Informatics Institute
Federal University of Rio Grande do Sul
Porto Alegre, Brazil

April 20, 2020

Dear Editor-in-Chief:

Please find enclosed our manuscript, *Understanding Memory Prefetcher Performance over Parallel Applications: From Real to Simulated*, by Girelli et al., which we would like to submit for publication in the special issue of WSCAD 2019 in Concurrency and Computation: Practice and Experience.

In this paper, we provide a better understanding of the memory prefetcher's role in the performance of HPC applications, considering the prefetcher algorithms offered by both the real hardware and the simulators. We performed a careful experimental campaign, executing the NAS parallel benchmark (NPB) on a real Skylake machine, and as well in a simulated environment with the ZSim and Sniper simulators, taking into account the prefetcher algorithms offered by both Skylake and the simulators. We obtained many interesting findings about the prefetcher's behavior on the NPB applications. Some highlights of these findings are:

- We show experimental evidence that an L2 memory prefetcher is more efficient in comparison with an L1 prefetcher, since avoiding excessive L3 cache accesses better contributes to performance, when comparing to accessing the L2 cache;
- We show evidence that the prefetchers' contribution to performance is limited by the level of parallelism of the application, mainly due to the increase in communication and memory contention as the level of parallelism increases;
- The Skylake and NPB simulation with ZSim and Sniper had poor accuracy in predicting the NPB applications performance, with and without simulating prefetcher algorithms, mainly due distinctions of the models and prefetcher algorithms present in ZSim and Sniper when compared to Skylake.

This article is an extension of our previous paper *Impacto do Prefetcher na Precisão de Simulações de Arquiteturas Paralelas*, published in the WSCAD 2019 conference. Compared to the previously published paper, the paper has undergone major changes.

These include:

- We add the Sniper simulator, which supports prefetcher algorithms;
- We used a newer hardware model in the system used in our evaluation;
- We add more detailed analysis of the NPB benchmark performance on both real and simulated architecture executions;
- We adopt the instructions per cycle (IPC) metric to evaluate the performance of the NPB benchmark.

We believe our findings would appeal to the readership of the Concurrency and Computation: Practice and Experience. This manuscript has not been published elsewhere and is not under consideration by another journal.

We look forward to hearing from you at your earliest convenience.

Sincerely,

Valéria S. Girelli
Francis B. Moreira
Matheus S. Serpa
Danilo Carastan-Santos
Philippe O. A. Navaux