# Assessing the Prefetcher's Role in High Performance Computing

Valéria S. Girelli[1], Francis B. Moreira[2], Matheus S. Serpa[1], Danilo Carastan-Santos[1], Philippe O. A. Navaux[1]
[1]Institute of Informatics, Federal University of Rio Grande do Sul (UFRGS) – Porto Alegre, Brazil
[2]Department of Informatics, Federal University of Paraná (UFPR) – Curitiba, Brazil
{vsgirelli, msserpa, danilo.csantos, navaux}@inf.ufrgs.br, fbmoreira@inf.ufpr.br

***Abstract*—Memory prefetchers have been broadly used in modern processors to mitigate the performance gap between processor and memory. In High Performance Computing (HPC) systems, many complications arise from parallelism, which are crucial to understand the prefetcher's impact over parallel applications' performance. In this work, we aim to evaluate the behavior of the prefetchers available in a Skylake machine and in two simulators of parallel architectures, ZSim and Sniper. Our results demonstrate that prefetching from the L3 to the L2 cache presents the best performance gains, and the prefetcher's efficiency is restrained by the memory contention that emerges from increasing the parallelism. Moreover, both ZSim and Sniper only simulate inclusive L3 caches, while Skylake's L3 is non-inclusive. Consequently, the outcome is a poor simulation of both the memory contention and the prefetcher behavior.**

***Index Terms*—Computer Architecture, Parallel Architecture, Prefetcher, Parallel Simulation.**

## ACKNOWLEDGMENT

This work has been partially supported by Petrobras (2016/00133-9, 2018/00263-5) and Green Cloud project (2016/2551-0000 488-9), from FAPERGS and CNPq Brazil, program PRONEX 12/2014. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Experiments presented in this paper were carried out using the Grid'5000 experimental testbed, being developed under the INRIA ALADDIN development action with support from CNRS, RENATER and several Universities as well as other funding bodies (see https://www.grid5000.fr).

## DISCLAIMER

This abstract describes the paper entitled "Understanding Memory Prefetcher Performance over Parallel Applications: From Real to Simulated" from the same authors, that has been submitted and is under review for publication on the journal Concurrency and Computation: Practice and Experience (CCPE).