EPJ Web of Conferences **337**, 01147 (2025) https://doi.org/10.1051/epjconf/202533701147
CHEP 2024

Experience with ARM WNs at the WLCG Tier1 GridKa

Armin Alfredo Krull[1,*], *Matthias Jochen* Schnepf[1,**], *Max* Fischer[1,***], and *Andreas* Petzold[1,****]

[1]Karlsruhe Institute of Technology

Abstract. Computing centers are always looking for new worker nodes (WNs) that can reduce operational costs, especially energy consumption, and provide higher performance. ARM CPUs promise higher power efficiency than x86 CPUs. The WLCG Tier1 center GridKa will therefore partially use WNs with ARM CPUs. It has already conducted various energy consumption and performance benchmarkings for several such ARM WNs based on the benchmark HEPScore23 as well as a CPU frequency scan to determine their best performance and power efficiency. We present the results of these benchmarkings in comparison to those of the x86 WNs currently operated at GridKa and summarize the progress made at GridKa in providing ARM WNs to the HEP community.

1 Introduction

For computing grids that process high-energy physics (HEP) data, such as the Worldwide LHC Computing Grid (WLCG), energy consumption is a significant operational cost [1]. The computing centers in the WLCG are therefore always looking for worker nodes (WNs) with CPUs that offer a better power efficiency.

In recent years, the question of whether ARM or x86 CPUs are the better choice in terms of power efficiency has been subject to increasing discussion. For applications in mobile devices, several comparative benchmarking studies for ARM and x86 CPUs of different generations have already been published, which mainly came to the conclusion that the ARM CPUs have a better power efficiency than their x86 competitors [2–6]. In 2020, the discussion gained further momentum triggered by Apple's decision to switch the architecture of its CPUs from x86 to ARM. Apple communicated its step as an advantage in terms of power efficiency [7]. Several comparative benchmarking studies for Apple's ARM-based M1 CPUs and its x86 competitors at the time, as well as x86 CPUs previously used by Apple, have confirmed that the change has significantly improved power efficiency [5, 8, 9].

In this context, the WLCG has also become increasingly interested in clarifying whether WNs with ARM CPUs can help to improve power efficiency in HEP computing. The Scottish Grid Service (ScotGrid) has already conducted several benchmarking studies comparing x86 and ARM CPUs since 2022 [10, 11]. Their findings show that the ARM CPU under consideration had significantly lower power usage and energy consumption and often slightly

*e-mail: armin.krull@student.kit.edu
**e-mail: matthias.schnepf@kit.edu
***e-mail: max.fischer@kit.edu
****e-mail: andreas.petzold@kit.edu

better performance than its x86 competitors, resulting in better performance and power efficiency overall. Due to these promising results and the intention to provide WNs with ARM CPUs to the HEP community in the future, WLCG's Tier 1 Grid Computing Centre Karlsruhe (GridKa) purchased several ARM WNs at the end of 2023 to have ARM resources for test experiences with data from the ATLAS, CMS, ALICE, LHCb, and BELLE 2 experiments. With these WNs various benchmarkings were conducted for comparisons with x86 WNs of the same thread count that are currently operated at GridKa and were purchased in 2022. The experiences gained from these benchmarks were used to provide ARM WNs that meet the computing needs of the HEP community in the GridKa batch farm.

2 Comparisons

2.1 Benchmarked WNs

Table 1. WN configurations (Conf.) under comparison

Conf.	CPU model	Threads	CPU frequency	Ref.
x86-1	AMD EPYC 7742	256	2.25 GHz (default)	[12]
x86-2	AMD EPYC 7702	256	2.0 GHz (default)	[13]
ARM-A	Ampere Altra M128-30	256	3.0 GHz (default)	[14]
ARM-B	Ampere Altra M128-30	256	2.2 GHz (see section 2.4)	[14]

All WNs are dual-socket systems with all eight memory channels per socket populated. This results in at least 2.5 GB RAM per thread, which meets the standard requirement for the benchmark HEPScore23 (HS23). RAM bottlenecks affecting the test results can thus be excluded.

2.2 Metrics

The metrics applied in the comparisons were:

- The HS23 score as a dimensionless quantity for the CPU performance, calculated with the HEPScore23 in the HEP Benchmark Suite, which includes 7 workloads (WLs) from the 5 experiments listed in the introduction [16].
- The average power usage in W calculated from the power values of the WN logged by IPMItool [17] in the continuous power reading mode every 10 seconds during the benchmarking runtime. The average power usage was calculated analogous to Kozik 2024 using the 85% quantile [18].
- The power efficiency in 1/W as ratio of HS23 score vs. average power usage.
- The CPU frequency in GHz, as logged by IPMItool every 10 seconds during the benchmarking runtime.
- The CPU temperature in °C, calculated as the mean of the maximum CPU temperatures across all benchmarking runs considered in a comparison. The maximum CPU temperatures were taken from the temperature values logged by IPMItool for each run every 10 seconds during the runtime.

2.3 HS23 Scores and Power Efficiencies

The power efficiencies of the x86 configurations under comparison were determined using HS23 score, average power usage, and benchmarking runtime data from the HEPiX Benchmarking Working Group database [15]. These data were initially generated in benchmarking runs carried out by GridKa between November 2023 and February 2024 using the HEP Benchmark Suite with HEPScore23 v1.5 and IPMItool v1.8.19.

Calculations of the arithmetic means and standard deviations for the HS23 score, average power usage, and total energy consumption of the x86-1 and the x86-2 WNs were based on the data from 36, and 15 available benchmarking runs respectively.

The power efficiencies of the ARM configurations under comparison were calculated using HS23 score, average power usage and benchmarking runtime data generated in benchmarking runs carried out by the authors of this study between April 2024 and January 2025 using the HEP Benchmark Suite with HEPScore23 v2.0rc8 and IPMItool v1.8.19. Calculations of the means and standard deviations for the ARM-a and ARM-b configurations were based on 60 runs, and 10 runs respectively.

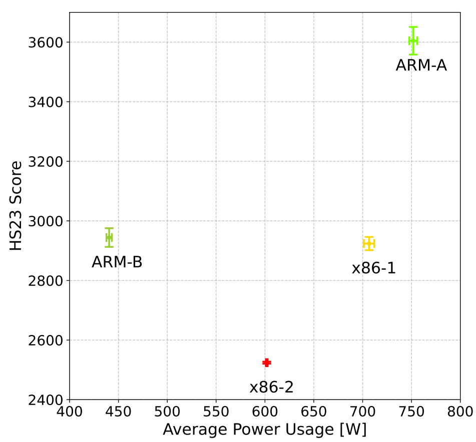

Figure 1. HS23 score vs. average power usage for all WN configurations under comparison. The HS23 score for ARM-a is significantly higher than that of the x86 WNs. Since its average power usage also increases significantly, the power efficiency does not improve to the same extent.

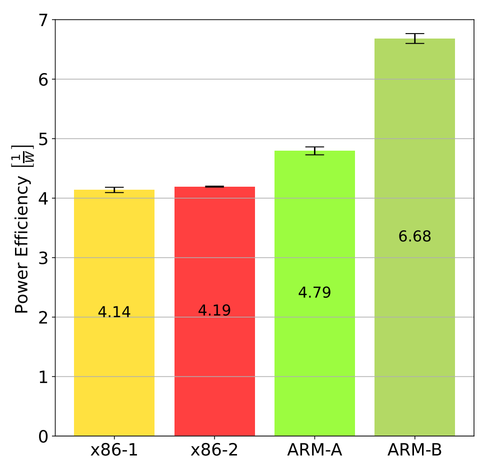

Figure 2. Power efficiencies for all WN configurations under comparison.The power efficiency of the ARM WN exceeds that of the x86-2 WN, which is the x86 WN with the best power efficiency, by 14% for ARM-a and 59% for ARM-b.

Table 2. HS23 score, average power usage (APU), total energy consumption (TEC), power efficiency (PE), and average benchmarking runtime (ABR) of the WN configurations.

Conf.	HS23 score	APU [W]	TEC [kWh]	PE [1/W]	ABR [h]
x86-1	2924±22	706.6±5.3	3.49±0.11	4.14±0.04	5.66±0.35
x86-2	2524.0±3.3	601.87±0.88	3.38±0.03	4.19±0.01	6.13±0.06
ARM-a	3605±46	751.8±4.0	2.97±0.02	4.79±0.07	4.83±0.03
ARM-b	2944±31	440.5±2.8	2.23±0.01	6.68±0.08	5.86±0.03

Table 2 and the plots of the HS23 score vs. average power usage (figure 1) and the power efficiencies (figure 2) show that ARM-A has a significantly better HS23 score than x86-2, which is the x86 configuration with the best power efficiency. The HS23 score has increased by 43%. This improvement, however, comes at the cost of 25% higher average power usage. The improvement in power efficiency is thus 14%.

2.4 Frequency Scan

For the ARM WN we also determined the HS23 score and average power usages for CPU frequencies from 1.2 to 3.0 GHz in 0.2 GHz steps (figure 3). For each frequency, 10 benchmarking runs analogous to section 2.3 were carried out. The aim of this benchmarking was to identify the frequency at which the power efficiency is at its maximum.

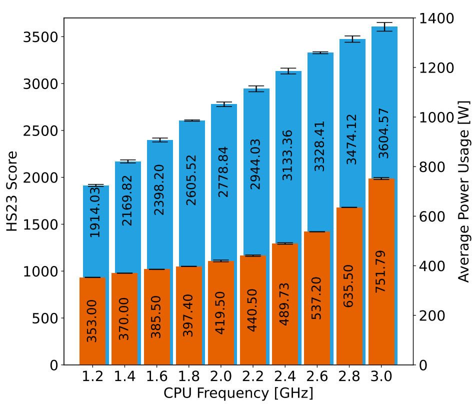

Figure 3. HS23 score (upper bars, left axis) and average power usages (lower bars, right axis) for the ARM WN at CPU frequencies from 1.2 to 3.0 GHz in 0.2 GHz steps. While the HS23 score follows the frequency with a constant gradient, the gradient of the average power usage steepens exponentially.

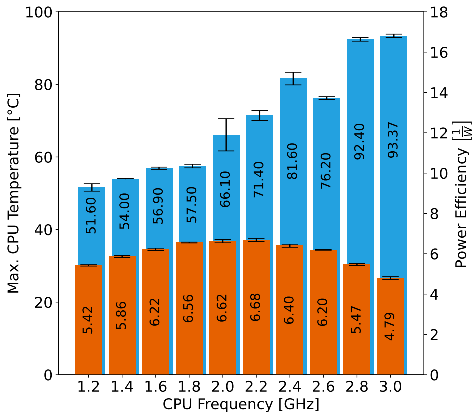

Figure 4. Maximum CPU temperature (upper bars, left axis) and power efficiency (lower bars, right axis) for the ARM WN at CPU frequencies from 1.2 to 3.0 GHz in 0.2 GHz steps. The power efficiency has a maximum at 2.2 GHz.

The plot for the power efficiency vs. CPU frequency shows a maximum at the frequency of 2.2 GHz (figure 4). This characteristic is expected with almost all CPUs [16, 18]. The plots of the HS23 score vs. CPU frequency and the average power usage vs. CPU frequency show that the power efficiency maximum at 2.2 GHz results from their divergent behavior (figure 3). The HS23 score follows the frequency with a constant gradient. This is not the case for the average power usage. Its gradient steepens exponentially.

2.5 Individual Workloads

For a subset of the data obtained from the benchmarking runs conducted in section 2.3 with all four WN configurations, additional comparisons of the performance and efficiencies were carried out for the 7 WLs included in the HEPScore23. For each WN configurations, 10 runs of data were taken. The aim of the comparisons was to find out whether the individual WLs show different performances and power efficiencies depending on the WN configuration and CPU architecture, respectively. The metrics WL runtime, WL score (WLS), WL average power usage and WL power efficiency correspond analogously to the metrics defined in section 2.2. In addition, the following metric was applied: The relative workload performance as a dimensionless quantity that indicates how well the WL performs per HS23 score and thus how suitable the WN is for executing the WL. The quantity is given by the following equation:

$$\text{Relative Workload Performance} = \frac{\text{WL Score} \cdot HS23_{\text{ref}}\ \text{Score}}{\text{HS23 Score} \cdot WL_{\text{ref}}\ \text{Score}} \quad (1)$$

where $HS23_{ref}$ and WL_{ref} are the reference scores of the reference machine specified in the HEPScore23.

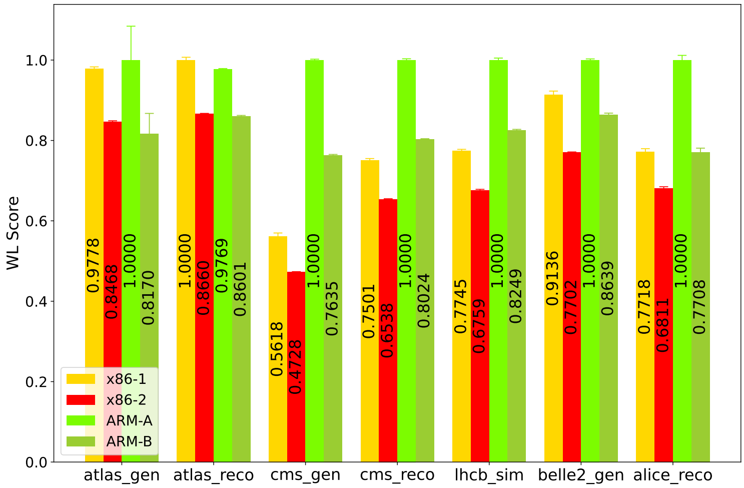

Figure 5. Comparison of the WL scores determined by the HEP Benchmark Suite for the 7 WLs included in the HEPScore23 from the benchmarking runs conducted in section 2.3. The WL scores were normalized to the maximum value of the respective WL for easier comparison.

The plot for the WL Scores (figure 5) shows that the WLs of the ARM-A configuration, with one exception, perform better than those of the x86 configurations. The performance of the ARM-A WLs is on average 18% better compared to those of the x86-1 WN, whose WLs consistently perform better than those of the x86-2 WN. This result is in line with those of Simili et al. 2022 [10]. Their comparisons of two x86 configurations (single-socket AMD EPYC 7643 2.3GHz, dual-socket AMD EPYC 7513 2.6GHz) with an ARM WN (single-socket Ampere Altra Q80-30 3.0GHz) also showed that the performance of the WLs of the ARM WN was significantly better. They found an average improvement in performance of 20%.

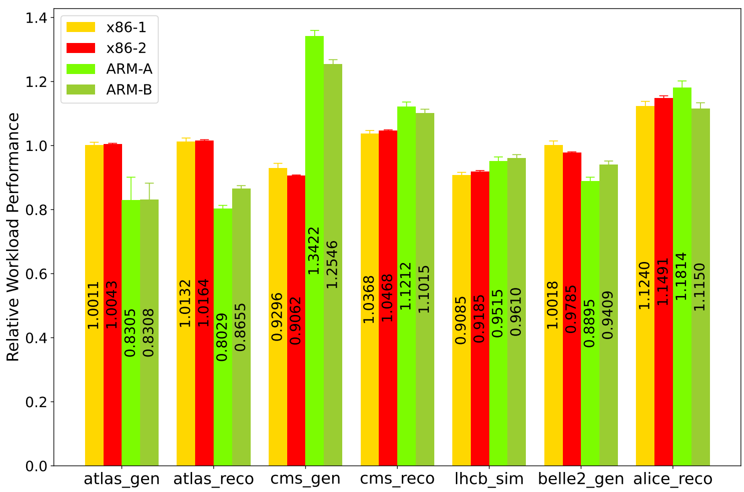

Figure 6. Comparison of relative performance for the individual WL for all WN configurations under comparison. Ratios greater than 1 indicate that the WN configuration is well suited for the WL. It shows that the x86 WNs are better suited for the WLs of the ATLAS and BELLE-2 experiments and the ARM WN for the WLs of the CMS and LHCb experiments.

The plot for the relative performance (figure 6) of the individual WLs shows that the x86 WNs perform better than the ARM WN for the WLs of the ATLAS and BELLE 2 experiments. Vice versa, the ARM WN performs better than the x86 WN for the WLs of the CMS and LHCb experiments. For the ALICE experiment, there is no clear favorite.

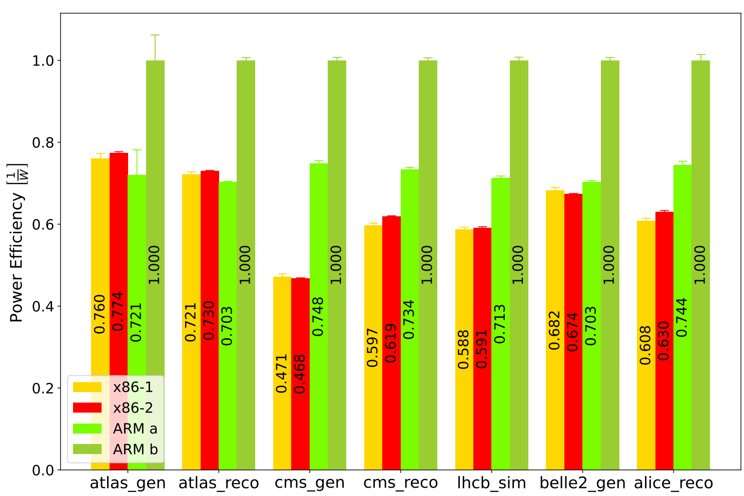

Figure 7. Comparison of the power efficiency for the individual WLs. The values were normalized to the maximum value of the respective WL for easier comparison. It shows that the WLs of the ARM WN, predominantly those of the ARM-B configuration, are significantly more efficient than those of the x86 WNs.

The plots for the power efficiency (figure 7) show that the WLs of the ARM WN, predominantly those of the ARM-B configuration, are significantly more efficient than those of the x86 configurations. The power efficiency of the ARM-B WLs are on average 35% better, respectively, than those of the x86-1 WN, whose WLs are consistently more efficient than those of the x86-2 WN.

3 Conclusion, Outlook

The results of our benchmarkings show that the ARM WN tested has advantages over the x86 WNs currently in operation at GridKa in terms of performance, as well as power efficiency. GridKa therefore concludes that it is attractive, in particular for some WLs, to provide computing resources with ARM WNs on a larger scale.

Following the results of this study, GridKa has committed to providing ARM WNs in a production environment in order to foster the experience of both local and experiment experts. Since June 2024, GridKa has integrated 12 ARM WNs in its production batch farm, making them available via the standard GridKa Compute Elements. The CMS and ALICE collaborations have successfully run some validation WLs on these ARM WNs. The ATLAS collaboration also performed some validation WLs and is using the ARM WNs in production for different WLs since then.

References

[1] D. Britton, S. Campana & B. Panzer-Stradel, A holistic study of the WLCG energy needs for the LHC scientific program, EPJ Web of Conferences **295**, 04001 (2024). https://doi.org/10.1051/epjconf/202429504001

[2] R. V. Aroca & L. M. G. Gonçalves, Towards green data centers: A comparison of x86 and ARM architectures power efficiency, J. Parallel Distrib. Comput. **72**, 1770–1780 (2012). https://doi.org/10.1016/j.jpdc.2012.08.005

[3] E. Blem, J. Menon & K. Sankaralingam, Power struggles: Revisiting the RISC vs. CISC debate on contemporary ARM and x86 architectures, 2013 IEEE 19th International Symposium on High Performance Computer Architecture (HPCA), 1-12 (2013). https://doi.org/10.1109/HPCA.2013.6522302

[4] J. Maqbool, S. Oh & G. C. Fox, Evaluating ARM HPC clusters for scientific workloads, Concurrency Computation: Practice and Experience **27**, 5390-5410 (2015). https://onlinelibrary.wiley.com/doi/10.1002/cpe.3602

[5] K. Gupta & T. Sharma, Changing trends in computer architecture: A comprehensive analysis of ARM and x86 processors, International Journal of Scientific Research in Computer Science, Engineering and Information Technology **7**, 619-631 (2021). https://ijsrcseit.com/paper/CSEIT2173188.pdf

[6] D. Kumar et al., Performance evaluation of ARM-based versus x86-based processors in high performance computing clusters, Journal of Independent Studies and Research Computing **21**, 32-41 (2023). https://jisrc.szabist.edu.pk/ojs/index.php/jisrc/article/view/58/146

[7] Apple, Apple unleashes M1, Press Release, (2020). https://www.apple.com/newsroom/2020/11/apple-unleashes-m1/

[8] B. Funk, Mac mini 2020 review: Apple M1 silicon performance deep dive, HotHardware, (2020). https://hothardware.com/reviews/mac-mini-2020-apple-silicon-m1

[9] J. Salter, Hands-on with the Apple M1 - a seriously fast x86 competitor, Ars Technica, (2020). https://arstechnica.com/gadgets/2020/11/hands-on-with-the-apple-m1-a-seriously-fast-x86-competitor/

[10] E. Simili et al., Power efficiency in HEP (x86 vs. ARM), To be published in the Proceedings of the 21st International Workshop on Advanced Computing and Analysis Techniques (ACAT) in Physics Research Conference, (2022). https://indico.cern.ch/event/1106990/papers/4991256/files/11720-PoW_ESimili.pdf

[11] E. Simili et al., ARMing HEP for the future, EPJ Web of Conferences **295**, 11007 (2024). https://www.epj-conferences.org/articles/epjconf/pdf/2024/05/epjconf_chep2024_11007.pdf

[12] AMD, EPYC 7742 Datasheet. https://www.amd.com/en/support/downloads/drivers.html/processors/epyc/epyc-7002-series/amd-epyc-7742.html#amd_support_product_spec

[13] AMD, EPYC 7702 Datasheet. https://www.amd.com/en/support/downloads/drivers.html/processors/epyc/epyc-7002-series/amd-epyc-7702.html#amd_support_product_spec

[14] Ampere, Ampere Altra M128-30 Datasheet. https://amperecomputing.com/assets/Altra_Max_Rev_A1_DS_v1_25_20240130_73cfcc518a_4705c00046.pdf

[15] HEPiX Benchmarking Working Group. https://w3.hepix.org/benchmarking.html

[16] N. Szczepanek et al., HEP Benchmark Suite: Enhancing efficiency and sustainability in worldwide LHC computing infrastructures, arXiv, 2408.12445 (2024). https://arxiv.org/pdf/2408.12445

[17] D. Laurie, IPMItools: Intelligent Platform Management Interface (2024), https://github.com/ipmitool/ipmitool

[18] K. Kozik, Power consumption time series: Comparison of various aggregation metrics, Presentation, (2024). https://indico.cern.ch/event/1433496/contributions/6031430/attachments/2917567/5120217/comparison_of_various_aggregation_metrics_kacper_kozik.pdf