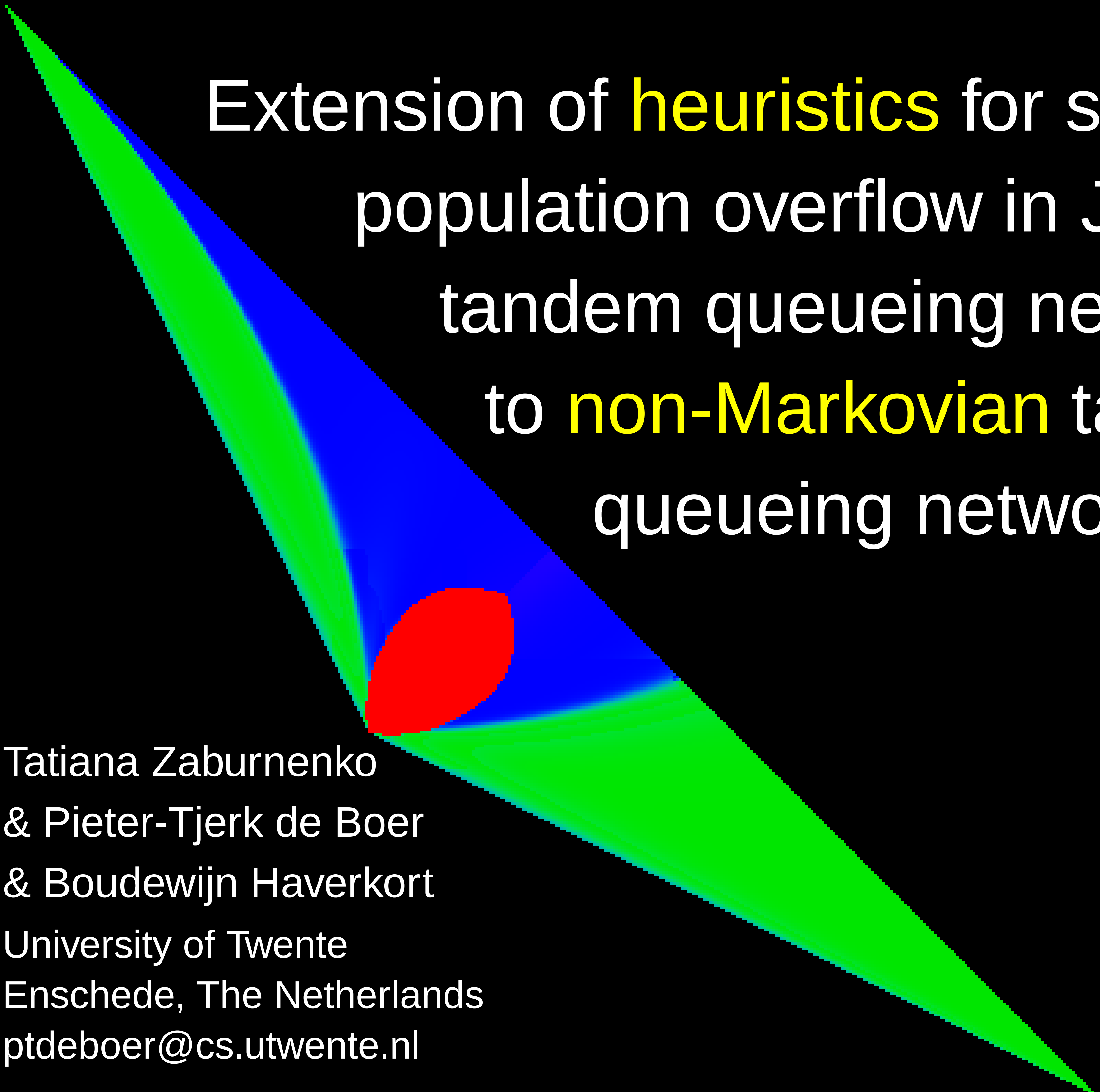
Extension of heuristics for simulating population overflow in Jackson tandem queueing networks to non-Markovian tandem queueing networks



Tatiana Zaburnenko
& Pieter-Tjerk de Boer
& Boudewijn Haverkort
University of Twente
Enschede, The Netherlands
ptdeboer@cs.utwente.nl

Background

Recent progress for importance sampling for Markovian tandem queues:

- State-independent c.o.m. is not sufficient (Glasserman&Kou 1995, De Boer 2004)
- Adaptive state-dependent c.o.m. using cross entropy (De Boer & Nicola 2002)
- Linear state-dependent heuristics (Zaburnenko & Nicola 2005)
- Provably efficient c.o.m. based on game theory (Dupuis & Sezer & Wang 2007)

But what about the *non*-Markovian case?

State-dependent importance sampling for non-Markovian systems

Should c.o.m. depend only on queue lengths (discrete) or also on ages (continuous)?

Re-schedule future events to take into account state change since they were scheduled?

How to determine the c.o.m.?

State-dependent importance sampling for non-Markovian systems

Should c.o.m. depend only on queue lengths (discrete) or also on ages (continuous)?
dependence on only queue lengths is simpler and seems sufficient

Re-schedule future events to take into account state change since they were scheduled?

How to determine the c.o.m.?

State-dependent importance sampling for non-Markovian systems

Should c.o.m. depend only on queue lengths (discrete) or also on ages (continuous)?
dependence on only queue lengths is simpler and seems sufficient

Re-schedule future events to take into account state change since they were scheduled?
rescheduling is necessary for good speedup

How to determine the c.o.m.?

State-dependent importance sampling for non-Markovian systems

Should c.o.m. depend only on queue lengths (discrete) or also on ages (continuous)?
dependence on only queue lengths is simpler and seems sufficient

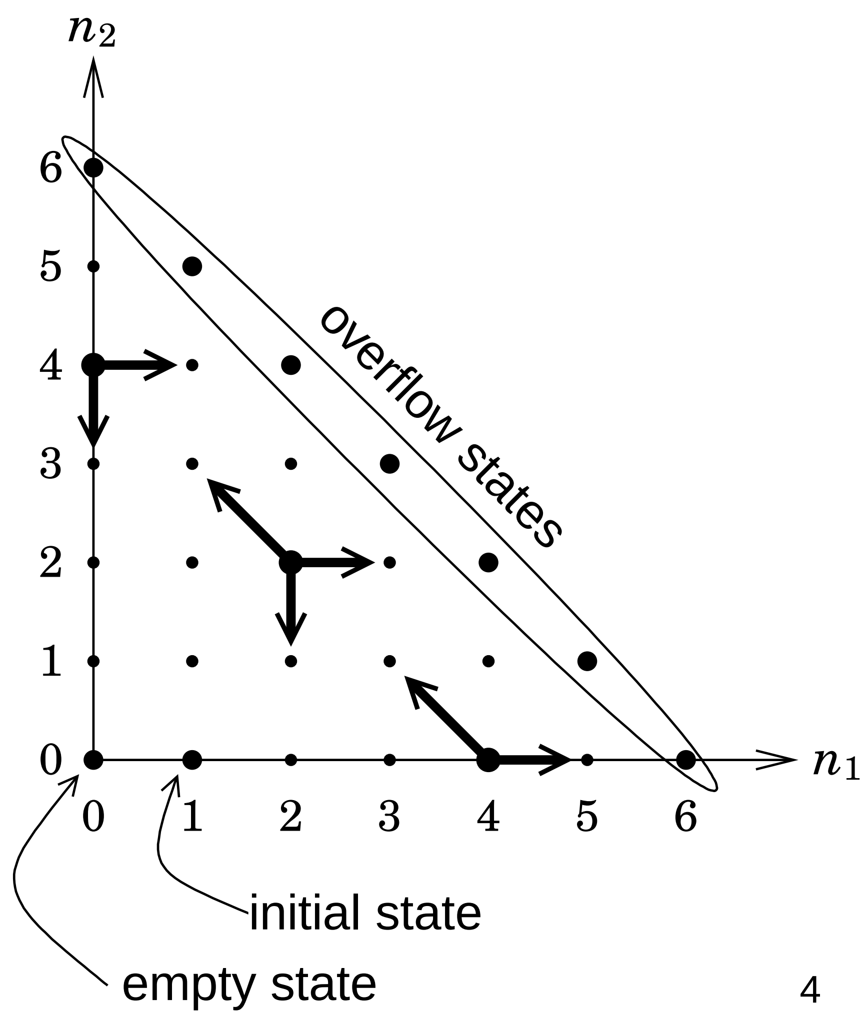
Re-schedule future events to take into account state change since they were scheduled?
rescheduling is necessary for good speedup

How to determine the c.o.m.?
in this work: heuristic

Heuristic for Markovian tandem queue

Idea: (assuming second queue is bottleneck)

- Parekh&Walrand heuristic: interchange arrival and second service rate
  $\rightarrow$ problems caused at state space boundary where second queue is empty.
- Near this boundary, interchanging arrival and *first* service rate is better.
- Do a smooth (linear) transition between those two.
- Possible improvement: also linearly turn off this c.o.m. near the other boundary.



Mathematically...:

$$COM_{x_1,x_2} =$$

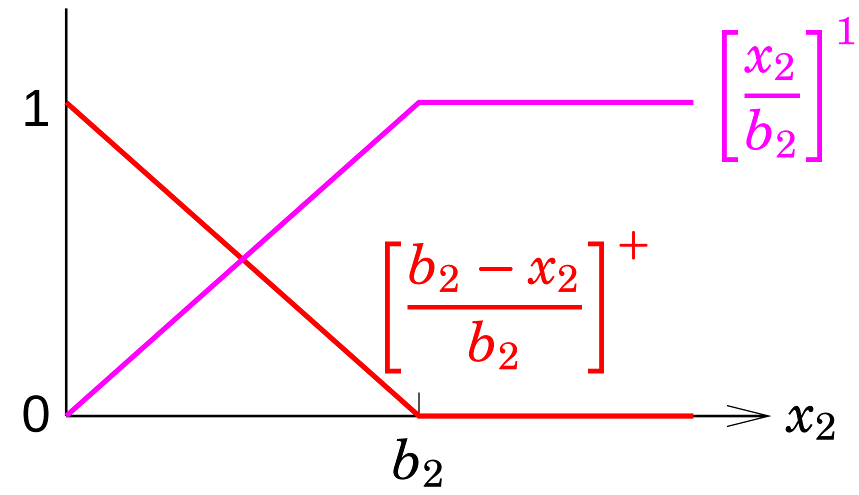
$$\left[\frac{x_2}{b_2}\right]^1 COM_2 + \left[\frac{b_2 - x_2}{b_2}\right]^+ \left(\left[\frac{x_1}{b_1}\right]^1 COM_1 + \left[\frac{b_1 - x_1}{b_1}\right]^+ COM_0 \right)$$

where:

COM_0 is the original measure
COM_1 pushes the first queue
COM_2 pushes the second queue

$[a]^+ = \max(a, 0)$
$[a]^1 = \min(a, 1)$.

x_i is the length of queue i
b_i are boundary layer widths, to be chosen



How to tilt non-Markovian queues?

For estimating overflow probabiliy of a *single* queue, exponential tilting of the interarrival and service time distributions is optimal (Parekh& Walrand 1989; Sadowsky 1991).

Exponential tilting means:

$$d\tilde{F}(x) = \frac{e^{\theta x} dF(x)}{M(\theta)}$$

where $F(x)$ is the original distribution and $M(\theta) = \mathbb{E}e^{\theta x}$.

Optimal θ solves:

$$M_{\text{arrival}}(-\theta) \cdot M_{\text{service}}(\theta) = 1.$$

Heuristic for the non-Markovian case

$$COM_{x_1,x_2} =$$

$$\left[\frac{x_2}{b_2}\right]^1 COM_2 + \left[\frac{b_2 - x_2}{b_2}\right]^+ \left(\left[\frac{x_1}{b_1}\right]^1 COM_1 + \left[\frac{b_1 - x_1}{b_1}\right]^+ COM_0 \right)$$

Let COM_i represent the exponential tilting for pushing node i.

Two possibilities:

- COM_i are vectors containing exponential tilting parameters;
- COM_i are vectors containing the distributions' "natural" parameters.

(note that these are identical in the Markovian case)

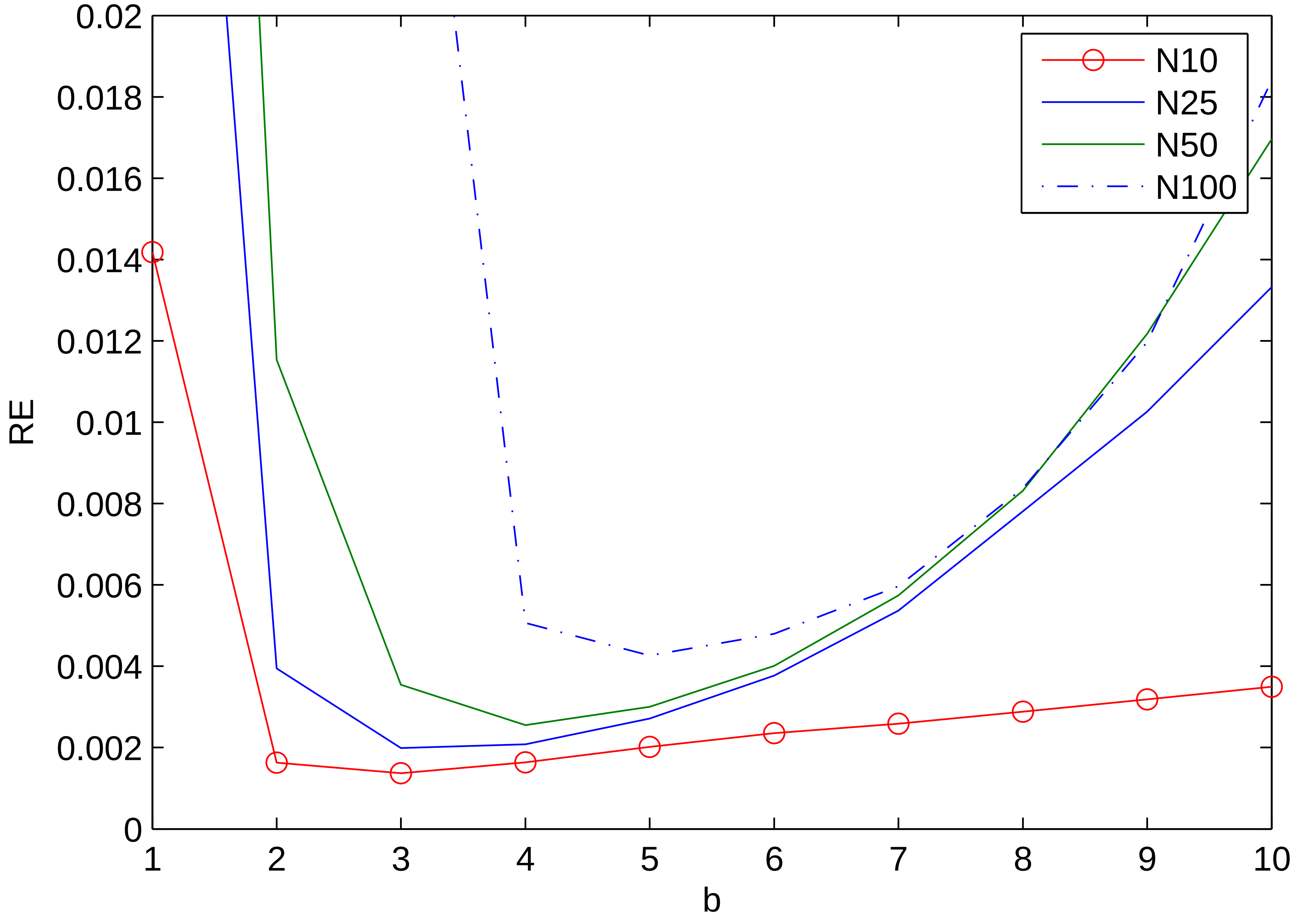
Preliminary experiments

- two queues in tandem
- arrival process: hyper-exponential interarrival times, rate 4000 with probability 0.9, rate 200 otherwise.
- service processes: bimodal, packet length is 12000 bits (“data”) with probability 0.7, 400 bits (“ack”) otherwise.
- first server: $9.01 \cdot 10^7$ bits/sec
- second server: $9.00 \cdot 10^7$ bits/sec

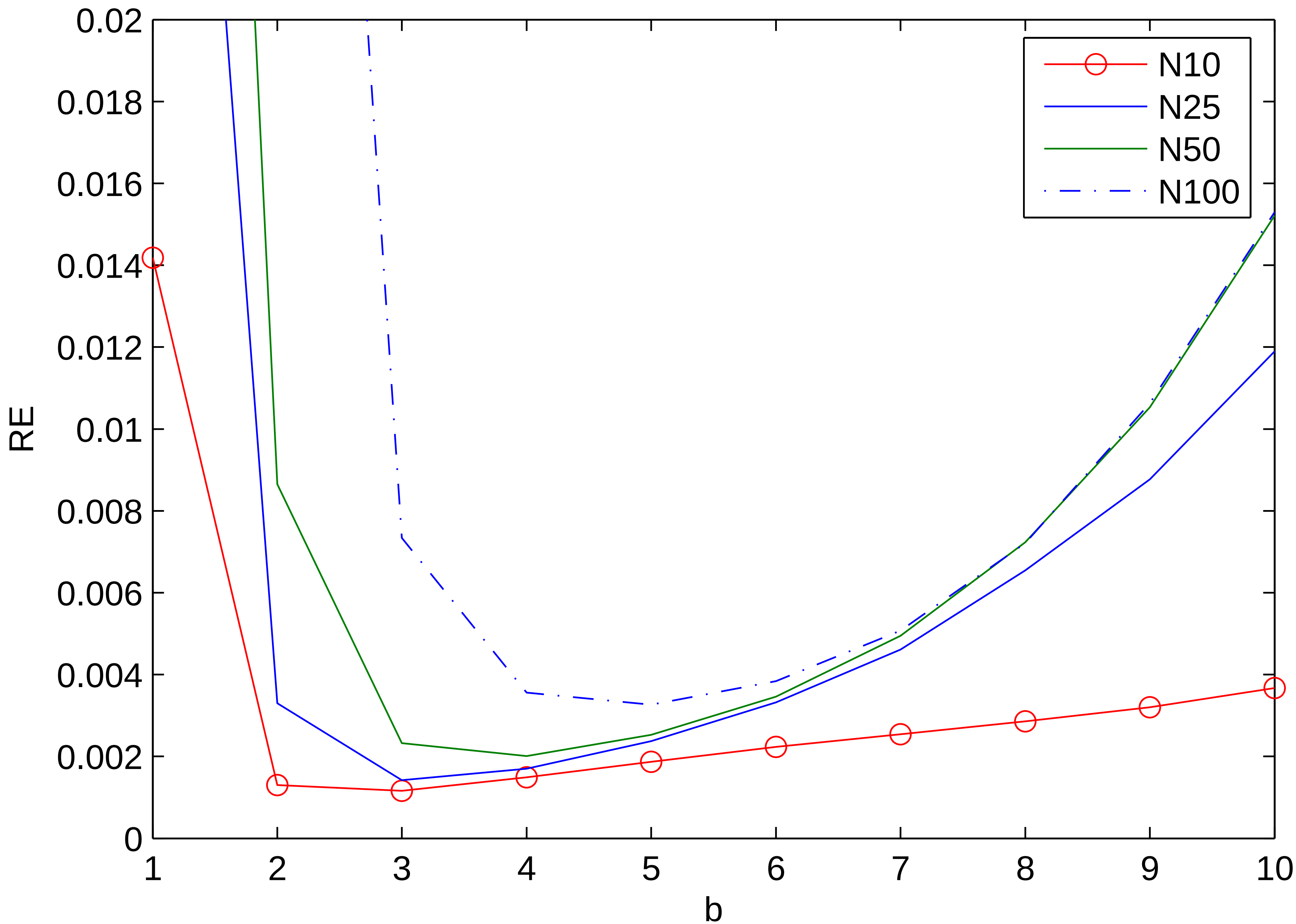
Results, asymptotic efficiency

N	linear in parameters		linear in θ	
	b_{opt}	estimate ± rel.error %	b_{opt}	estimate ± rel.error %
10	3	1.3068e-05 ± 0.12	3	1.30985e-05 ± 0.14
25	3	8.1645e-16 ± 0.14	3	8.11988e-16 ± 0.20
50	4	3.8084e-33 ± 0.20	4	3.81196e-33 ± 0.25
100	5	4.4188e-68 ± 0.32	5	4.45316e-68 ± 0.43

Results, linear-in-theta



Results, linear-in-parameters



Conclusions

- Generalization of linearly state-dependent heuristic to non-Markovian networks is feasible.
- Preliminary experimental results are promising, showing asymptotic efficiency.
- More experiments needed for solid conclusions.