Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

16th IEEE/ACM International Symposium on Networks-on-Chip

SynFull-RTL: Evaluation Methodology for RTL NoC Designs

Neiel Leyva, Alireza Monemi and Enrique Vallejo

October 13, 2022

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

UNIVERSIDAD DE CANTABRIA

Outline

- 1. Introduction
- 2. SynFull Limitations
- 3. SynFull-RTL Methodology
- 4. Evaluation Results
- 5. Conclusions

1. Introduction

Synthetic Workloads: Network traffic is synthetically generated using transposed and random traffic patterns.

SynFull [1] is an attractive solution to use as a traffic generator with synthetic workloads. It avoids long simulation times and gets workloads from a method that involves realistic application patterns.

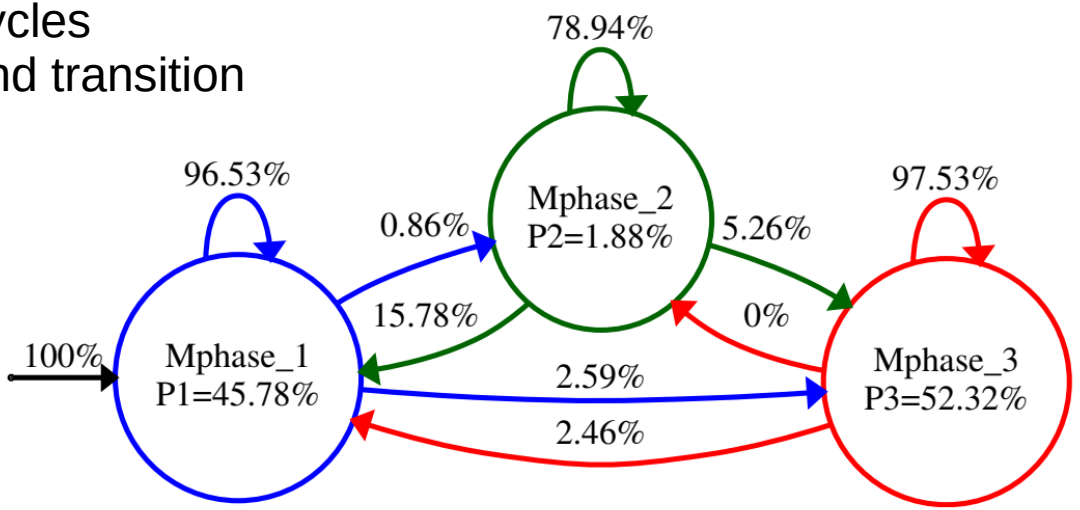
[1] M. Badr and N. E. Jerger, “SynFull: Synthetic traffic models capturing cache coherent behaviour,” in ISCA’14.

1. SynFull Methodology

SynFull splits the simulations into Macro-phases; each model based on a real application has a different number of Macro-phases.

Each Macro-phase has the following characteristics:

- Simulation time: 500,000 cycles
- Probability of occurrence and transition
- Packet injection rate
- Pattern of destinations

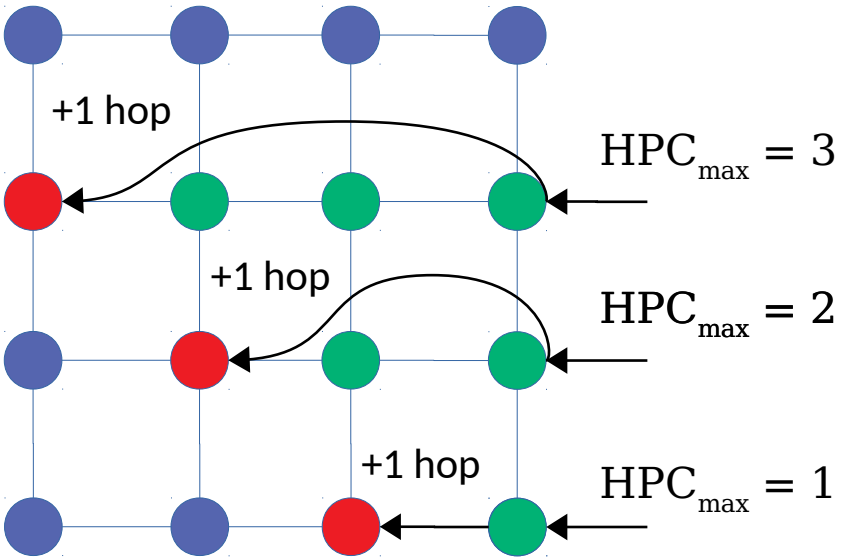


Macro-phases and probabilities of the Barnes Model in SynFull

1. ProSMART

ProSMART [2] is a fully parametrizable NoC router design written in SystemVerilog, integrated with a module that supports multihop bypass.

ProSMART allows injected flits to skip multiple routers in a dimension within a single cycle.

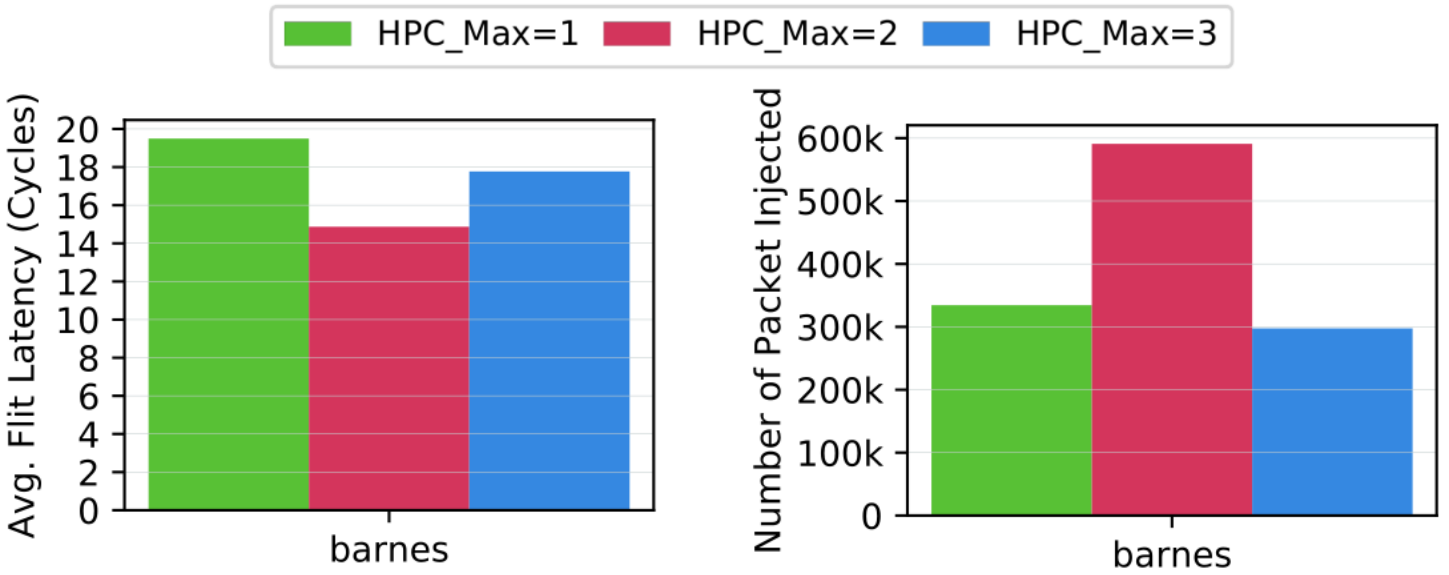


Multi-hop bypass configurations in ProSMART for a 4x4 mesh.

[2] A. Monemi, *et al* "PlugSMART: a pluggable open-source module to implement multihop bypass in Networks-on-Chip," in NOCS'21.

2. SynFull Limitations

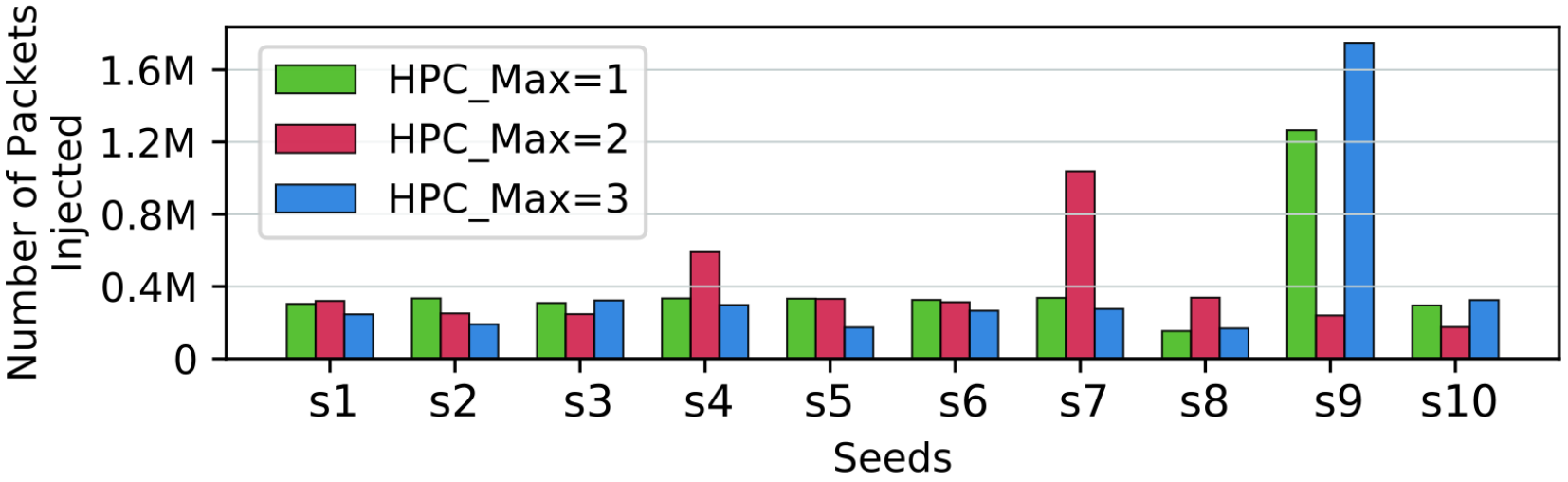
SynFull uses a pseudo-random number generator, based on a hard-coded seed value. High variability in the results for different NoC models was identified using one seed and more different seeds in 10 millions of cycles.



High variability using a single seed value.

2. SynFull Limitations

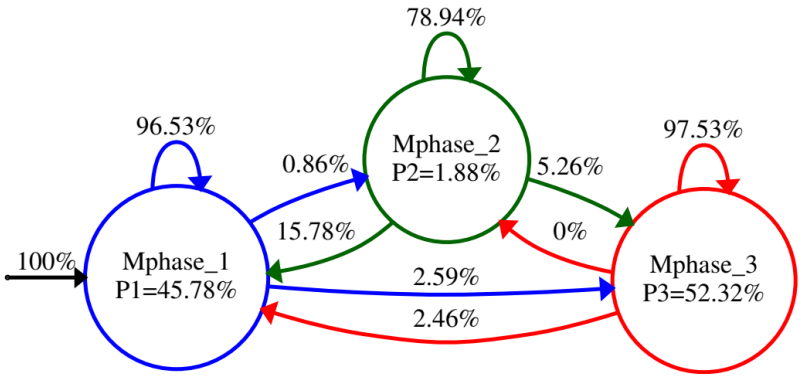
SynFull uses a pseudo-random number generator, based on a hard-coded seed value. High variability in the results for different NoC models was identified using one seed and more different seeds in 10 millions of cycles.



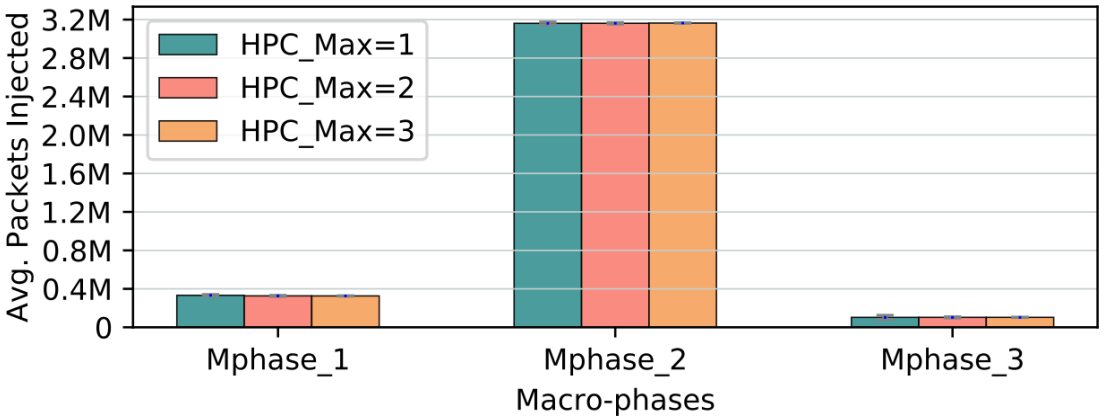
High variability using 10 different seeds.

2. SynFull Limitations

We observe that in SynFull this does not guarantee the same sequence of macro-phases and, thus, an equal or very similar traffic pattern for each model.



Barnes model probabilities



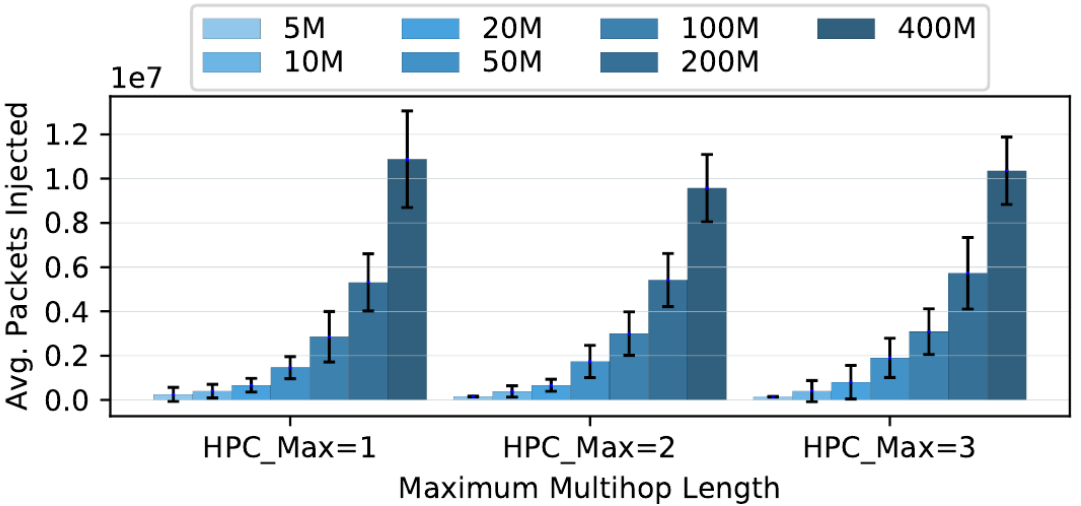
Number of packets injected per macro-phase in Barnes model.

2. SynFull Limitations

The execution time was increased to guarantee the same sequence of macro-phases executed.

Running in a typical HPC cluster, we measured an average 5.4 hours of simulation using SynFull with Booksim.

However, using SynFull integrated with an RTL design, we measured average simulation times of 39 hours, **7.2× slower**.



Number of injected packets for different lengths of the simulation in Barnes model.

3. SynFull-RTL Methodology

All Macro-phases *M* are simulated separately.

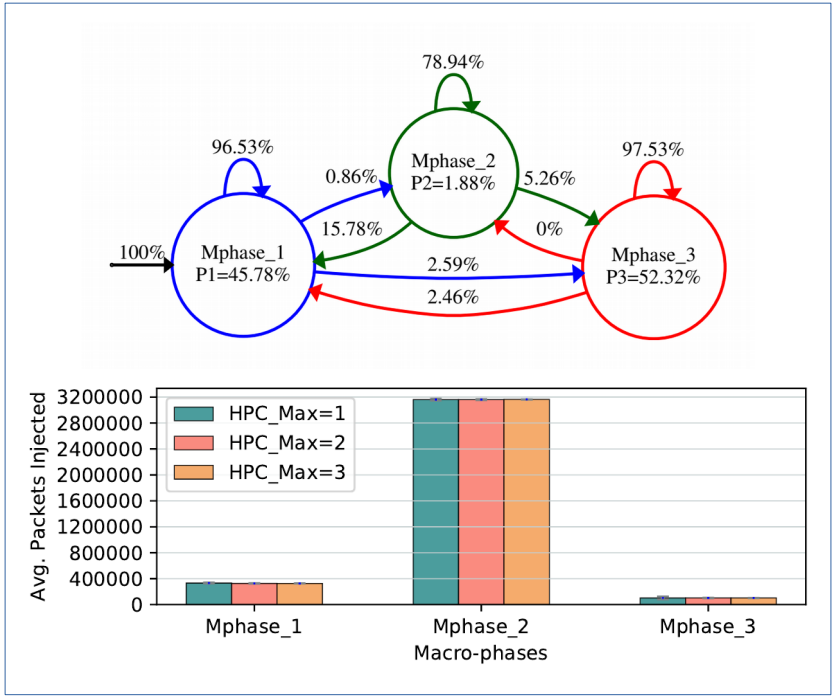
For each Macro-phase *m* with *N* simulations:

- Average packet latency
- Standard deviation of the packet latency
- Average number of packets

To obtain the overall results, a weighted average is employed.

The weight w_m of each *m* depends on:

- Its probability of occurrence.
- Its load intensity.

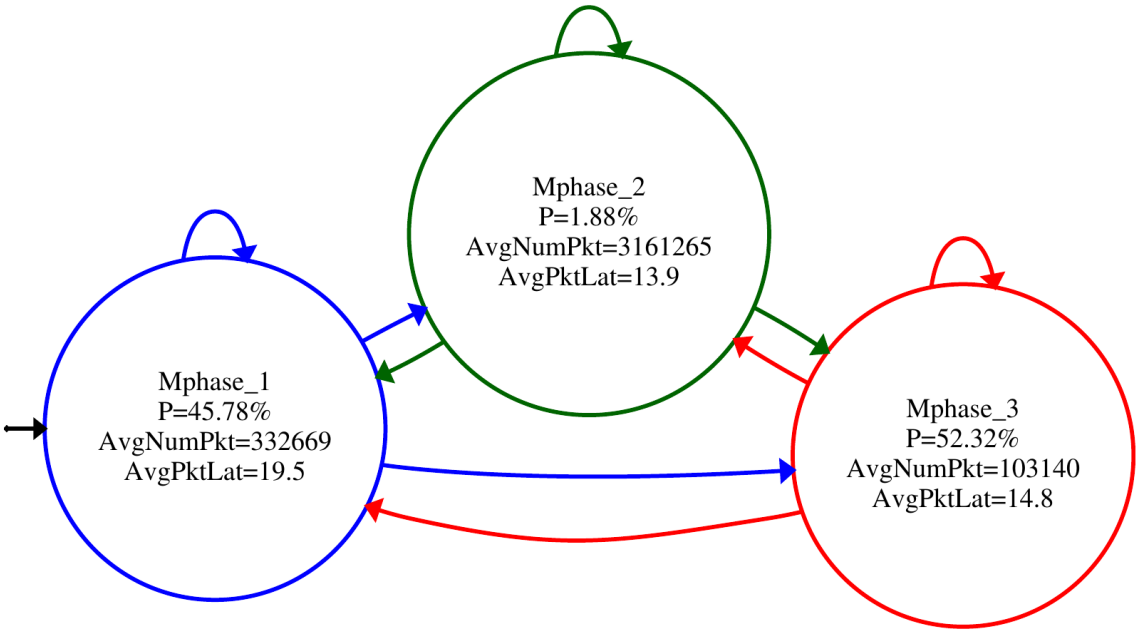


Probabilities and injection ratio of each Macro-phase in the Barnes model

3. SynFull-RTL Methodology

To obtain the weight of each Macro-phase *m*, w_m:

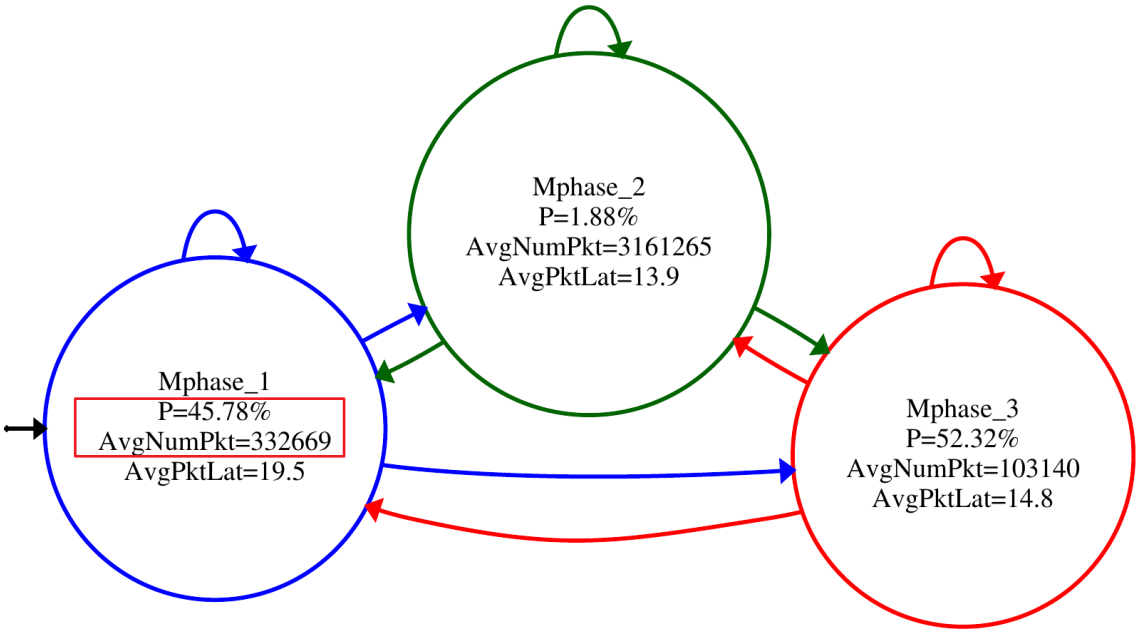
$$w_m = \frac{AvgNumPkt_m \times P(m)}{\sum_{m=1}^{M}(AvgNumPkt_m \times P(m))} \qquad (1)$$



3. SynFull-RTL Methodology

To obtain the weight of each Macro-phase *m*, w_m:

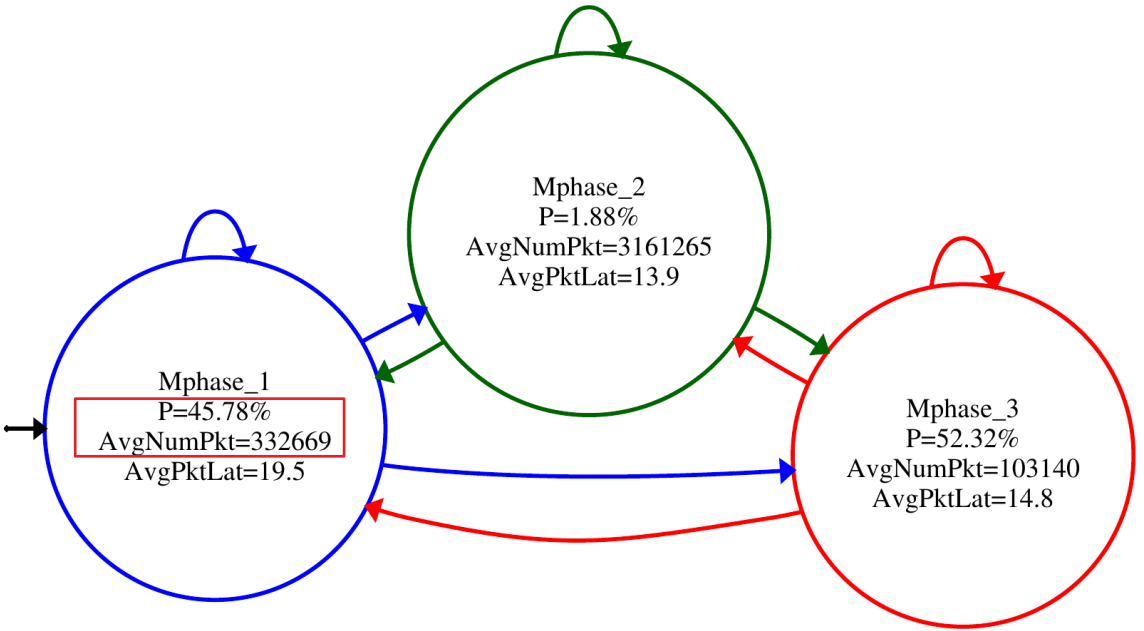
$$w_m = \frac{AvgNumPkt_m \times P(m)}{\sum_{m=1}^{M}(AvgNumPkt_m \times P(m))} \qquad (1)$$



3. SynFull-RTL Methodology

To obtain the weight of each Macro-phase *m*, w_m:

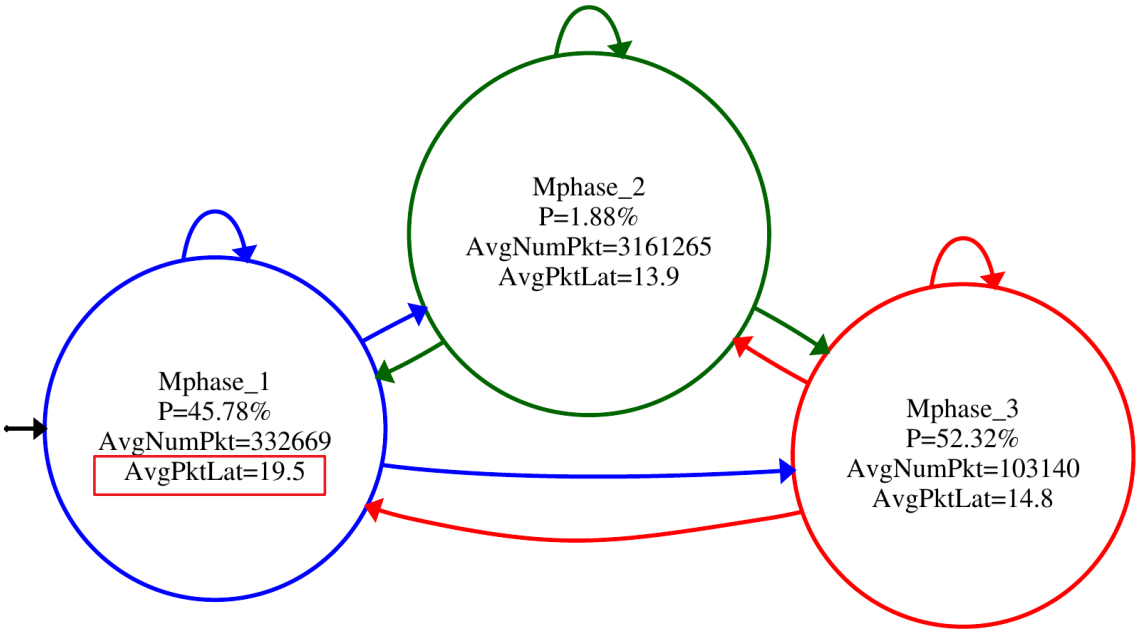
$$w_m = \frac{AvgNumPkt_m \times P(m)}{\sum_{m=1}^{M}(AvgNumPkt_m \times P(m))} \qquad (1)$$



3. SynFull-RTL Methodology

To obtain the average packet latency:

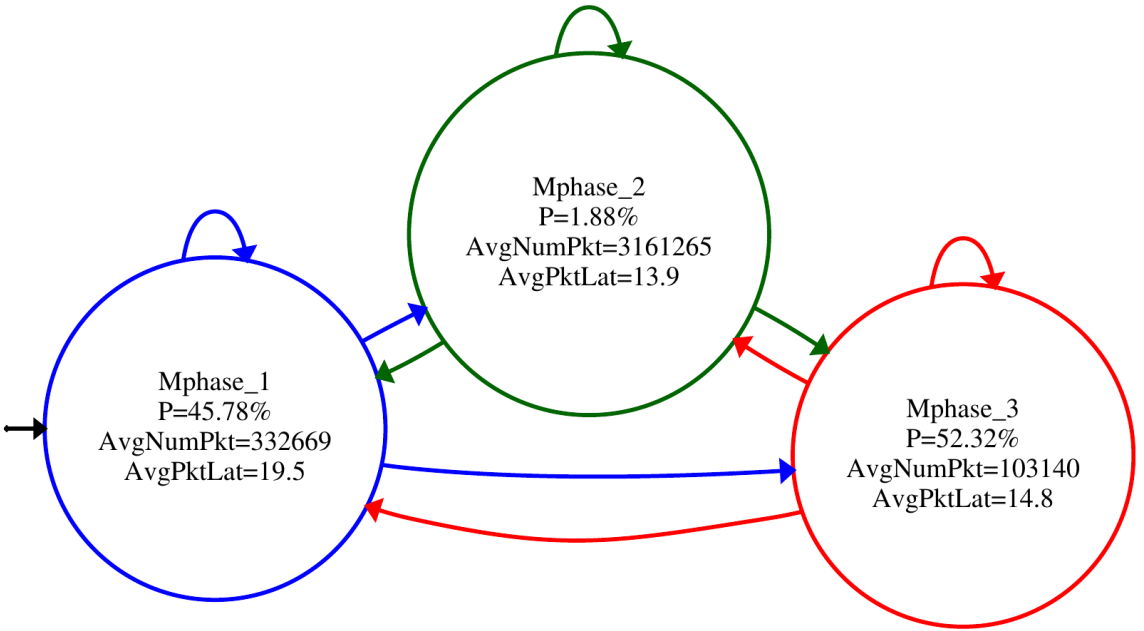
$$AvgPktLat = \sum_{m=1}^{M} (w_m \times AvgPktLat_m) \qquad (2)$$



3. SynFull-RTL Methodology

To obtain the average packet latency:

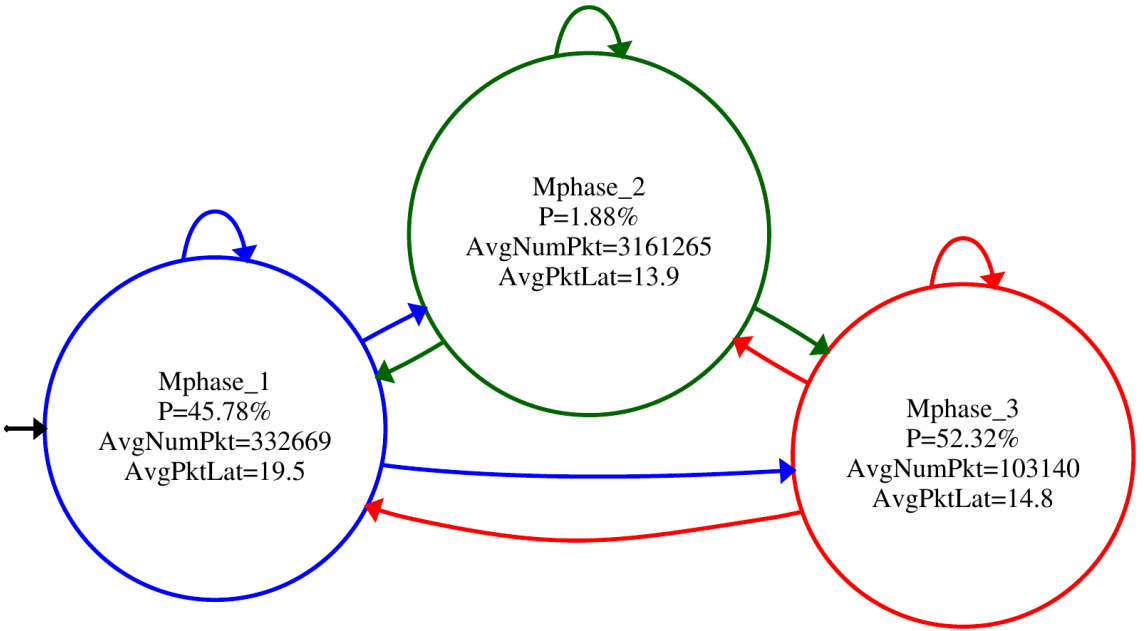
$$AvgPktLat = \sum_{m=1}^{M} (w_m \times AvgPktLat_m) \qquad (2)$$



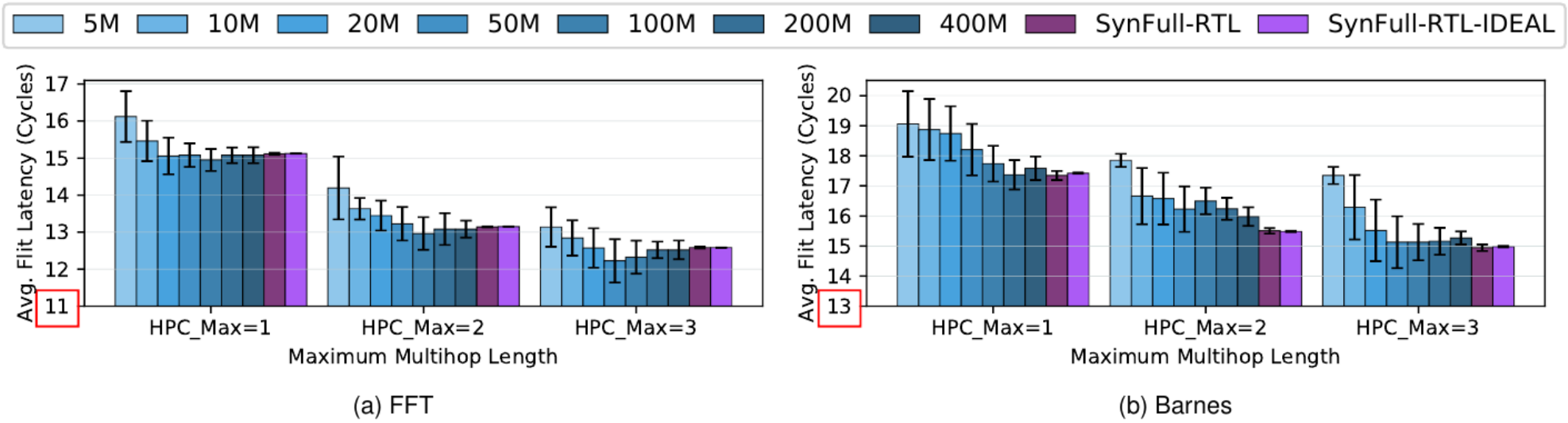
3. SynFull-RTL Methodology

To obtain the average packet latency:

$$AvgPktLat = \sum_{m=1}^{M} (w_m \times AvgPktLat_m) \qquad (2)$$



4. Evaluation Results



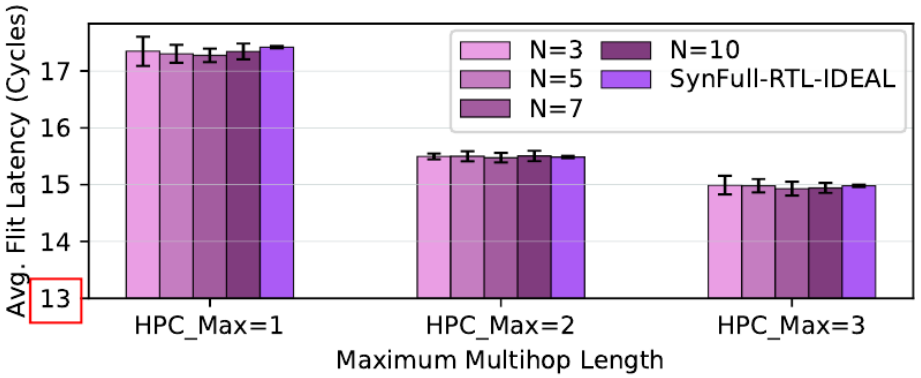
SynFull (blue) and SynFull-RTL (purple) results for average latency.

SynFull : Cycles 5M to 400M , seeds 10.

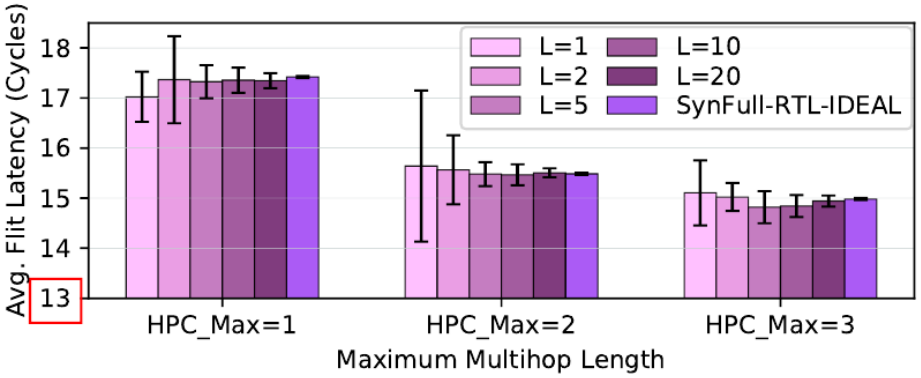
SynFull-RTL : Cycles 10M, seeds 5.

SynFull-RTL-IDEAL: Cycles 200M, seeds 20.

4. Evaluation Results



(a) number of seeds, N ($L = 20$ in all cases)

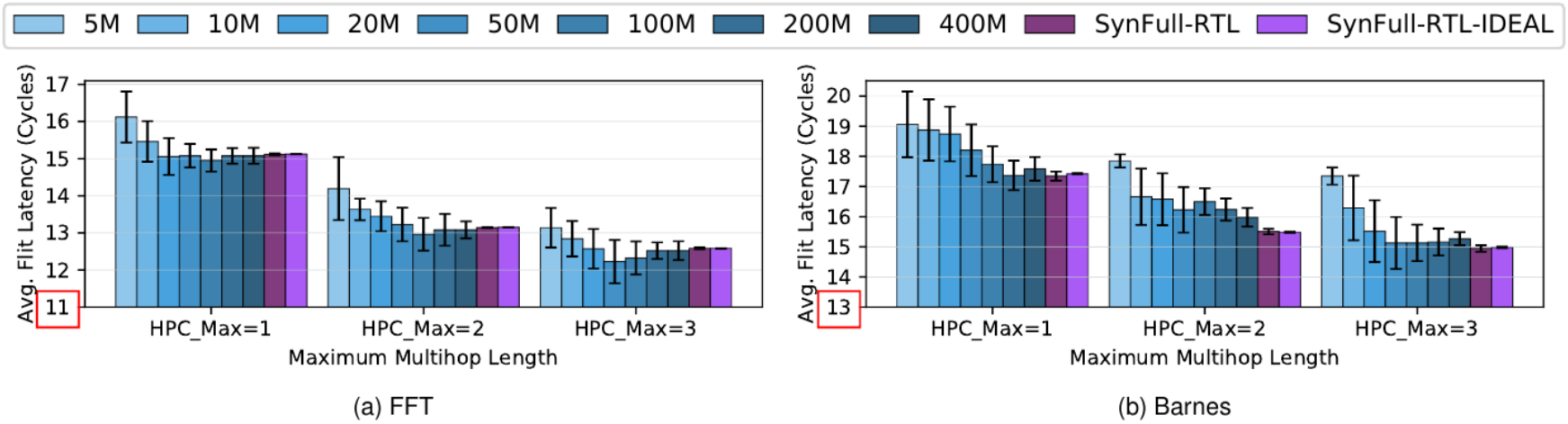


(b) macro-phase iterations, L ($N = 5$ in all cases)

The impact of the number of seeds (*N*) and the number of iterations of each macro-phase (*L*) was evaluated by simulation.

Where L = 1 is equal to 500 000 cycles.

4. Evaluation Results

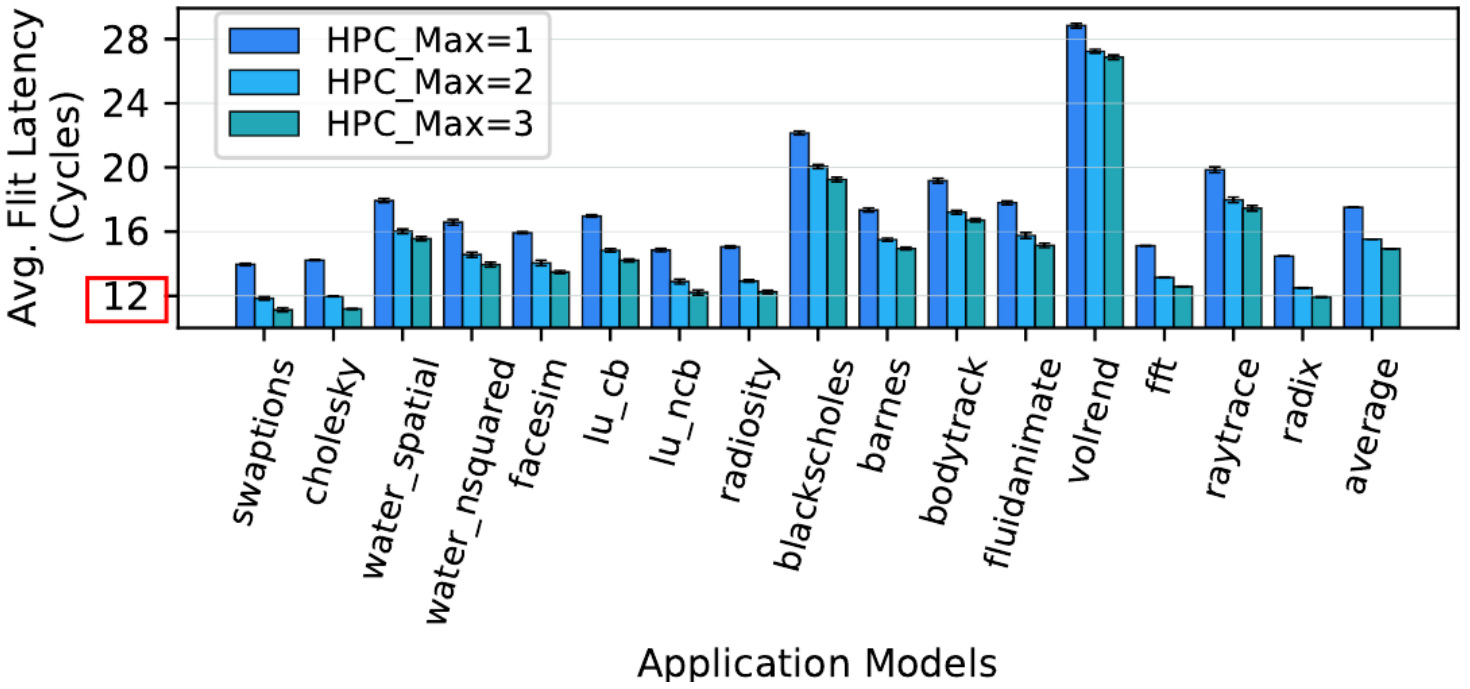


SynFull (blue) and SynFull-RTL (purple) results for average latency.

SynFull-RTL average latency differ by up to 0.58% of the ideal value, whereas SynFull values differ by up to 3.2%.

The results are obtained up to **40× faster**.

4. Evaluation Results



ProSMART average flit latency running the complete set of available models using SynFull-RTL.

5. Conclusions

This work analyzed the **SynFull** methodology and significant limitations were found.

The generated traffic has a large variability. To mitigate this variability, long execution time are required.

SynFull-RTL simulates each macro-phase in isolation and averages the results according to the steady-state probability of occurrence and the measured traffic.

SynFull-RTL methodology reduces time-to-solution by up to 40×.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

16th IEEE/ACM International Symposium on Networks-on-Chip

SynFull-RTL: Evaluation Methodology for RTL NoC Designs

Neiel Leyva, Alireza Monemi and Enrique Vallejo

Thank you!

October 13, 2022

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

UNIVERSIDAD DE CANTABRIA