Fuzzy Flow Regulation for Network-on-Chip based Chip Multiprocessor System

ROYAL INSTITUTE OF TECHNOLOGY

Yuan Yao and Zhonghai Lu
KTH Royal Institute of Technology, Stockholm

14th ASP-DAC Conference
19-23, January, 2014, Singapore

Outline

- Introduction
- Problem statement
- Design overview
- Fuzzy logic basics
- Design details
- Experimental results
- Conclusions

Introduction

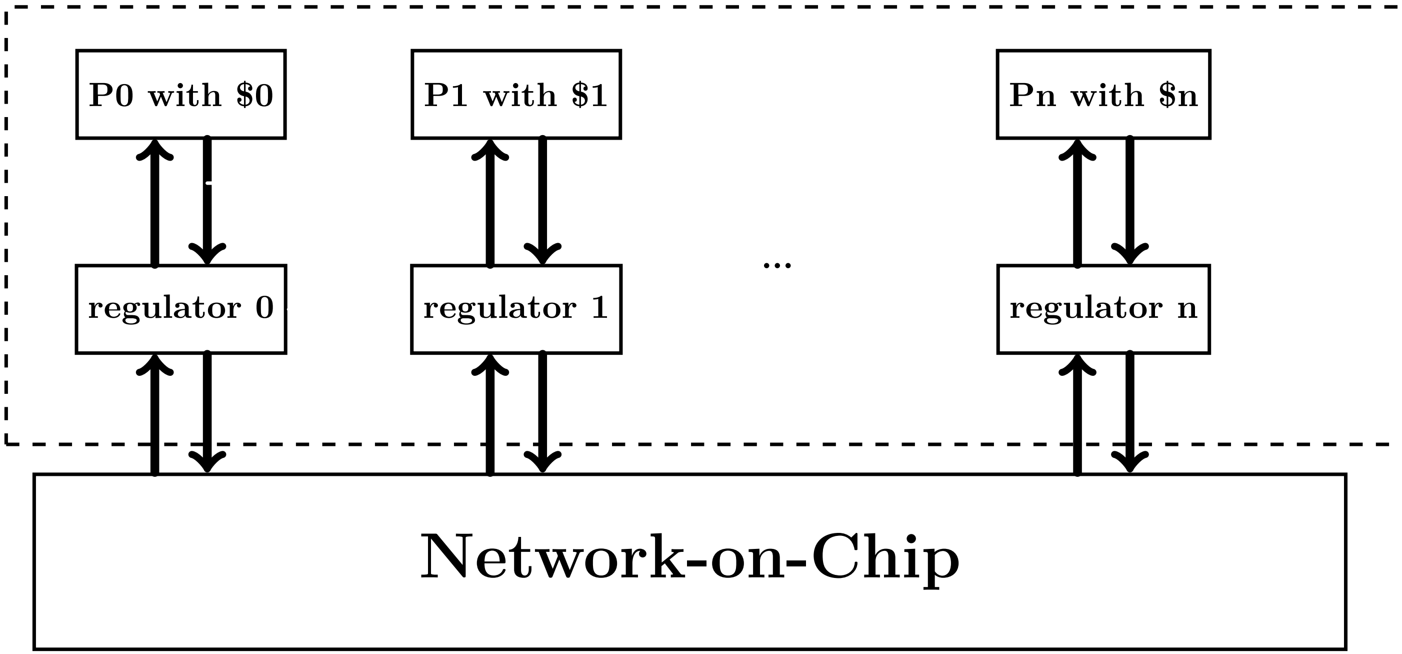
- Chip multiprocessors (CMP) become the mainstream to build high-performance computers.
- Integrating more processors onto a CMP infrastructure presents challenges.
  - Traffic flows from processors are diverse,
  - The impact of interferences among traffic flows may introduce high communication delay and reduce the CMP performance (in terms of CPI).

Flow Regulation

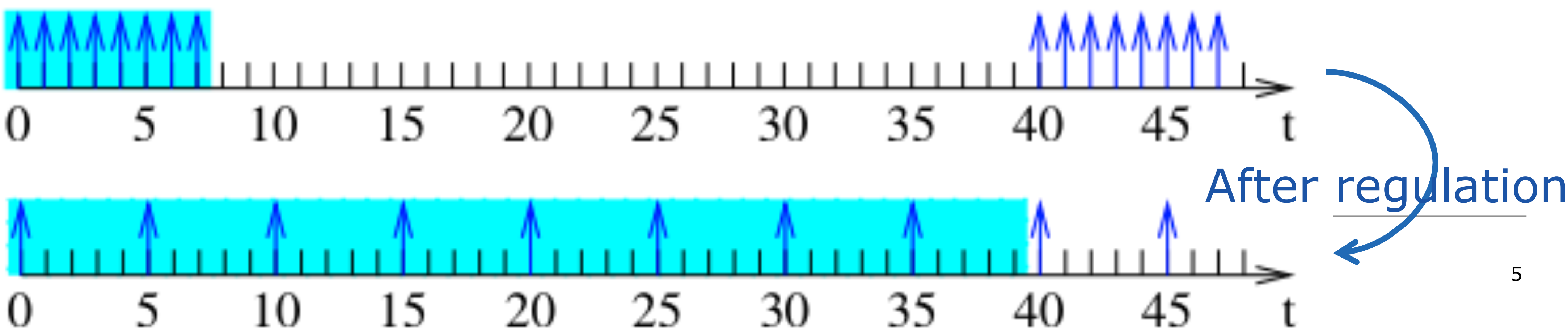
- The on-chip (σ, ρ) flow regulation[1] has been introduced to achieve proactive flow congestion control.
  - Based on *network calculus*[2], a queuing theory for worst-case performance analysis in communication networks.
  - σ constrains the maximum burst of a flow and ρ the long-term average rate.
- It can avoid traffic congestion and further decrease communication delay and reduce buffer requirements.

1. Z. Lu, M. Millberg, A. Jantsch, A. Bruce, P. Wolf and T. Henriksson. "Flow Regulation for On-Chip Communication", *Proceedings of the 2009 Design, Automation and Test in Europe Conference (DATE'09)*, Nice, France, April.
2. J.-Y. L. Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet.* No. 2050 in LNCS, 2004.

Flow Regulation Example



- An example of (1, 0.2) regulation.



ROYAL INSTITUTE OF TECHNOLOGY

Previous Works

- Two approaches in literature to achieve on-chip (σ, ρ) regulation: static[1] and partially dynamic[2].
- Static (σ, ρ) regulation has (σ, ρ) offline configured. It is easy to implement, especially in providing real-time performance guarantees.
- Partially dynamic (σ, ρ) regulation can further adaptively adjust (σ, ρ) parameters online in response to real traffic workload scenarios.

1. Z. Lu, M. Millberg, A. Jantsch, A. Bruce, P. Wolf and T. Henriksson. "Flow Regulation for On-Chip Communication", *Proceedings of the 2009 Design, Automation and Test in Europe Conference (DATE'09)*, Nice, France, April.

2. Z. Lu, Y. Wang, "Dynamic Flow Regulation for IP Integration on Network-on-Chip", *Proceedings of the Sixth IEEE/ACM International Symposium on Network-on-Chip*, May, 2012.

Problem Statement

- However, both the static or partially dynamic regulation lack the adaptivity to different network states. Thus, over-regulation or under-regulation may happen.

Network utilization is not optimized

1. Over-regulation: the network is empty, but the admission control policy is tight.
2. Under-regulation: the network is saturated, but the admission control policy is loose.

- In CMP environment, a desirable flow admission control policy should make regulation decisions according to the traffic dynamism as well as to the state of interconnection network.

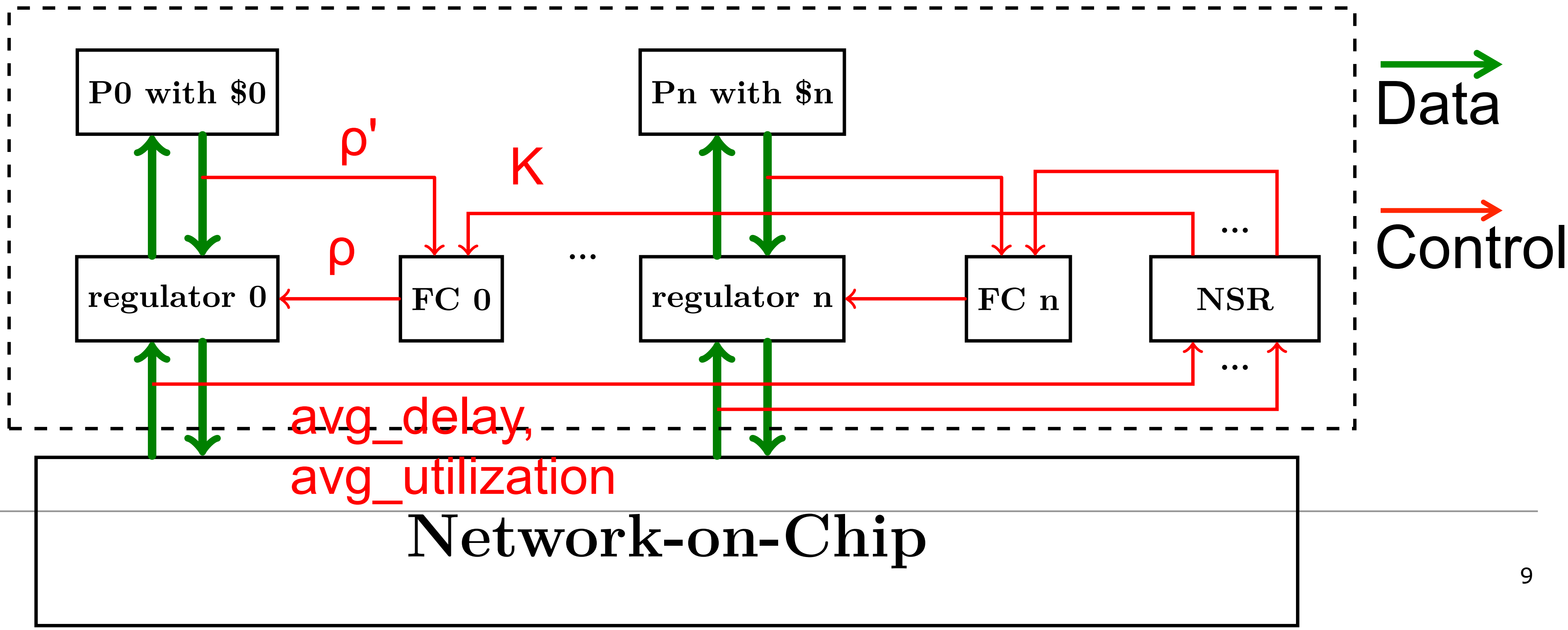
ROYAL INSTITUTE OF TECHNOLOGY

Problem Statement - Cont.

- It may not be appropriate to model the network states using exact logics which defines threshold to distinguish different states.
- **IF** *average_delay*≥x **THEN** the network is *saturated*. How about *average_delay*=x-1? The network is unsaturated at all?
- **The network state is better recognized using fuzzy logic rather than exact logic.**
- **Central idea:** use fuzzy logic to recognize the network status and then intelligently control the admission of input flows.

Design Overview

- We design two kinds of component: network state recognizer (NSR) and fuzzy controller (FC), to achieve fully dynamic flow control.
- The control loop is marked.



KTH VETENSKAP OCH KONST
ROYAL INSTITUTE OF TECHNOLOGY

Design Overview - Cont.

- In CMP, σ is equal to the size of a memory data block, thus remains changed.
- The FC generates new flow injection rate ρ according to:
  1. The cache missing rate of the local CPU (ρ’),
  2. The network state indicator (K) from NSR.
- The NSR monitors the network state through two metrics: average packet delay and average link utilization.
- K reflects the network state, and is assigned one of the following three conditions: empty, normal, saturated.

ROYAL INSTITUTE OF TECHNOLOGY

Design Overview - Cont.

- In order to tolerate the control latency, our design is based on a sampling window mechanism. The new flow injection rate is updated periodically instead of continually.

- All the CPUs share the same state of interconnect network. There is only one NSR per CMP chip. However, FC is attached to each CPU.

ROYAL INSTITUTE OF TECHNOLOGY

Fuzzy Logic Basics - Principle

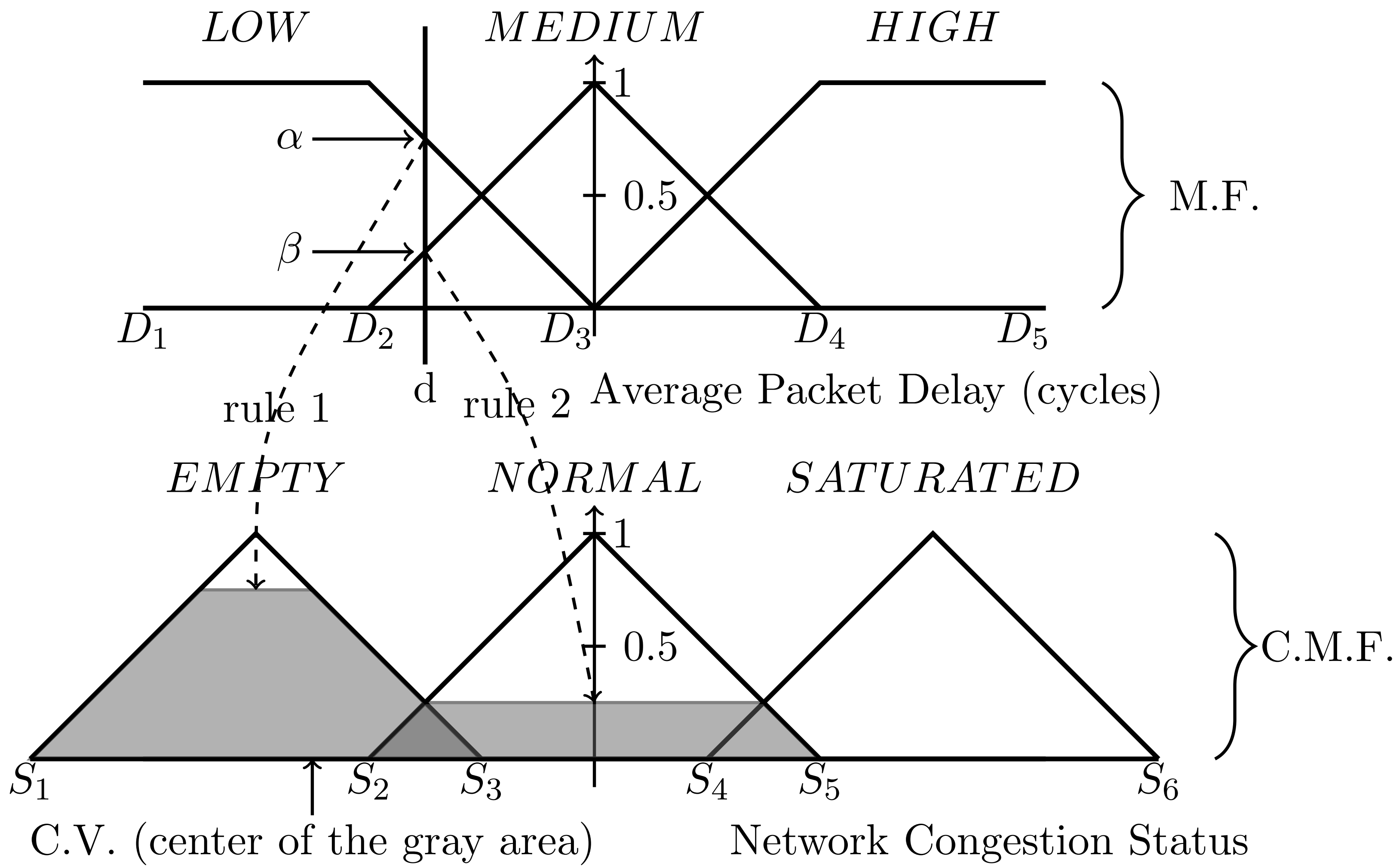
- Instead of using exact true/false, fuzzy logic uses degree of truth, which denotes how confidently we believe a statement is true. Degree of truth is in [0, 1], with “1” indicates totally true and “0” totally false.

- Degree of truth represents that a proposition might be more or less true, rather than simply true or simply false. For instance, the proposition <u>1 + 1 = 2</u> is simply true, while <u>The network is *saturated*</u> is neither simply true nor simply false.

Fuzzy Logic Basics - Terminology

- **Membership Function (MF)**: The MF determines the degree of truth value of a proposition.
- **Fuzzy Rule**: A fuzzy rule is defined as a statement in the form: **IF** x is A **THEN** y is B. For example, **IF** *average_delay*≥x **THEN** the network is *saturated*.
  - Note that a fuzzy rule can also be in the following form: **IF** x is A **AND** y is B **THEN** z is C.
- **Rule Base**: A set of fuzzy rules.
- **Consequence Membership Function (CMF):** Select the "most true" proposition.

Fuzzy Logic Basics - An Example



1. **IF** *avg. packet delay* is LOW, **THEN** *network status* is $EMPTY$
2. **IF** *avg. packet delay* is $MEDIUM$, **THEN** *network status* is $NORMAL$
3. **IF** *avg. packet delay* is $HIGH$, **THEN** *network status* is $SATURATED$

} Rule Base

Reference of Fuzzy Logics

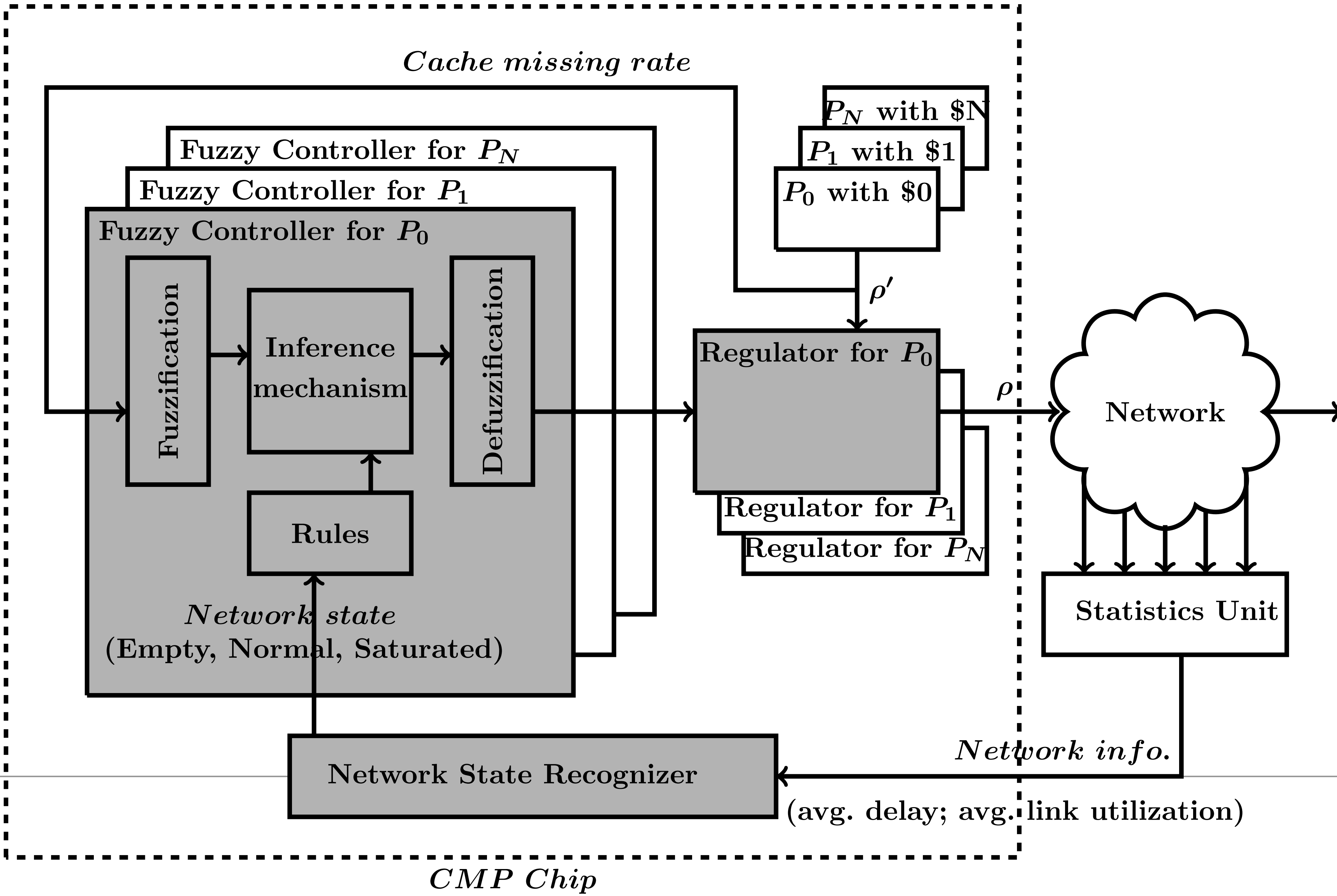
- For more information about fuzzy logics and fuzzy control (for example: shape/ number of MF/CMF, design of rule base) you are encouraged to read the following articles/books:

1. L.A. Zadeh, "Fuzzy Logic", *Stanford Encyclopedia of Philosophy*, Stanford University, 2006.
2. K. M. Passino, S. Yurkovich, *Fuzzy Control*, Addison-Wesley, 1997.

Design Details

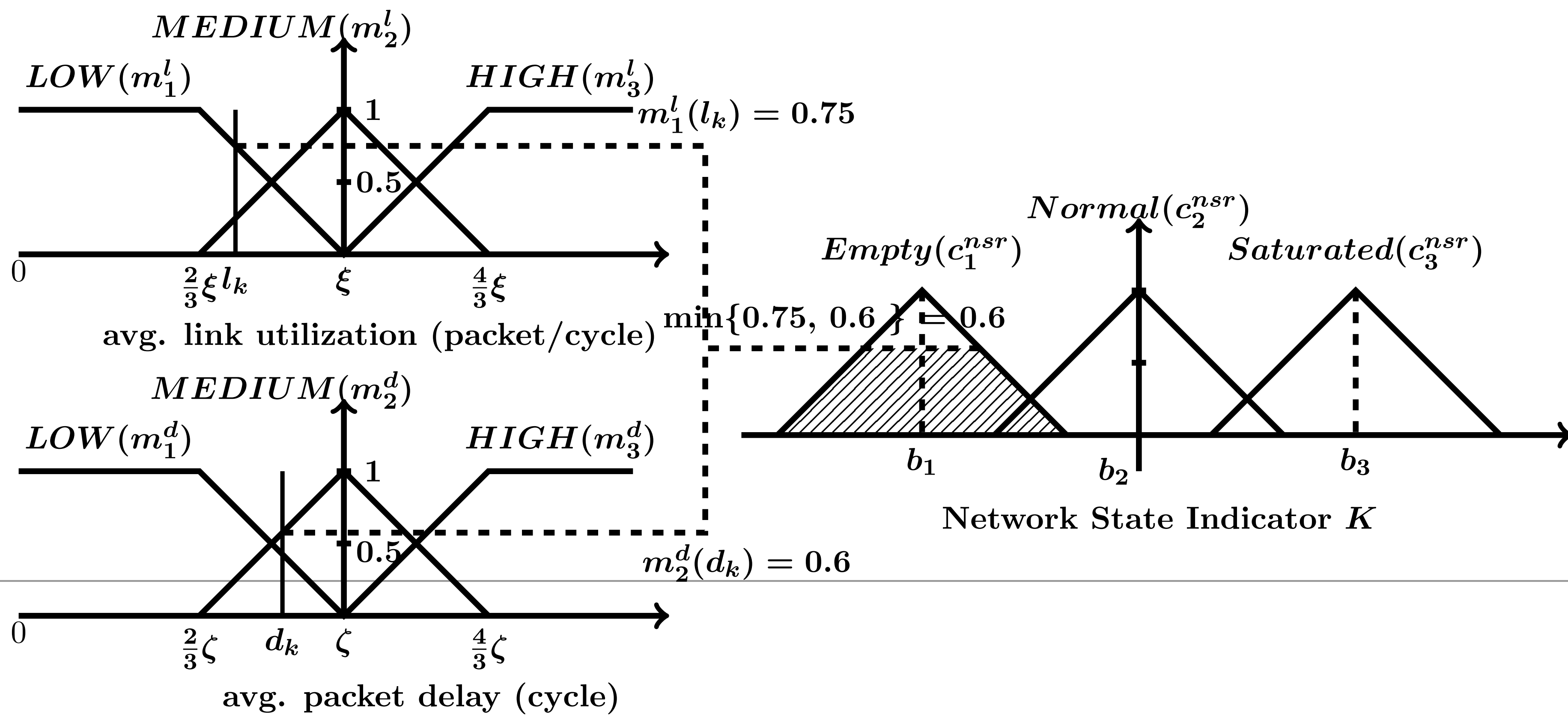
- Both NSR and FC are designed with fuzzy logic, which are composed of the following four elements:
  1. Fuzzification interface, which converts input variables into degree of truth values.
  2. Rule base, which contains a set of fuzzy rules.
  3. Inference mechanism, which makes decision by interpreting and applying rules in the rule base. Inputs for CMF are generated.
  4. Defuzzification interface, which converts the conclusions of the inference mechanism into actual output signals.

Design Details - Cont.



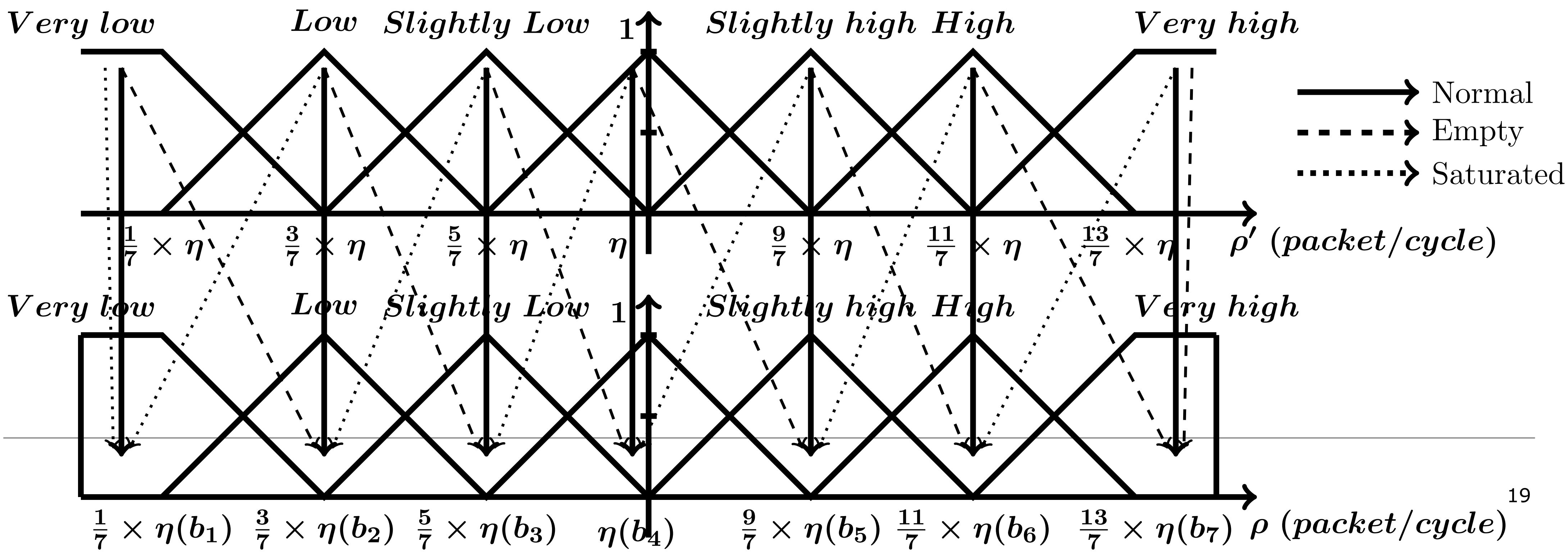
Design Details - NSR

Network state		Packet latency (cycle)		
		Low	Medium	High
Link utilization (packet/cycle)	Low	Empty	Empty	Saturated
	Medium	Empty	Normal	Saturated
	High	Saturated	Saturated	Saturated



Design Details - FC

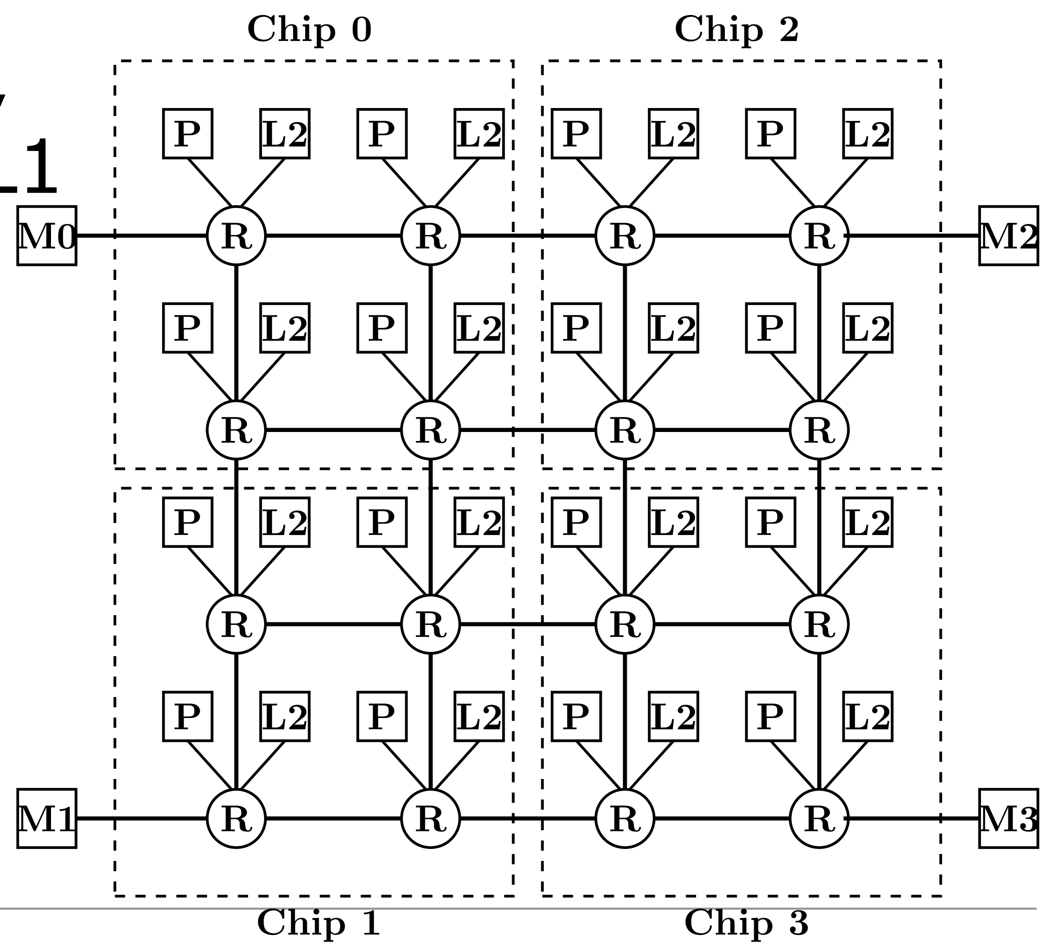
- In fuzzy controller, the network state affects the mapping of the membership functions.
- To avoid both over-regulation and under-regulation: if the network is saturated, then a tight flow control policy is produced. Otherwise, a loose policy is produced.



Experiment Setup

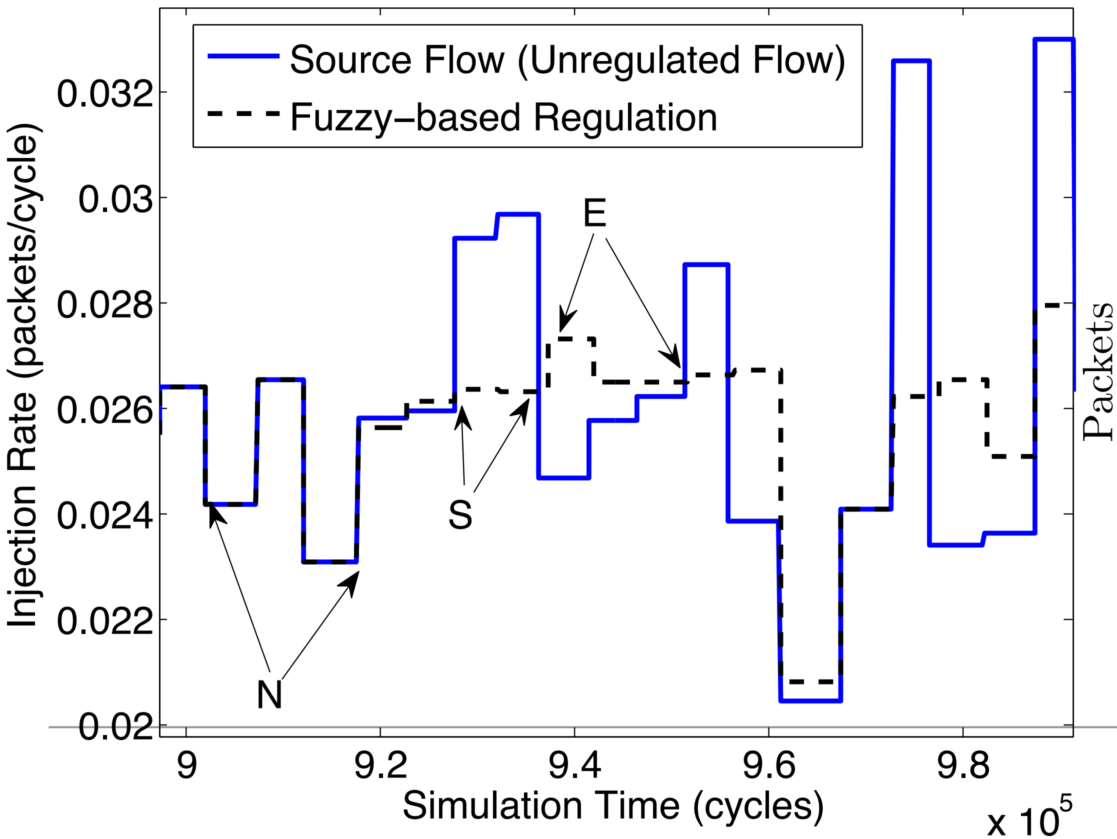
- Our design is implemented using C++ in the cycle accurate full system simulator GEMS.
- We simulate 4 CMP chips, with16 CPUs, 16 private L1 cache and 16 shared L2 cache totally.
  - Synthetic flow patterns:

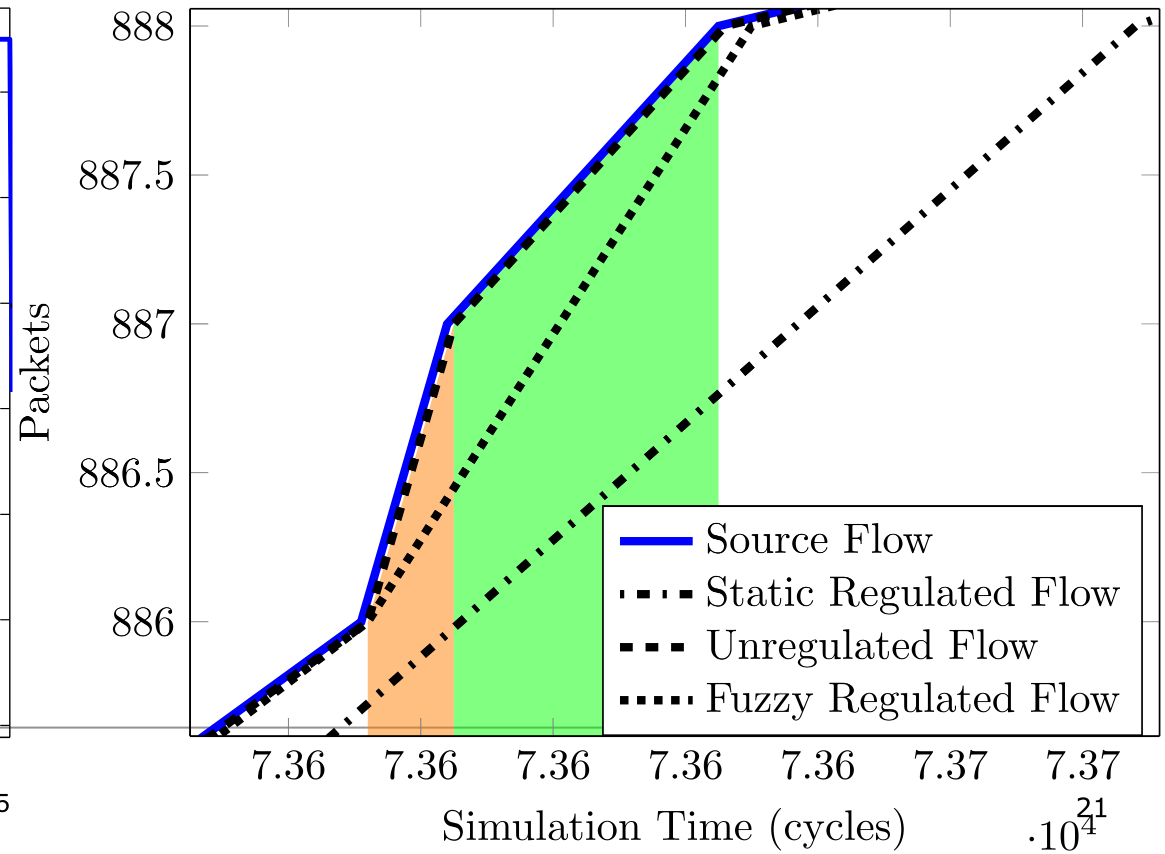
    uniform random and bit-permutation.
  - Benchmark program SPLASH2.



Synthetic Flows I: Uniform traffic

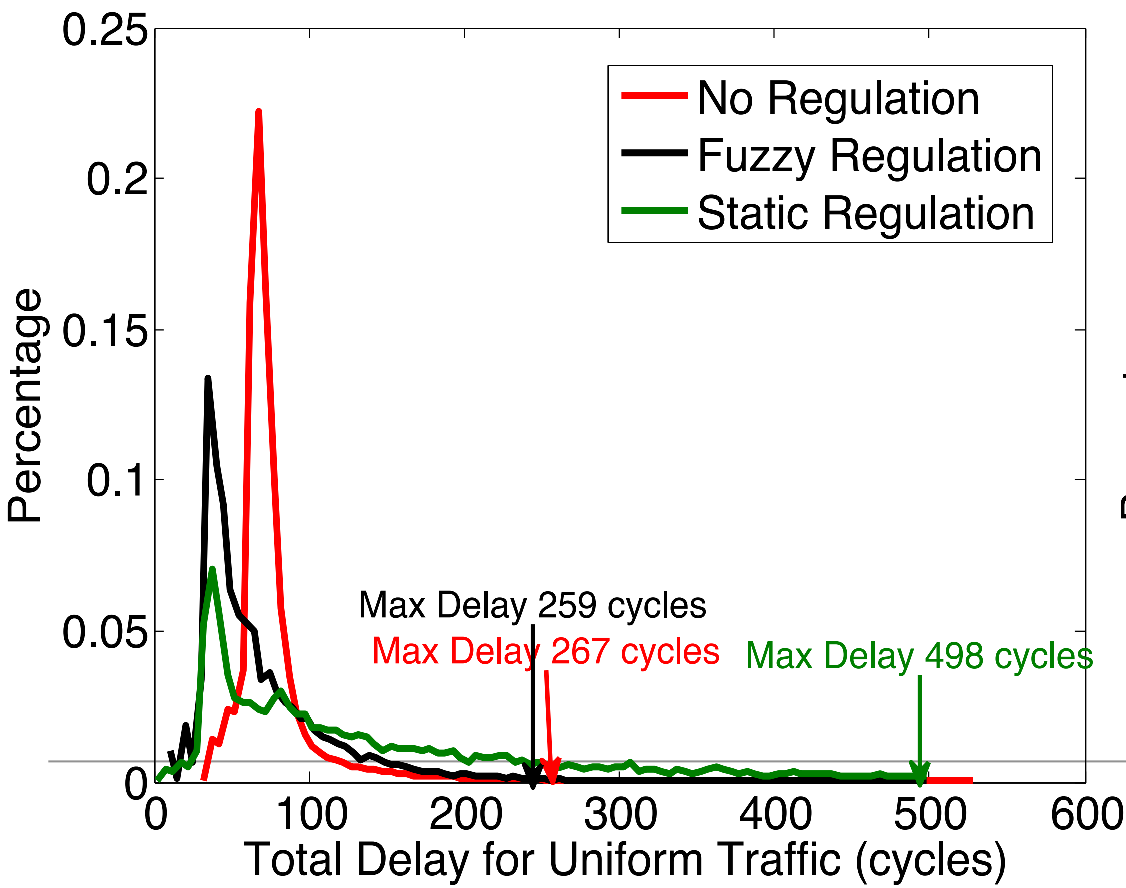
- The experiment shows that the fuzzy flow regulation method can efficiently regulates flows. Both over-regulation and under regulation are avoided.

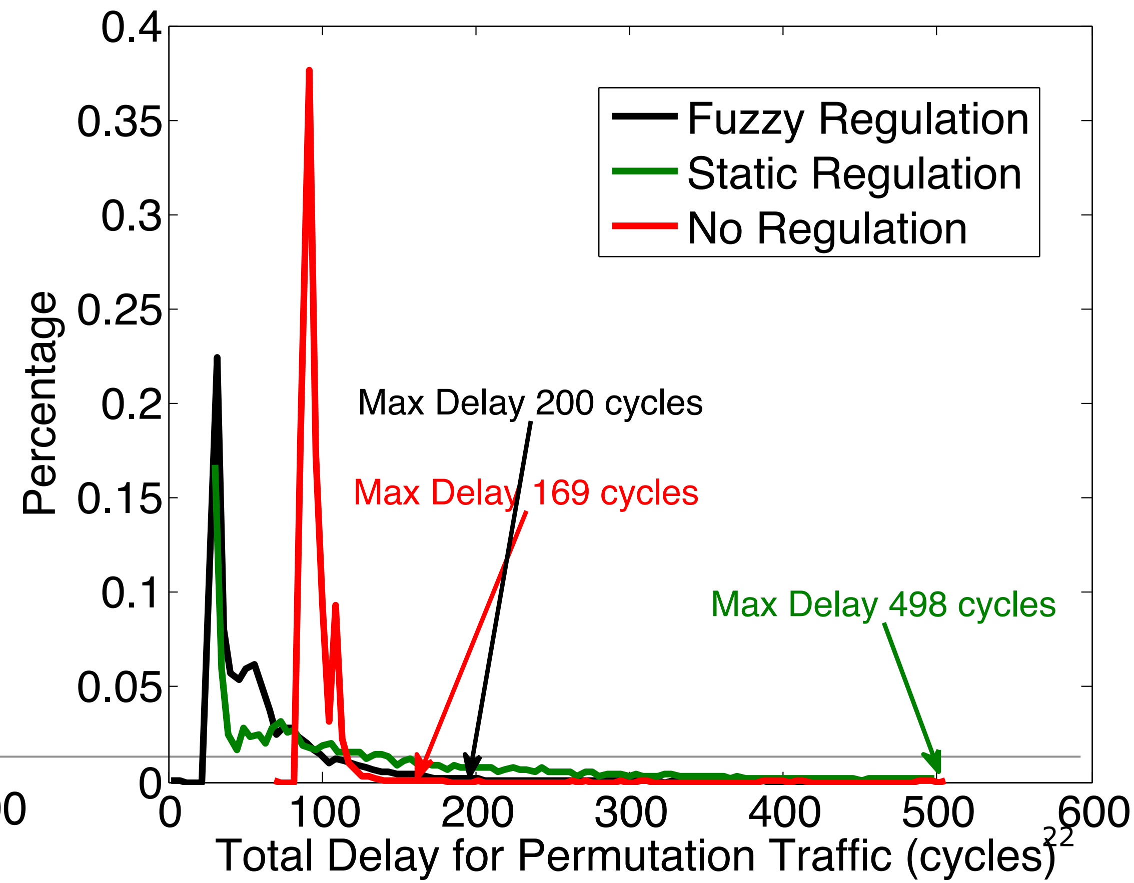




Synthetic Flows II

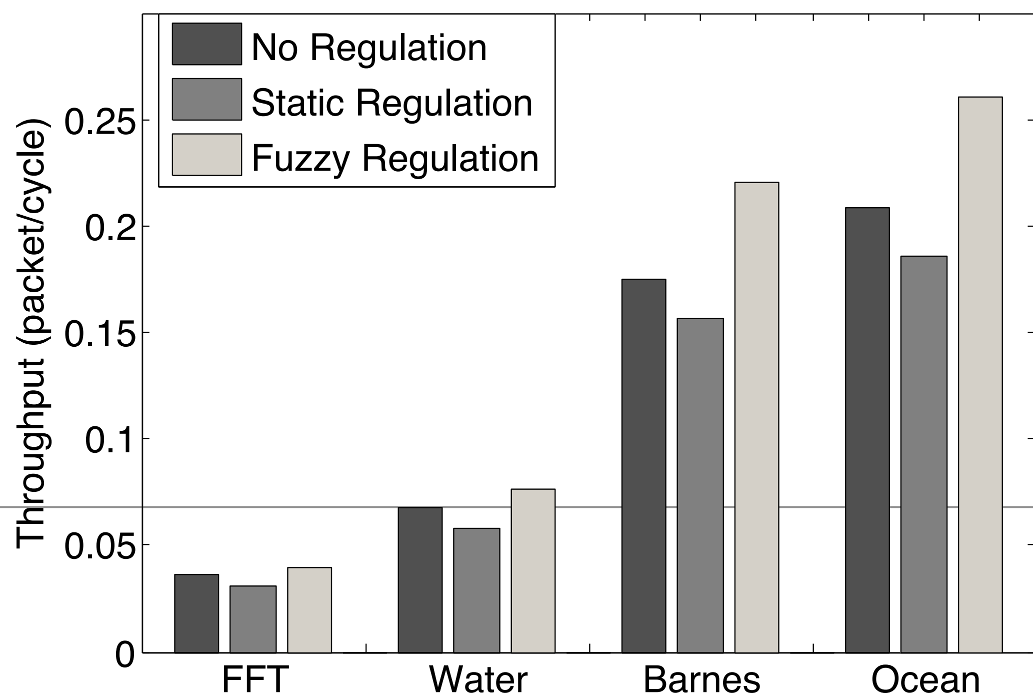
- The fuzzy flow regulation consistently improves the results of static regulation and no-regulation for both traffic patterns.





Results with SPLASH-2

- We experimented with the SPLASH-2 benchmark to confirm the performance benefits brought by the fuzzy flow regulation system.
- Since in a real system the performance is measured by CPI, we choose to report the results for the network throughput here.



Conclusion

- The central idea of this work is to make network status aware flow regulation through a fuzzy logic approach.
- On GEMS, our experiments with both synthetic traffic and SPLASH-2 benchmark traces show that the fuzzy regulation can flexibly adjust regulation strength on demand.
- As a result, it makes more effective use of the system interconnect, achieving significant improvement in average packet delay and throughput.

ROYAL INSTITUTE OF TECHNOLOGY

ACKNOLEDGEMENT

- The research is sponsored in part by Intel® Corporation through a research gift.

ROYAL INSTITUTE OF TECHNOLOGY

Thank you very much!

Q & A