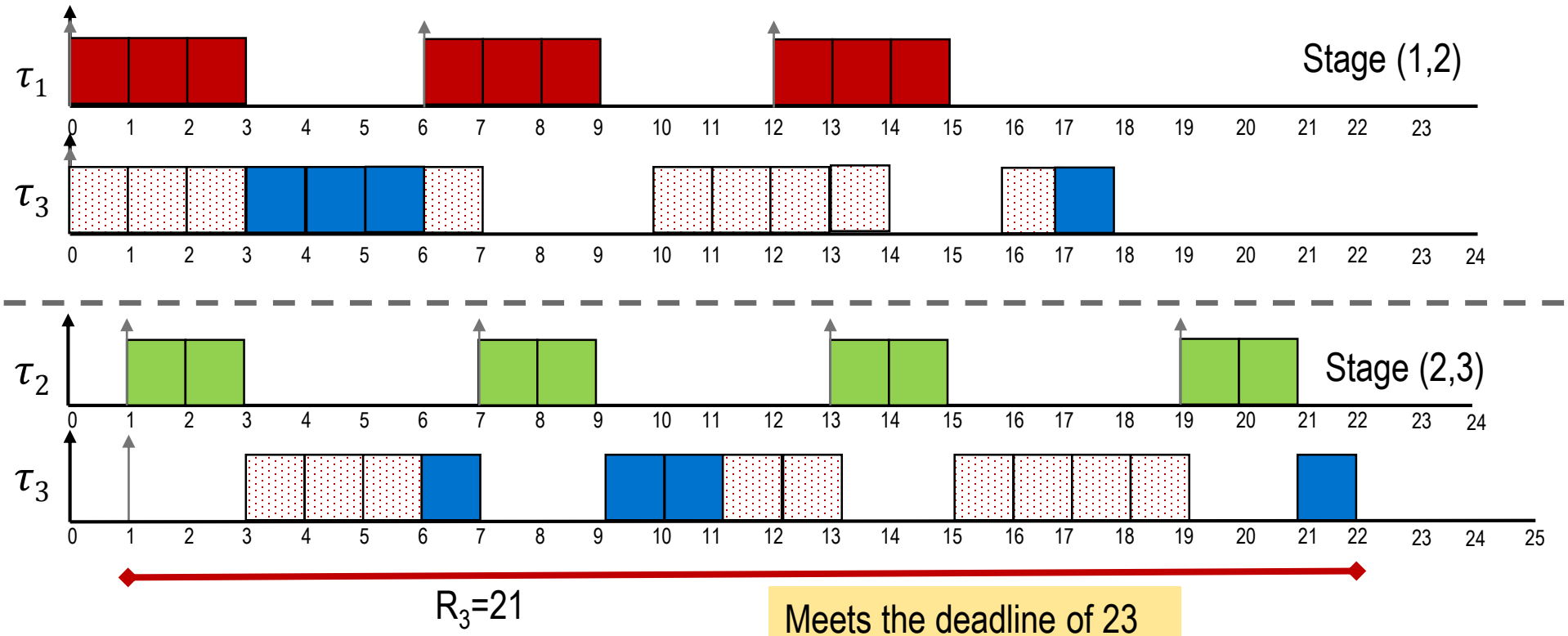


H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Buffer Space Allocation

- Increment VC on stage (2,3) to 3
- Re-compute R_3 on (1,2) and (2,3)

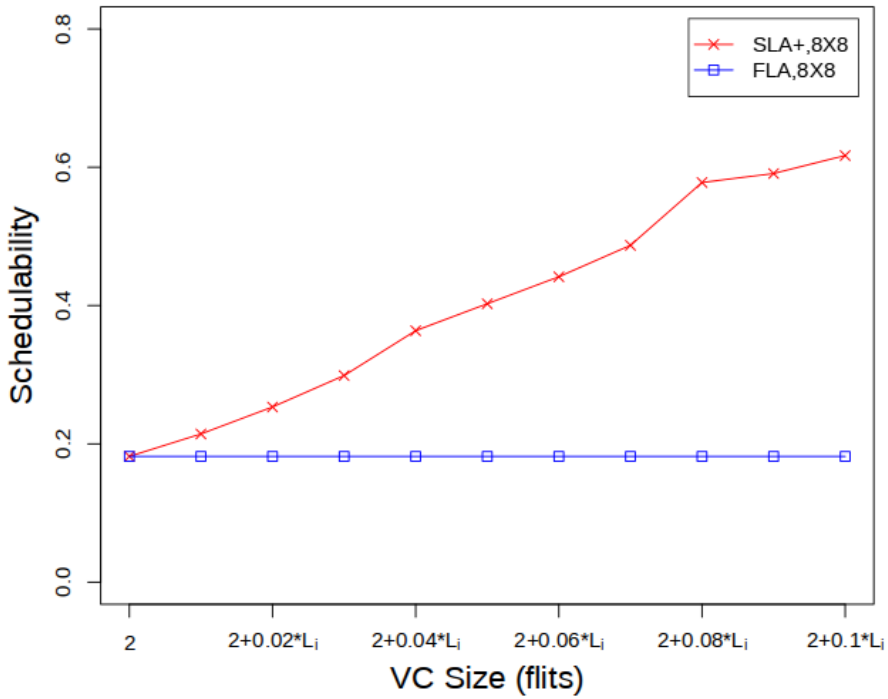


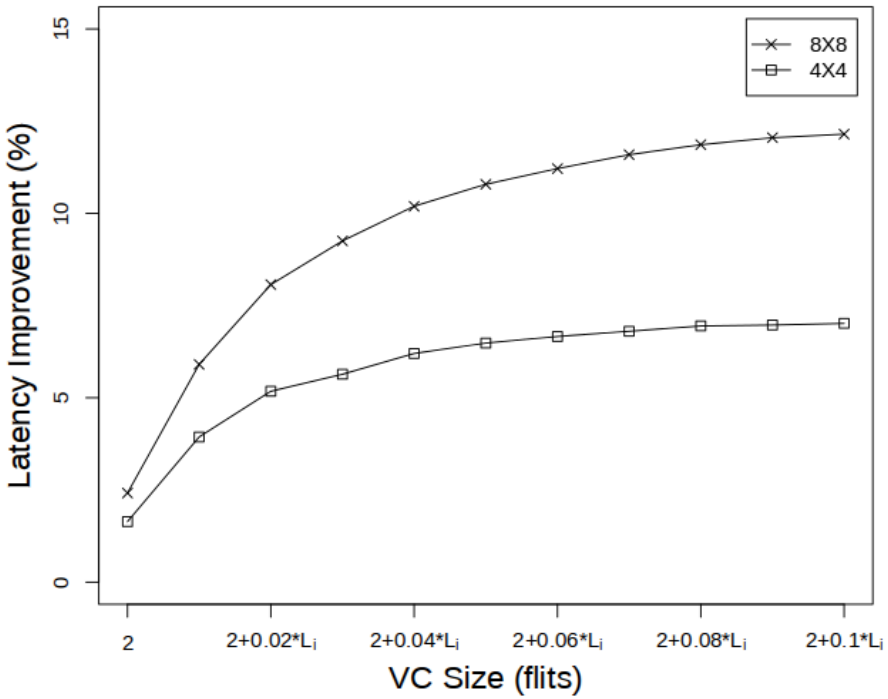
H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Experimental Evaluation

- As more buffer space is available SLA with buffer constraints offers more schedulable configurations



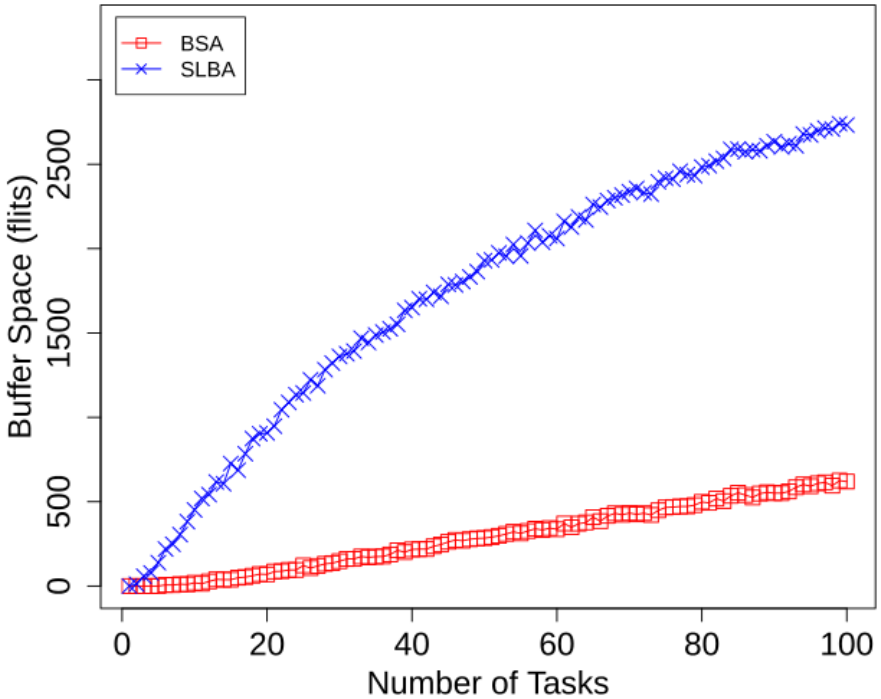


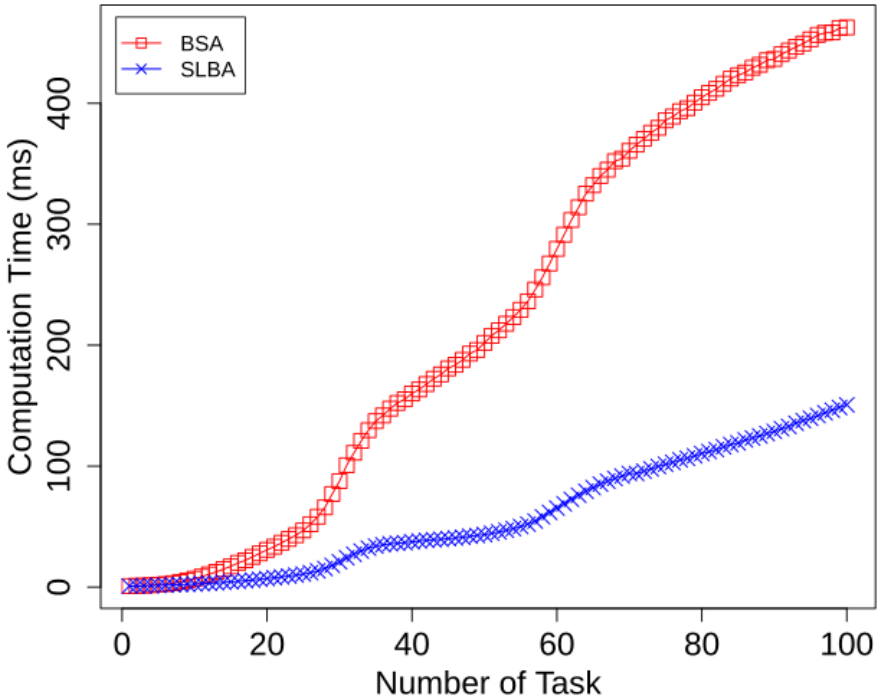
H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Experimental Evaluation

- For schedulable configurations
  - BSA uses less buffer space than SLA (~85%)
  - But, computation time is higher (~4x)





H. Kashif and H. Patel (RTAS 2016)

UNIVERSITY OF WATERLOO

Conclusions

- Extended SLA
  - Buffer constraints and credit feedback delay
- Buffer space allocation
  - Reduces required buffer space
- Future work
  - Hardware implementation of a PAR NoC

END

- ☺