Activity Report

Danilo Carastan dos Santos

(✉ dancarastan@gmail.com - danilo-carastan-santos.github.io)

23rd January 2023

Contents

1 Activity Report

My research activity takes place in the context of large-scale parallel and distributed infrastructures such as supercomputers and computing clusters and their applications, with a recent emphasis on limiting their energy consumption and environmental impacts. Such infrastructures are complex in terms of high-level parallelism and heterogeneity, are shared by numerous users, and rapidly evolve over time. Such complexity raises many fundamental questions and technical challenges in order to exploit such infrastructures as efficiently as possible.

I organize my research activity in the following three research axes:

- Data Science and Machine Learning applied to High-Performance Computing infrastructures' resource management
- Frugal methods to limit the energy consumption and environmental impacts of distributed computing
- Analysis and development of parallel and distributed infrastructures applications and architectures

In the sections below, I describe my contributions among each axis and the underlying collaborations.

1.1 Axis 1: Data Science and Machine Learning applied to High-Performance Computing infrastructures' resource management (2016-present)

This research axis is about finding heuristics for High-Performance Computing (HPC) resource management. Optimal resource management typically requires solving instances of an NP-Hard problem called online parallel job scheduling. In contrast to many theoretical advances in this problem, practitioners use naive First-Come-First-Served (FCFS) based heuristics, arguably because the simplicity of FCFS-based heuristics is crucial for real-world use cases. The challenge lies therefore in constructing efficient online parallel job scheduling heuristics while keeping a similar simplicity and theoretical guarantees of FCFS.

In [3], under supervision of Professor Raphael Y. de Camargo (Associate Professor at Federal University of ABC, Brazil), I played a part in paving the way to use Machine Learning methods for online parallel job scheduling. I innovated in proposing ways to use regression methods to generate scheduling heuristics for HPC platforms, creating a novel link between Machine Learning (ML) methods and simple HPC scheduling heuristics. Before this contribution, most works relied purely on job execution logs. I went beyond and explored HPC simulation to generate novel derivative scheduling data to feed ML algorithms. The contribution's novelty lies in exploiting HPC simulation to generate scheduling data not present in HPC logs and how to feed this data into regression algorithms to create efficient and simple scheduling heuristics. The proposed scheduling heuristics remain among the most efficient known heuristics in the literature. The above work received nominations for Best Paper and Best Student Paper awards of a highly prestigious, international High-Performance Computing conference (see Sections 2.1 and 3.1).

We observed from the heuristics obtained in [3] that the jobs "area" (i.e., the multiplication of the requested processing time and the number of processors of the jobs) is an important factor for efficient heuristics. In [5], in a conjoint work with former colleague Salah Zrigui, and under supervision of Raphael Y. de Camargo and Denis Trystram (Distinguished Professor at Grenoble INP *Institut d'ingenierie et de management*), we further investigated this factor by proposing novel, Data Science based ways to evaluate the HPC job scheduling performance. We highlighted a known heuristic called Shortest estimated Area First (SAF), which consists of prioritizing the jobs in increasing order of their "area". Adding a job threshold mechanism to SAF provided the same theoretical guarantees of FCFS (i.e., no job starvation, all jobs will execute at some point) while being up to 80% more efficient than FCFS. After this contribution, it is still an open challenge of how to be more efficient than SAF by a large margin. This work was the winner of the Best Paper Award at a prestigious international High-Performance Computing conference (see Sections 2.1 and 3.1).

The work in [3] shed light on the hypothesis that linear regression methods are promising for exploiting scheduling data to create simple and transparent scheduling heuristics with lesser computational overhead than state-of-the-art learning methods. In this regard, I collaborated with the University of São Paulo (USP) by co-supervising with Alfredo Goldman (Associate Professor at University of São Paulo, Brazil) the student Lucas Rosa. In [14], we explored ways to increase the scheduling performance of regression-obtained scheduling heuristics, which allowed us to discover non-trivial findings regarding obtaining scheduling heuristics with linear regression, notably the relationship between regression multicollinearity effects and scheduling performance. I also collaborated with my former colleague Luis Sant'ana by evaluating other Machine Learning methods, such as logistic regression and support vector machines, to perform real-time scheduling heuristics selection [15, 16]. The idea was to change the scheduling heuristic in real-time according to the properties of the HPC infrastructure and the waiting queue.

1.2 Axis 2: Frugal methods to limit the energy consumption and environmental impacts of distributed computing (2021-present)

This research axis addresses the critical issue limiting the energy consumption and greenhouse gas (GHG) emissions of information and communications technology – whose global share of GHG emissions is in the order of 3%[1], and in par with the aviation industry[2] – in the context of large-scale distributed computing.

A disruptive aspect approached in this axis is that I leverage my expertise in Machine Learning, Data Science, and visualization to conceive simple, low-cost, frugal, and efficient methods for real-world needs. I pursue to show that there are viable alternatives to the costly black-box approaches of deep neural networks. This axis is in an early stage of advancement. Below I present some preliminary results of this axis, and the research program presents how I plan to further advance in this axis.

In [9], we constructed a benchmarking tool to measure the energy consumption of training and inference of numerous Artificial Intelligence (AI) models. This work tackles a challenge faced by AI engineers: choosing which AI model is better suited for the available hardware regarding energy consumption. The dimensions of different hardware generations, especially for GPUs, and the different computations performed by the AI models hinder understanding the energy efficiency of AI algorithms given particular hardware. Our benchmark tool allows a lightweight evaluation of numerous AI algorithms in a given hardware (CPU and GPU). It enables AI engineers to evaluate their target hardware efficiently and choose an energy-efficient AI model for their hardware. This work was done by my former M1 student Thi Hoang Thi Pham. This work has earned me an invitation as a lecturer at the Second Inria-DFKI European Summer School on AI (IDESSAI 2022)[3]. In the lecture, I explained how to use software tools to measure the energy consumption of AI code, with practical examples[4] on how we can use this energy information to guide the AI model training. I also did a Portuguese version of this lecture at the *II Escola Regional de Alto Desempenho/Aprendizado de Máquina e Inteligência Artificial Norte 2*[5].

I am interested in the environmental impacts of distributed systems but also in the impacts of research activities in this field. I had the opportunity to collaborate with the International European Conference on Parallel and Distributed Computing (Euro-Par) steering committee to estimate the CO2 emissions, as a measure of environmental impact, of the 2020 and 2021 editions of the conference, which took place virtually. We can see virtual conferencing as a distributed computing infrastructure, with a data center sharing data between end devices (i.e., laptops). In our paper [2], my contribution was in the challenge of giving accurate CO2 emissions estimations considering the conference participant's information in the

[1] Charlotte Freitag et al. *"The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations"*. In: Patterns 2.9 (2021), p. 100340.

[2] Our World in Data. *Climate change and flying: what share of global CO2 emissions come from aviation?* 2020. url: `https://ourworldindata.org/co2-emissions-from-aviation` (visited on 09/07/2022).

[3] `https://idessai.eu/track-b-sustainable-ai-2022/`

[4] `https://github.com/danilo-carastan-santos/ai-energy-consumption`

[5] `https://www2.sbc.org.br/erad_eramia-no2/programa.html`

temporal and geographical dimensions and considering the energy mix of the participant's country. I also was the leading writer of this paper, and I collaborated with processing, analyzing, and interpreting results.

1.3 Axis 3: Analysis and development of parallel and distributed applications and architectures (2013-2018 and 2020-2021)

This axis compiles additional research works in parallel and distributed systems. I organized this axis into two components, (i) analysis of parallel and distributed infrastructures and (ii) development of parallel and distributed applications.

1.3.1 Analysis of parallel and distributed infrastructures (2020-2021)

In [10], we explored a novel distributed computing infrastructure consisting of smart heaters, which are computing machines that work as heaters in a smart building. We used Qarnot Computing's[6] implementation of a smart heater as a use case for using Machine Learning forecasting methods to model and predict the heating provided by such machines. This work was done by Anderson Andrei da Silva at the end of his undergraduate studies under my supervision. I supervised his work in collaboration with the University of São Paulo (USP), Grenoble Informatics Laboratory (LIG), and Qarnot Computing within the ANR Greco project[7].

During my first post-doc, in collaboration with Philippe Navaux[8] (Professor Emeritus of the Federal University of Rio Grande do Sul, Brazil) and the students at the Parallel and Distributed Processing Group (GPPD) of the Informatics Institute (INF) of the Federal University of Rio Grande do Sul (UFRGS, Brazil), we took a step towards understanding better the contribution of memory prefetcher algorithms in the performance of parallel applications [11]. Memory prefetcher algorithms help mitigate memory latency. Still, since numerous threads can access different parts of the memory in parallel applications, the gains in performance provided by prefetcher algorithms for highly parallel applications are yet to be fully understood. The rapid parallelism increase of processor technology (i.e., from two cores to tens of cores) may increase the uncertainty in the memory accesses, thus reducing the efficiency of prefetcher algorithms. We verified this hypothesis, and we showed evidence that the prefetchers' contribution to performance is limited by the level of parallelism of the application, reaching negligible performance increase at high parallelism levels, mainly due to the increase in communication and memory contention as the level of parallelism increases. In this first post-doc, I complemented the GPPD group with my data science and visualization expertise to synthesize experimental results in the context of distributed applications and architectures, which led to fruitful collaborations beyond the aforementioned one [1, 13, 12].

1.3.2 Development of parallel and distributed applications (2013-2018)

I worked on highly parallel exact algorithms for Minimal Hitting Set problems in developing parallel and distributed applications. Important applications, notably in bioinformatics, are modeled as instances of a SAT-like problem called Minimal Hitting Set (MHS). Finding minimal hitting sets is NP-hard, and often finding all minimal hitting sets given a particular problem instance is envisaged in bioinformatics, thus requiring exact algorithms. I employed several algorithmic innovations to develop a highly parallel, exact Minimal Hitting Set algorithm, using accelerator devices such as GPUs [6, 4, 17], and collaborative computing of CPU, GPU, and MIC (former Intel Xeon Phi) [7]. One of these innovations is using a combinatorial numbering system to encode possible MHS solutions. This encoding allows representing possible MHS solutions as single integer numbers, which facilitates a distributed exhaustive search of solutions with negligible

[6] `https://qarnot.com/en`
[7] `https://anr.fr/Project-ANR-16-CE25-0016`
[8] `https://www.researchgate.net/profile/Philippe-Navaux`

communication, resulting in an embarrassingly parallel algorithm that scales linearly in the function of the number of accelerator devices.

2 List of Publications

In agreement with the research teams I worked with, my publication strategy value high-quality conferences (e.g., Core[9] A-ranking conferences such as SC and CCGrid) and journals (e.g., FGCS, CCPE). As a junior researcher, I tend to prioritize conference publications over journals since conferences can leverage me to establish new scientific connections and collaborations.

When I work with undergraduate and Master's students, I also seek to foster their interaction with the research community and reward their research work by accepting publishing in B-ranking international conferences (e.g., ISCC) and regional conferences and workshops (e.g., CARLA and WSCAD). In this case, the goal is maximizing the student's research experience, giving them the advantage to reach higher research positions, notably by acquiring a PhD student position.

In working with specific research teams, the publications list the authors alphabetically. I spot publications whose authors are sorted by alphabetic order in the below list.

- Link Google Scholar: `https://scholar.google.com/citations?user=HZWnciwAAAAJ&hl=en`
- Link DBLP: `https://dblp.org/pid/165/2057.html`

2.1 Most meaningful publications

In this section I highlight my five most meaningful publications. These publications are sorted in order of importance.

Paper one

[3] **Danilo Carastan-Santos**, and Raphael Y. de Camargo. "*Obtaining dynamic scheduling policies with simulation and machine learning*". In the International Conference for High Performance Computing, Networking, Storage and Analysis (SC), ACM Press, **2017**.

- Core Conference Ranking A (top 16.07% of ranked venues); Low acceptance rates (19% in 2017); more than ten thousand attendees
- Nominated for the **Best Paper** and **Best Student Paper** awards
  - Disclosure of the nomination: `https://team.inria.fr/datamove/889-2/` - Video: `https://youtu.be/Z0kJv9-Zbvs?t=679`
  - In 2017, the SC conference accepted 61 papers, with an acceptance rate of around 19%[10]. Among these 61 papers, five were nominated for the Best Paper Award, and three papers for the Best Student Paper award. The conference nominated my paper as finalist for both awards.
- 34 citations (30 November 2022). Close to 2000 paper downloads, from INRIA's HAL and ACM digital library repositories
- **Pre-print openly available** (`https://hal.inria.fr/hal-01618940`) and **companion material** (`https://github.com/hpcsched/gen-sched-policies`), with **description** (last two pages of the paper)

[9] `http://portal.core.edu.au/conf-ranks/`
[10] `https://dl.acm.org/doi/proceedings/10.5555/3291656`

This publication resulted in Chapter 3 of my thesis manuscript[11], and I made this work under the supervision of Raphael Y. de Camargo. This work pioneered in proposing regression-based methods to generate scheduling heuristics for HPC platforms, creating a novel link between Machine Learning (ML) and simple HPC scheduling heuristics. Previous works used ML in scheduling only to predict the applications' execution time, and not the execution order of the applications. One of the main challenges was how to use ML with scheduling data: exploiting HPC simulation to generate data not present in HPC logs and feeding this data into regression algorithms to create efficient and simple scheduling heuristics. The HPC scheduling heuristics obtained in this work remain among the most efficient heuristics in the literature.

Paper two

[5] **Danilo Carastan Santos**, Raphael Y. de Camargo, Denis Trystram, Salah Zrigui. "*One can only gain by replacing EASY Backfilling: A simple scheduling policies case study.*". 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), **2019 (authors sorted by alphabetic order)**.

- Core Conference Ranking A (top 16.07% of ranked venues); Low acceptance rates (between 20% and 30%)
- **Best Paper Award winner**
  - Certificate: `https://www.dropbox.com/s/t0c22bgjqkeq6ux/Best_Paper_Award.pdf?dl=0`
  - Disclosure of the nomination:
    `https://team.inria.fr/datamove/ccgrid-2019-best-paper/`
  - Each edition of CCGrid awards only one paper as the best paper of the conference.
- 19 citations (30 November 2022)
- **Pre-print openly available**: `https://hal.archives-ouvertes.fr/hal-02237895`

This publication resulted in Chapter 4 of my thesis manuscript[12]. I made this work under supervision of Raphael Y. de Camargo and Denis Trystram, and with collaboration with my former colleague Salah Zrigui. This work's challenge was dealing with the cognitive gap between theory and practice in scheduling jobs in HPC platforms. We then proposed by design only simple scheduling heuristics similar to the ones used in practice. We made statements on properly designing and evaluating scheduling algorithms, also providing convincing experimental results to validate claims. We showed that we could use other scheduling heuristics as simple as the ones used in practice and still have significant performance improvements without losing the simplicity and theoretical guarantees. We advocated for the Shortest estimated Area First (SAF) as a benchmark in the literature, as it is an open challenge to beat SAF by a significant margin.

Paper three

[2] **Danilo Carastan-Santos**, Krzysztof Rzadca, Leonel Sousa, Denis Trystram. "*A community-guided discussion about social and environmental effects of post-COVID-19 Computer Science conferencing* ", **2022 (authors sorted by alphabetic order, working preprint)**.

- **Preprint openly available:** `https://hal.archives-ouvertes.fr/hal-03903632`
- **Companion material openly available:** `https://github.com/danilo-carastan-santos/europar-travel-emissions`

[11] `https://theses.hal.science/tel-02928077`
[12] `https://theses.hal.science/tel-02928077`

This paper is one of the scientific outputs of my second post-doc. In the Computer Science case, we used the Euro-Par conference to combine the knowledge and experience regarding conferencing social and environmental effects post-COVID-19. The challenge of this work was gathering these community experiences to give directions on how to keep the learned lessons post-COVID-19. With reactions from the Euro-Par community, we open the debate to rethink the conference's utility according to our objectives (scientific and ecologic) and to be aware of social/geographical biases are essential factors in participating and organizing post-COVID-19 conferences. We also showed practical results that reinforce the marginal emissions of virtual conferencing compared to in-person conference travel. I was the leading writer of this paper, and I collaborated with processing, analyzing, and interpreting results.

Paper four

[15] Luis Sant'Ana, **Danilo Carastan-Santos**, Daniel Cordeiro and Raphael Y. de Camargo. "*Real-Time Scheduling Policy Selection from Queue and Machine States.*". 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), **2019**.

- Core Conference Ranking A (top 16.07% of ranked venues); Low acceptance rates (between 20% and 30%)
- 8 citations (30 november 2022)

I collaborated with my former colleague Luis Sant'ana's PhD work under the supervision of Raphael Y. de Camargo and Daniel Cordeiro. The methodology I introduced in [3] inspired this work. At the time of this paper, no work had yet to attempt to use logistic regression and support vector machines to predict the best scheduling heuristic from a portfolio. Therefore, the challenges of this work were to assess if (i) logistic regression and support vector machines could predict the best scheduling heuristic to use at specific points in time and (i) what HPC platform and waiting queue data are good features for classification. My contributions were insights about efficient scheduling heuristics to compare and features to use in the classification, notably by comparing with the Shortest estimated Area First (SAF) and using the "area" of the jobs as features in the classification.

Paper five

[4] **Danilo Carastan Santos**, Raphael Y. de Camargo, David C. Martins-Jr, Siang W. Song, and Luiz C.S. Rozante. "*Finding exact hitting set solutions for systems biology applications using heterogeneous GPU clusters.*". Future Generation Computer Systems-The International Journal of eScience, v. 67, p. 418-429, **2016**.

- Core Journal Ranking A (top 10% of ranked venues);
- 15 citations (30 november 2022)

This publication is one of the results of my Master's work. It tackles the problem of creating a large-scale parallel exact algorithm for the NP-Hard Minimal Hitting Set (MHS) problem in the use case of bioinformatics applications. A challenge of this work was that state-of-the-art exact algorithms for MHS employ techniques that compromise parallelism. An exhaustive search algorithm can help parallelization at the hindrance of exponential workload growth. We approached this hindrance by proposing a multi-GPU exhaustive search algorithm for MHS. One of my contributions was to minimize the communication overhead of offloading computations to the GPUs by using a combinatorial numbering system to encode the GPUs' workload, which resulted in linear performance scalability in the function of the number of GPUs.

2.2 Publications in peer-reviewed journals

[11] Valéria S. Girelli, Francis B. Moreira, Matheus S. Serpa, **Danilo Carastan-Santos**, and Philippe OA. Navaux. "*Investigating memory prefetcher performance over parallel applications: From real to simulated.*" Concurrency and Computation: Practice and Experience, **2021 (without PhD supervisor as co-author)**

- **Companion material openly available:** `https://gitlab.com/msserpa/prefetcher-ccpe`

[7] **Danilo Carastan Santos**, David C. Martins-Jr, Siang W. Song, Luiz C.S. Rozante, and Raphael Y. de Camargo. "*A hybrid CPU-GPU-MIC algorithm for minimal hitting set enumeration.*". Concurrency and Computation: Practice and Experience, **2018**.

2.3 Publications in peer-reviewed international conferences

[14] Lucas Rosa, **Danilo Carastan-Santos**, and Alfredo Goldman. "*A multiple linear regression approach for understanding the trade-offs in learning HPC job scheduling heuristics*". 23rd IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), **2023 (under review, without PhD supervisor as co-author)**.

[10] **Danilo Carastan-Santos**, Anderson A. Da Silva, Alfredo Goldman, Angan Mitra, Yanik Ngoko, Clément Mommessin, Denis Trystram"*Short-Term Ambient Temperature Forecasting for Smart Heaters*". 2021 IEEE Symposium on Computers and Communications (ISCC), IEEE, **2021**.

- **Companion material openly available:**
  `https://gitlab.com/andersonandrei/forecasting-smart-heaters-temperature`

[17] Tovar, C.R. Portocarrero, Eloi Araújo, **Danilo Carastan-Santos**, David C. Martins-Jr, and Luiz C.S. Rozante. "*Finding attractors in biological models based on boolean dynamical systems using hitting set.*" In 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE) IEEE, **2019**. **(without PhD supervisor as co-author)**.

[6] **Danilo Carastan Santos**, Raphael Y. de Camargo, David C. Martins-Jr, Siang W. Song, and Luiz C.S. Rozante. "*A multi-GPU hitting set algorithm for GRNs inference.*" 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. IEEE, **2015 (6 citations)**.

[13] Francis B. Moreira, **Danilo Carastan-Santos**, and Philippe OA. Navaux. "Attesting L-3 General Program Anomaly Detection Efficiency with SPADA." 2020 IEEE Symposium on Computers and Communications (ISCC). IEEE, **2020 (without PhD supervisor as co-author)**.

2.4 Publications in peer-reviewed regional conferences

[9] **Danilo Carastan-Santos** and Thi H.T. Pham. *"Understanding the Energy Consumption of HPC Scale Artificial Intelligence."* Latin American High Performance Computing Conference. **2022, (authors sorted by alphabetic order, without PhD supervisor as co-author)**.

- **Source-code available:** `https://github.com/phamthi1812/Benchmark-Tracker`

[1] Matheus W. Camargo, Matheus S. Serpa, **Danilo Carastan-Santos**, Alexandre Carissimi, and Philippe OA. Navaux "*Accelerating Machine Learning Algorithms with TensorFlow Using Thread Mapping Policies.*" Latin American High Performance Computing Conference. Springer, Cham, **2020 (without PhD supervisor as co-author)**.

2.5 Publications in peer-reviewed workshops

[16] Luis Sant'Ana, **Danilo Carastan-Santos**, Daniel Cordeiro, and Raphael Y. de Camargo "*Analysis of Potential Online Scheduling Improvements by Real-Time Strategy Selection.*" Symposium on High Performance Computing Systems (WSCAD). IEEE, **2018**.

[12] Félix Michels, Matheus S. Serpa, **Danilo Carastan-Santos**, Lucas Schnorr, and Philippe OA. Navaux. "*Otimização de Aplicações Paralelas em Aceleradores Vetoriais NEC SX-Aurora.*" Symposium on High Performance Computing Systems (WSCAD), pp. 311-322. SBC, **2020**.

[8] **Danilo Carastan-Santos**, David C. Martins-Jr, Luiz C.S. Rozante, Siang W. Song, Raphael Y. de Camargo. "*A Hybrid CPU-GPU-MIC algorithm for the hitting set problem*", Symposium on High Performance Computing Systems (WSCAD), SBC, **2017**.

3 Visibility

3.1 Prizes and Awards

- **Atos/GENCI Joseph Fourier Prize**. **Danilo Carastan-Santos** and Denis Trystram; **Bezons, France 2022**.
  - Press announce (in French): `https://atos.net/fr/2022/communiques-de-presse_2022_07_06/atos-et-genci-annoncent-les-vainqueurs-du-prix-joseph-fourier-2022`
  - This prestigious prize reward works in advanced computing (supercomputing, quantum computing, edge computing) and artificial intelligence. A jury of independent specialists and representatives of the French scientific and industrial world chooses and rewards several innovative projects. I, conjointly with Professor Denis Trystram, was awarded first place in the "Artificial Intelligence" category. My contribution to our contest submission, entitled *do better with less*, involves improving the operation of High-Performance Computing platforms, particularly by using Machine Learning to optimize the order of the applications to be executed (see Section 1.1).
- **Best PhD Thesis award**; XXI Symposium on High Performance Computing Systems (WSCAD); **Santo André, Brazil 2020**.
  - Certificate (in Portuguese): `https://www.dropbox.com/s/ez3yedrhww5el10/certificado_ctd_1.pdf?dl=0`
  - The contest judges and awards the best Brazilian theses in Computer Architecture and High-Performance Computing each year. It is a high-level scientific excellence contest since all yearly PhD theses in the disciplines above can apply for the contest. My thesis was awarded first place in the 2020 edition of this contest.
- **Best Paper award winner**; CCGRID 2019 - 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, **Larnaca, Cyprus 2019**.
  - See Section 2.1 for details
- **Best Paper and Best Student Paper awards nomination**; SC17 - The International Conference for High Performance Computing, Networking, Storage and Analysis, **Denver, United-States 2017**.
  - See Section 2.1 for details
- **Two times awarded (2015 and 2016)**; **contest of academic excellence of graduate students**; Federal University of ABC, **Santo André, Brazil**.

- Results announcements (in Portuguese):
  - 2015: https://propg.ufabc.edu.br/wp-content/uploads/3.-Resultado_1%C2%AA-fase_site.pdf
  - 2016: https://propg.ufabc.edu.br/wp-content/uploads/Resultado_II_PEAPG_final.pdf
- This contest aims to stimulate excellence in the research and development production at UFABC. This contest awarded me second place in 2015 and first in 2016, both in the Computer Science category.

3.2 Invited Presentations

- "*Measuring the energy consumption of AI*". **Danilo Carastan-Santos**. **2nd Inria-DFKI European Summer School on AI (IDESSAI 2022), Saarbrücken, Germany. 2022**[13]
  - Course material openly available: https://github.com/danilo-carastan-santos/ai-energy-consumption
  - I was an invited speaker at the track "Sustainable AI". I prepared a tutorial on measuring the energy consumption of Artificial Intelligence code. I first explained the different kinds of energy measurement methods (hardware and software), then how to instrument AI code with popular energy measurement software, and I showed how we could use energy-related information during AI model training.
- "*To travel or not to travel at Euro-Par, that is the question*". **Danilo Carastan-Santos**, Krzysztof Rzadca, Leonel Sousa and Denis Trystram. **International European Conference on Parallel and Distributed Computing (Euro-Par), Glasgow, United-Kingdom 2022**.
  - Invited presentation at the Miscellaneous session.
  - I was in charge of the presentation content, notably the emission estimations of Euro-Par conference travel *versus* virtual. The talk was given by Denis Trystram, on behalf of all authors.
- "*Towards AI frugality by Edge Distributed Computing and Simulation*". Danilo Carastan-Santos. **Franco-german research and innovation network on AI. INRIA Rocquencourt, France 2022**.
  - Invited presentation at the "Resource aware AI" session.
- *"Entender e medir o consumo energético de aplicações de Inteligência Artificial"*. Danilo Carastan-Santos. ***II Escola Regional de Alto Desempenho/Aprendizado de Máquina e Inteligência Artificial Norte 2***. Belém, Brazil[14].
  - I was invited to give a Portuguese version of my course, "*Measuring the energy consumption of AI*" (see above), in a Brazilian regional school. Brazilian regional schools rarely cover the subject of energy consumption awareness of computing and AI. I introduced this awareness of computing and AI's energy and environmental costs to the students.

4 Collaborations

- Participant of the Edge Intelligence chair[15] of Multidisciplinary Institute in Artificial Intelligence (MIAI Grenoble-Alpes)[16]

[13] https://idessai.eu/track-b-sustainable-ai-2022/
[14] https://www2.sbc.org.br/erad_eramia-no2/programa.html
[15] https://edge-intelligence.imag.fr/index.php#
[16] https://anr.fr/ProjetIA-19-P3IA-0003

  - I am a participant of the chair since the beginning of my second post-doc. My work so far involves measuring and estimating the energy consumption and CO2 emissions of distributed platforms, and mentoring of Master's and PhD students. We have so far one paper [2] and one regional conference publication [9] as outcomes of this participation.

- Scientific collaborator with University of São Paulo (USP, Brazil), in the project "Trends on high performance computing, from resource management to new computer architectures[17]".

  - I collaborate with this project in the Machine Learning applied to High-Performance computing resource management by co-supervising with Alfredo Goldman, the student Lucas Rosa. We already have one peer-reviewed international conference submission [14] as an outcome of this collaboration. This collaboration is ongoing.

- Participant in the Distributed Processing Group (GPPD) of the Informatics Institute (INF) of the Federal University of Rio Grande do Sul (UFRGS, Brazil)[18]

  - As one of the main duties of my first post-doc, I collaborated with Undergraduate, Master's, and PhD students by advising and transferring my Machine Learning and Data Science expertise into their works. This collaboration resulted in one peer-reviewed international journal publication [11], one publication in a peer-reviewed international conference [13] and two publications in regional conferences and workshops [1, 12].

- Scientific collaborator of the ANR project "Resource manager for cloud of Things" (GRECO)[19]

  - I was a scientific collaborator of this project, as a parallel task during my PhD thesis work. I contributed by working on temperature-aware simulation of distributed infrastructures, and by advising the student Anderson Andrei da Silva. We have one peer-reviewed international conference publication [10] as an outcome of this collaboration.

- Scientific collaborator of the laboratory *Metodologias e Técnicas de Computação* (LMTC)[20] of the Federal University of ABC (UFABC, Brazil).

  - I was a scientific collaborator of this research laboratory, as a parallel task during my PhD thesis work. I collaborated with the PhD work of my former colleague Lucas Sant'ana by helping to use Machine Learning models and setting up simulation experiments. I also contributed during the Master's work of Carlos Reynaldo Portocarrero Tovar, in which I helped to transfer my proposed Hitting Set algorithms (see Section 1.3.2) to his research context. We have two peer-reviewed international conference publications [15, 17] and one peer-reviewed regional workshop [16] as outcomes of this collaboration.

5 Research Administration

5.1 Organizing committee, and program committee

- Organizing committee of *Journee sur la Recherche en Apprentissage Frugal* 2022[21]

  - I was in charge of the workshop call in mailing lists and amphitheater preparations.

[17] https://bv.fapesp.br/en/auxilios/108817/trends-on-high-performance-computing-from-resource-management-to-new-computer-architectures/
[18] https://www.inf.ufrgs.br/gppd/site/
[19] https://anr.fr/Project-ANR-16-CE25-0016
[20] https://poscomp.ufabc.edu.br/pesquisa/laboratorios/
[21] https://edge-intelligence.imag.fr/workshop/index.php#mu-about

- Organizing committee of New Challenges in Scheduling Theory workshop[22]
  - I was in charge of the audiovisual management (i.e., projector, microphone, and videoconference preparations) and badge preparation.
- Program committee of the Symposium on High Performance Computing Systems (WSCAD) 2022[23]
  - This task involved reviewing a selection of paper. I reviewed a total of three papers.

5.2 Journal Reviewer

- Future Generation Computer Systems: Certificate: `https://www.dropbox.com/s/fhbc6961jc5wilv/Certificate_FUTURE_Recognised.pdf?dl=0`
- Parallel Computing: Certificate: `https://www.dropbox.com/s/okbxahdjf0nyli1/Certificate_PARCO_Recognised.pdf?dl=0`
- Journal on Parallel and Distributed Computing: Certificate: `https://www.dropbox.com/s/v91nf8rumu9u1ib/Certificate_YJPDC_Recognised.pdf?dl=0`
- Concurrency and Computation: Practice and Experience: certificate: `https://www.dropbox.com/s/4p0upd2ci7q1rgv/CPE_Reviewer_Certificate_2021.pdf?dl=0`

5.3 International Conference Reviewer

- **IEEE Cluster 2020**[24]: IEEE Cluster Conference
- **ICS 2020**[25]: International Conference on Supercomputing
- **SBAC-PAD 2020**[26]: International Symposium on Computer Architecture and High Performance Computing
- **Euro-Par 2019**[27]: International European Conference on Parallel and Distributed Computing
- **SBAC-PAD 2019**[28]: International Symposium on Computer Architecture and High Performance Computing
- **Euro-Par 2018**[29]: International European Conference on Parallel and Distributed Computing

6 Supervision, teaching

6.1 Research supervision

- Anderson Andrei da Silva[30](Master's student): Final Undergraduate Project[31]: Big Data Driven temperature control of Intelligent Heaters (co-advised with Alfredo Goldman and Denis Trysrtam). Current position: PhD student at University Grenoble-Alpes. Resulting publications: [10].

[22] `http://aussois2022.imag.fr/`
[23] `https://wscad.ufsc.br/chamada-trilha-principal/`
[24] `https://clustercomp.org/2020/`
[25] `https://ics2020.bsc.es/`
[26] `https://sbac2020.dcc.fc.up.pt/`
[27] `https://2019.euro-par.org/`
[28] `https://www.computer.org/csdl/proceedings/sbac-pad/2019/1fHkqCE6iQg`
[29] `https://europar2018.org/`
[30] `https://andersonandrei.github.io/`
[31] `https://andersonandrei.github.io/Anderson%20Andrei%20-%20TCC.pdf`

- Valéria Soldera Girelli[32] (Undergraduate student): Investigating memory prefetcher performance over parallel applications: From real to simulated (co-advised with Matheus Serpa, Francis Moreira and Philippe Navaux). Current Position: Research Engineer at Barcelona Supercomputing Center - BSC. Resulting publications: [11].
- Lucas de Sousa Rosa (Undergraduate student): On limits of Machine Learning techniques in the learning of scheduling policies (co-advised with Alfredo Goldman, funded by the Brazilian Government[33]). Resulting publications: [14].
- Ning Tang (M1 student): Big Data Driven temperature control of Intelligent Heaters (co-advised with Clément Mommessin and Denis Trystram). Current Position: PhD student at LIP6.
- Thi Hoang Thi Pham (M1 student): Understanding the energy consumption of Artificial Intelligent algorithms (co-advised with Denis Trystram). Resulting publications: [9].
- Félix Dal Pont Michels Jr. (Master's student): Optimization and adaptation of applications for vector processors (co-advised with Philippe Navaux). Resulting publications: [12].
- Matheus Camargo (Undergraduate student): Accelerating Machine Learning Algorithms with TensorFlow using Thread Mapping Policies (co-advised with Matheus Serpa and Philippe Navaux). Resulting publications: [1].

6.2 Teaching

- **Part-time Teacher (French)**; *Systèmes d'Exploitation et Programmation Concurrente* (20 Undergraduate students), University Grenoble-Alpes, **2022**.
- **Part-time Teacher (English)**; Technical Writing and Speaking in English (30 Master students), University Grenoble-Alpes, **2021 and 2022**.
- **Part-time Teacher (French)**; *Systèmes d'Exploitation et Réseaux* (20 Undergraduate students), University Grenoble-Alpes, **2021**.
- **Part-time Teacher (French)**; *Algorithmique Avancée* (20 Undergraduate students), University Grenoble-Alpes, **2018**.
- **Part-time Teacher (English)**; Algorithmic Problem Solving (20 Master students), University Grenoble-Alpes, **2018**.
- **Part-time Teacher (French)**; *Agorithmique et Programmation Fonctionnelle* (30 Undergraduate students), University Grenoble-Alpes, **2017**.

References

[1] Matheus W. Camargo, Matheus S. Serpa, Danilo Carastan-Santos, Alexandre Carissimi, and Philippe O. A. Navaux. Accelerating machine learning algorithms with tensorflow using thread mapping policies. In Sergio Nesmachnow, Harold Castro, and Andrei Tchernykh, editors, *High Performance Computing - 7th Latin American Conference, CARLA 2020, Cuenca, Ecuador, September 2-4, 2020, Revised Selected Papers*, volume 1327 of *Communications in Computer and Information Science*, pages 62–70. Springer, 2020.

[32] https://scholar.google.com/citations?user=syCkK-MAAAAJ&hl=en

[33] https://bv.fapesp.br/en/bolsas/202861/on-limits-of-machine-learning-techniques-in-the-learning-of-scheduling-policies/

[2] Danilo Carastan dos Santos, Krzysztof Rzadca, Leonel Sousa, and Denis Trystram. A community-guided discussion about social and environmental effects of post-COVID-19 Computer Science conferencing. preprint, January 2023.

[3] Danilo Carastan-Santos and Raphael Y. de Camargo. Obtaining dynamic scheduling policies with simulation and machine learning. In Bernd Mohr and Padma Raghavan, editors, *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC 2017, Denver, CO, USA, November 12 - 17, 2017*, page 32. ACM, 2017.

[4] Danilo Carastan-Santos, Raphael Y. de Camargo, David Correa Martins Jr., Siang W. Song, and Luiz C. S. Rozante. Finding exact hitting set solutions for systems biology applications using heterogeneous GPU clusters. *Future Gener. Comput. Syst.*, 67:418–429, 2017.

[5] Danilo Carastan-Santos, Raphael Y. de Camargo, Denis Trystram, and Salah Zrigui. One can only gain by replacing EASY backfilling: A simple scheduling policies case study. In *19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGRID 2019, Larnaca, Cyprus, May 14-17, 2019*, pages 1–10. IEEE, 2019.

[6] Danilo Carastan-Santos, Raphael Yokoingawa de Camargo, David Correa Martins, Siang Wun Song, Luiz Carlos Silva Rozante, and Fabrizio Ferreira Borelli. A multi-gpu hitting set algorithm for grns inference. In *15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid 2015, Shenzhen, China, May 4-7, 2015*, pages 313–322. IEEE Computer Society, 2015.

[7] Danilo Carastan-Santos, David Correa Martins Jr., Siang W. Song, Luiz C. S. Rozante, and Raphael Y. de Camargo. A hybrid CPU-GPU-MIC algorithm for minimal hitting set enumeration. *Concurr. Comput. Pract. Exp.*, 31(18), 2019.

[8] Danilo Carastan-Santos, David C Martins-Jr, Luiz CS Rozante, Siang W Song, and Raphael Y de Camargo. A hybrid cpu-gpu-mic algorithm for the hitting set problem. In *Anais do XVIII Simpósio em Sistemas Computacionais de Alto Desempenho*, pages 196–207. SBC, 2017.

[9] Danilo Carastan-Santos and Thi Hoang Thi Pham. Understanding the energy consumption of HPC scale artificial intelligence. In Philippe O. A. Navaux, Carlos J. Barrios, Carla Osthoff, and Ginés D. Guerrero, editors, *High Performance Computing - 9th Latin American Conference, CARLA 2022, Porto Alegre, Brazil, September 26-30, 2022, Revised Selected Papers*, volume 1660 of *Communications in Computer and Information Science*, pages 131–144. Springer, 2022.

[10] Danilo Carastan-Santos, Anderson Andrei Da Silva, Alfredo Goldman, Angan Mitra, Yanik Ngoko, Clément Mommessin, and Denis Trystram. Short-term ambient temperature forecasting for smart heaters. In *IEEE Symposium on Computers and Communications, ISCC 2021, Athens, Greece, September 5-8, 2021*, pages 1–6. IEEE, 2021.

[11] Valéria Soldera Girelli, Francis B. Moreira, Matheus S. Serpa, Danilo Carastan-Santos, and Philippe O. A. Navaux. Investigating memory prefetcher performance over parallel applications: From real to simulated. *Concurr. Comput. Pract. Exp.*, 33(18), 2021.

[12] Félix Michels, Matheus Serpa, Danilo Carastan-Santos, Lucas Schnorr, and Philippe Navaux. Otimização de aplicações paralelas em aceleradores vetoriais nec sx-aurora. In *Anais do XXI Simpósio em Sistemas Computacionais de Alto Desempenho*, pages 311–322. SBC, 2020.

[13] Francis B. Moreira, Danilo Carastan-Santos, and Philippe O. A. Navaux. Attesting L-3 general program anomaly detection efficiency with SPADA. In *IEEE Symposium on Computers and Communications, ISCC 2020, Rennes, France, July 7-10, 2020*, pages 1–6. IEEE, 2020.

[14] Lucas Rosa, Danilo Carastan-Santos, and Alfredo Goldman. A multiple linear regression approach for understanding the trade-offs in learning hpc job scheduling heuristics. In *IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGRID*, 2023, *submitted for peer-review.*

[15] Luis Sant'Ana, Danilo Carastan-Santos, Daniel Cordeiro, and Raphael Y. de Camargo. Real-time scheduling policy selection from queue and machine states. In *19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGRID 2019, Larnaca, Cyprus, May 14-17, 2019*, pages 381–390. IEEE, 2019.

[16] Luis Felipe Sant'Ana, Danilo Carastan-Santos, Daniel Cordeiro, and Raphael Y. de Camargo. Analysis of potential online scheduling improvements by real-time strategy selection. In *Symposium on High Performance Computing Systems, WSCAD 2018, São Paulo, Brazil, October 1-3, 2018*, pages 1–7. IEEE, 2018.

[17] Carlos Reynaldo Portocarrero Tovar, Eloi Araujo, Danilo Carastan-Santos, David Correa Martins, and Luiz C. S. Rozante. Finding attractors in biological models based on boolean dynamical systems using hitting set. In *19th IEEE International Conference on Bioinformatics and Bioengineering, BIBE 2019, Athens, Greece, October 28-30, 2019*, pages 235–239. IEEE, 2019.