ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUAI01

HIGH-THROUGHPUT DPDK-BASED FRAMEWORK FOR REAL-TIME APPLICATIONS IN ELETTRA 2.0

G. Gaio†, A. Bogani, G. Brajnik, R. De Monte, G. Scalamera, I. Trovarelli
Elettra-Sincrotrone Trieste S.C.p.A., Trieste, Italy

Abstract

The Data Plane Development Kit (DPDK) is a framework that enhances real-time communications by providing direct, high-speed access to network interfaces. This architecture centralizes acquisition and control in an HPC cluster, ensuring ultra-fast in-memory updates of all critical data, making it a viable choice for real-time feedback and machine control in particle accelerators.

This approach was chosen for Elettra 2.0 to enable premortem beam dump mitigation, more detailed post-mortem inspection, advanced correlation analysis and the implementation of complex control schemes. Time-sensitive applications implemented in C code interact with Beam Position Monitors (BPM), Low-Level RF (LLRF), Beam Loss Monitors (BLM) and Magnetic Power Supplies (PS) through simple memory read and write operations at megahertz rates, making the system competitive with high-level processing applications based on FPGA architectures.

This paper presents the system under test before the decommissioning of Elettra, highlighting its architecture, performance, and integration, while demonstrating the successful implementation of Fast Orbit Feedback in parallel with Turn-by-Turn (TbT) data acquisition from 7 BPM electronics and one LLRF system.

INTRODUCTION

Particle accelerators such as synchrotrons require precise real-time control to ensure beam stability, optimize performance, and avoid costly disruptions. Elettra, a third-generation light source in Trieste, is being upgraded to Elettra 2.0, a fourth-generation diffraction-limited storage ring aimed at higher brightness, coherence, and stability, demanding real-time systems capable of handling data at MHz rates.

The gradual adoption of DPDK at both FERMI and Elettra has replaced legacy real-time stacks such as Xenomai and RTAI, while building expertise in using DPDK as a scalable platform whose performance grows with CPU core counts and NIC bandwidths. At the same time, demand for online, high-frequency data has increased, exposing the complexity and rigidity of traditional schemes based on hardware triggers, local buffers, and offline processing.

With DPDK, TbT (1.15 MHz) and other high-rate signals can instead be handled as streaming data, processed in-memory in real-time for filtering, decimation, and analysis. This reduces latency, removes ad hoc buffering and heavy post-processing, and allows diagnostics and control loops to scale naturally on commodity server hardware.

DATA PLANE DEVELOPMENT KIT

The Data Plane Development Kit (DPDK) is an open-source project under the Linux Foundation, governed by a Governing Board and a Technical Board, with members as chipmakers market leaders such as Intel, AMD, Arm, NVIDIA, Marvell and Broadcom, telecommunications vendors such as Ericsson, Huawei and ZTE and cloud and software providers as Microsoft and Red Hat.

The DPDK community has over 200 contributors, with recent releases featuring around a thousand commits from approximately 150 authors each.

DPDK operates on principles of kernel bypass through poll-mode drivers (PMDs), core isolation to dedicate CPU cores for packet processing, user-space programming to avoid kernel overhead, and minimizing system calls for high performance. It is widely used in User Plane Functions (UPF) for 5G telephony and by hyper-scalers like AWS, Azure, and Google Cloud, where it is offered as an option for high-performance networking in virtual machines.

ROUND TRIP TIME

The well-established Round-Trip Time methodology (RTT) has been adopted to validate DPDK performance, testing it as a minimal model of a control system cell. In this configuration, two ports on the same host are connected through a loopback cable, with transmission and reception handled by separate processes pinned to dedicated CPU cores (Fig. 1). This arrangement abstracts the essential structure of a feedback loop, where the RTT provides a direct measure of end-to-end latency. In time-critical applications, one-way latency is conventionally taken as half of RTT, serving as a metric for system suitability in real-time control.

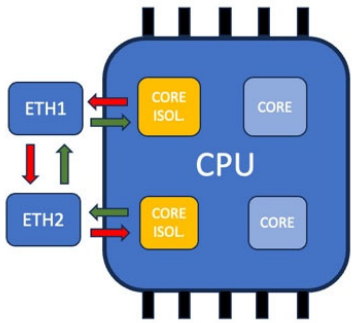

Figure 1: Round Trip Test setup on a single machine using two Ethernet ports.

ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUAI01

Tests have been carried out on a dedicated machine, a Dell PowerEdge 750 equipped with two Intel Xeon Gold 6346 processors and an Intel X710 quad-port NIC, running Voltumna Linux 2.6, a Yocto-based Linux distribution with Linux kernel 5.10 and DPDK 22.11, operated with a DPDK driver tuned to minimize latency.

To emulate realistic operating conditions, *stress-ng* was used to load the CPU, cache, and memory, while concurrent data transfers and a continuously running *top -d0* introduced pressure on the PCIe bus and the system call layer. Four tests were conducted: the first two compared a classical kernel-based configuration with a DPDK based one; the third highlighted the benefits of a dual-processor setup; and the fourth demonstrated that, ultimately, all measured latencies strongly depend on the specific hardware platform and must therefore be evaluated on a case-by-case basis.

DPDK vs. Raw Socket on Non-Isolated Core

DPDK showed significantly lower jitter than raw-socket implementation (Fig. 2), primarily because it bypassed system calls and operated in user space. Nonetheless, it remained affected by housekeeping threads and the sharing of hardware resources with the Linux kernel. Raw-socket performance can be regarded as a practical lower bound for control frameworks such as EPICS and TANGO. Time-sensitive applications, including feedback loops implemented as TANGO servers or EPICS IOCs, can therefore be expected to operate reliably up to a hundred hertz in controlled environments.

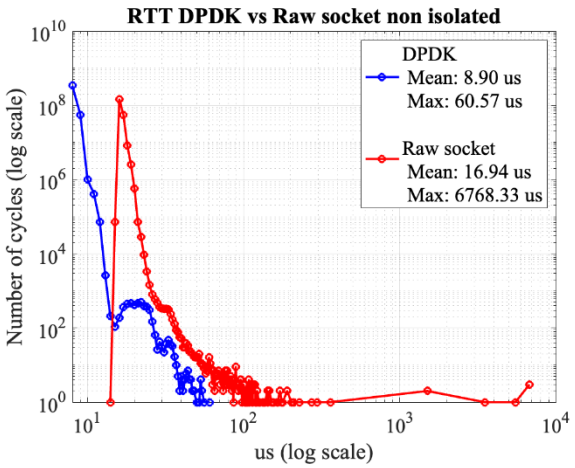

Figure 2: RTT in non-isolated environment.

DPDK vs. Raw Socket on Isolated Cores

When running on an isolated core, the raw-socket implementation achieved a mean latency of 24 µs, whereas DPDK reduced this to 8.04 µs (Fig. 3). For raw sockets, isolation can slightly increase latency, as scheduler optimizations are no longer available. With raw sockets, soft real-time feedback operation is feasible up to a few kHz once OS services are stripped down; with DPDK, hard real-time feedback rates approaching 100 kHz become realistic as the maximum latency is comparable to the mean latency.

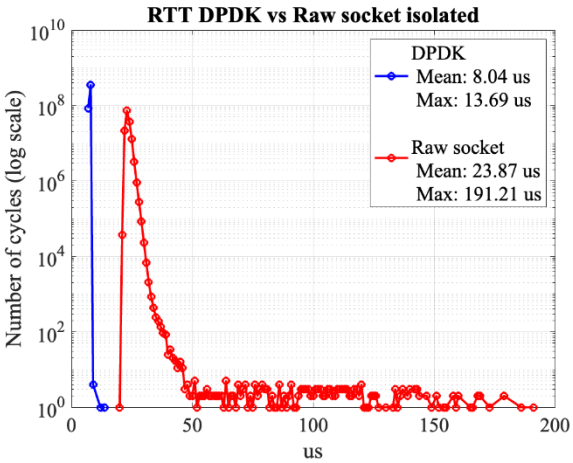

Figure 3: RTT in isolated environment.

The mean RTT corresponded to a one-way latency of 4 µs, consistent with values reported in literature [1].

DPDK on Full Isolated CPU vs. CPU Running Linux

When running RTT tests with DPDK on CPU 1, which shared the memory controller and PCIe subsystem with Linux, the average latency increased to 8.96 µs, about 1 µs higher than on CPU 2 (Fig. 4). This shows that Linux and DPDK applications running on the same subsystem introduces interference, whereas isolating DPDK on a separate CPU avoids this penalty.

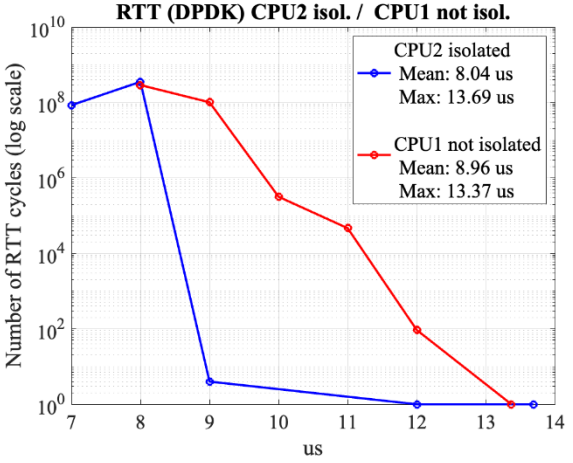

Figure 4: RTT on an isolated / non-isolated CPU.

DPDK on Different Hardware

DPDK tests were conducted across three different hardware setups (Fig. 5). Mean round-trip times ranged from a minimum of 8 µs, 13 µs and 22 µs, the latter two observed on a Supermicro rackmount equipped with a single Xeon D-1718T processor but with NICs (I350 and E823) connected to separate PCI roots. These results confirm that the performance strongly depends on the hardware platform.

ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUAI01

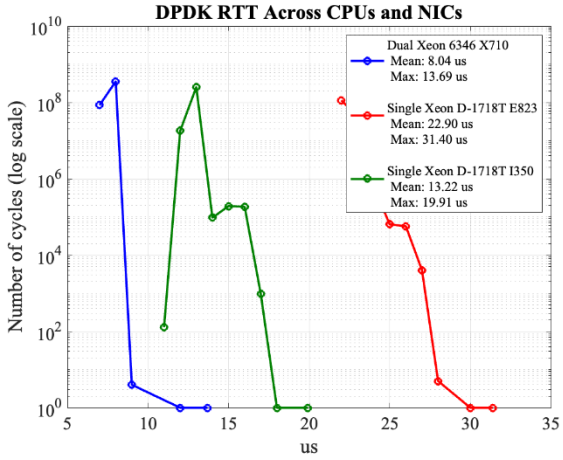

Figure 5: RTT on different architectures.

DPDK ON ELETTRA

The DPDK-based upgrade of the Global Orbit Feedback (GOF) system was specifically implemented on Elettra and validated over a period of one year, before its scheduled decommissioning in 2025. This long-term testing not only ensured the reliable operation of DPDK [2] but also allowed the integration of the new in-house BPM electronics that was not feasible with the legacy system. The upgrade proved invaluable in addressing the degradation of the power supply electronics, which, after 33 years of service, could no longer be repaired.

The old feedback system, based on a ring of 12 MVME6100 cards running at 10 kHz, was limited both in computational capacity to correct above 300 Hz and in its ability to integrate the new BPM electronics [3]. In contrast, the new implementation, with a dual-Xeon server collecting all BPM readings via Ethernet and driving the correctors through the legacy CPUs, provided virtually unlimited computing power relative to the 10 kHz feedback repetition rate.

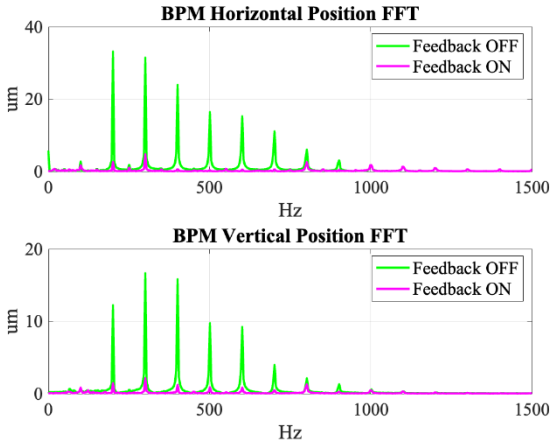

Figure 6: FFT of the orbit before and after switching ON the fast orbit feedback.

As shown in Fig. 6, this allowed the activation of thirteen notch filters at 50, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, and 900 Hz, effectively reducing peak orbit

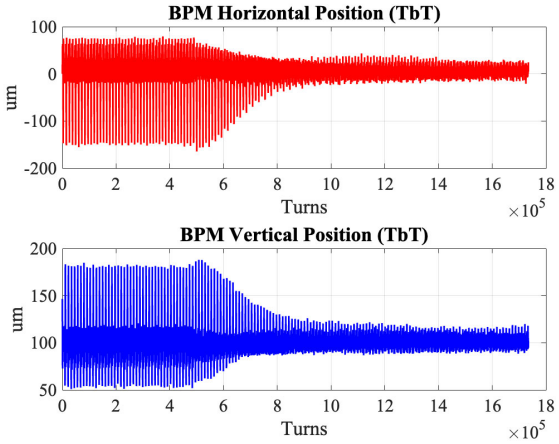

Figure 7: TbT data of a BPM when switching the feedback on.

distortions from over 150 µm (Fig. 7) during the last month of operation.

During the last year of Elettra operation, tune measurements for diagnostics and feedback were obtained from TbT data of the newly integrated BPMs. Using a moving FFT window of 8k samples, refresh rates up to 2 kHz were achieved. Even without any beam excitation (Fig. 8), this system has been proven faster than the bunch-by-bunch feedback and the old Libera Electron at injection.

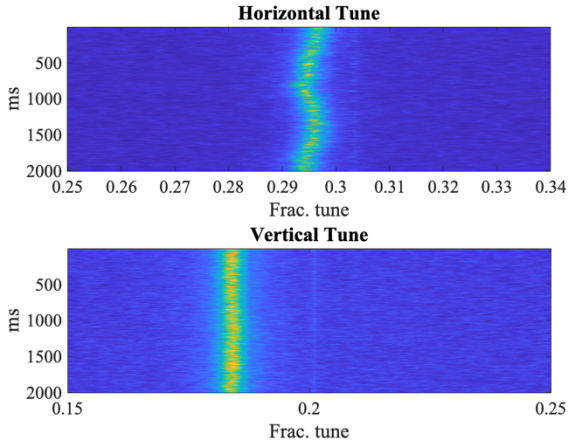

Figure 8: Spectrogram of the horizontal and vertical tune.

Confirming this, the high sampling rate allowed deterministic FFT computations on isolated cores to reveal harmonics at 50 Hz and 100 Hz (Fig. 9).

Beyond all the network traffic required for feedback control at 10 kHz, the server acquired a total of eight new TbT data streams: seven from new in-house BPM electronics and one from the Elettra 2.0 LLRF prototype. Data were first recorded and then decimated to 10 kHz and 10 Hz for high-level applications for a total data rate approaching 10 milions of packets per second (Mpps).

ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUAI01

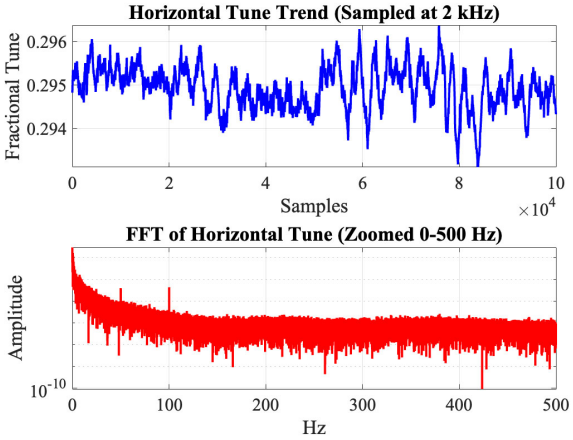

Figure 9: Horizontal tune and its FFT showing 50 Hz and 100 Hz peaks.

DPDK ON ELETTRA 2.0

For Elettra 2.0, the required TbT acquisition rate for all the diagnostics will exceed by an order of magnitude the 10 Mpps sustained in the DPDK system deployed in Elettra. The system must acquire data from 84 BPMs and 12 BLMs electronics (for a total of 168 BPMs and ~120 scintillators), 4 LLRF systems, the fast interlock electronics, and several photon diagnostics. Orbit feedback further increases the load, with 336 correctors embedded in sextupoles (300 Hz bandwidth), 48 correctors in the long straight sections (≈1 kHz bandwidth), and 144 air-core fast correctors (5 kHz bandwidth) controlled at the orbit feedback nominal frequency (100 kHz). The integration of all power supplies, including bending magnets, quadrupoles and sextupoles, adds only a small computational cost relative to diagnostics. Although the machine will employ more than 1200 power supplies, their organization into 96 daisy chains, rather than point-to-point connections as in the diagnostics, substantially reduces the communication overhead. Including all magnets in the real-time control system, even those not directly involved in orbit feedback, is worthwhile, as the additional computational cost is minimal compared to the benefits for high-speed optics procedures implementing LOCO and Beam-Based Alignment (BBA).

Under these conditions, the total traffic is estimated at 110–120 Mpps incoming and ~20 Mpps outgoing. Under these conditions, a conservative estimate indicates that the total memory load from network card transfer and circular buffer copying reaches nearly 80 GB/s.

Although the dual-Xeon system provides a nominal bandwidth of 204 GB/s per socket, time-sensitive applications cannot safely operate at such high utilization levels.

Test Facility

For Elettra 2.0, an AMD server with dual AMD EPYC Turin 9755 processors was selected as the RT central server. It provides a total of 128+128 physical cores, 1.5 TB of DDR5-6400 RAM, and a total bandwidth per socket of 576 GB/s. The server is connected to the LAN for booting (diskless system) and is equipped with three Intel E810 dual 100 GbE cards (Fig. 10), providing six ports connected to six of the eight available 100 GbE ports on an Extreme Networks 7520 switch.

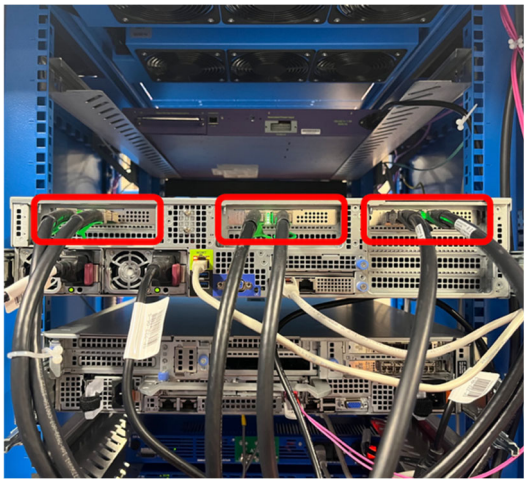

Figure 10: Rear view of the Supermicro server, showing the six 100 GbE ports highlighted in red with direct-attach cables.

To simulate the equivalent load of Elettra 2.0, 18 new BPM electronics are connected through 10 GbE ports to the same switch, transmitting data at TbT frequency. The switch, with a backplane capacity of 4 Tb/s, replicates all BPM traffic to the six ports connected to the server to reach the expected load. The server is configured in High Performance mode, with each socket set as a single NUMA node (NPS1), allowing transfer rates of up to 60 GB/s between cores on the same Core Chiplet Die (CCD), 36 GB/s for cores on different CCDs, and 22 GB/s for cores on different sockets. These numbers must be halved in case of heavy L3 cache misses. Six of the sixteen CCDs are pinned to match one-to-one with a single 100 GbE port, dedicating four cores out of eight (4 + 4 HT) per port.

The DPDK Abstraction Layer is configured to enable Receive Side Scaling (RSS), steering packets into eight queues. All packets are transferred from the card via DMA to ring buffers in memory, where cores fetch packets individually, extracting and tagging the payload with the server Time Stamp Counter (TSC) and the Turn Number (TN). The TN, a 64-bit counter available on each diagnostic, synchronized using the machine clock and a reference trigger, serves as the global timestamp, incremented at TbT frequency and included in each data stream. All meaningful readings are then filtered using the AVX-512 vector units to 100 kHz and 10 Hz and stored together with the full-rate data in circular buffers available for both fast feedback systems and high-level applications.

To avoid excessive pressure from a high number of idle cores spinning on the memory controller (which could cause memory contention and unnecessary thermal load), a feedback system within the application regulates sleep cycles for the network cards. On average, each core waits for a wakeup interval corresponding to every four-packet cycle. This feedback mechanism slightly increases the

ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUAI01

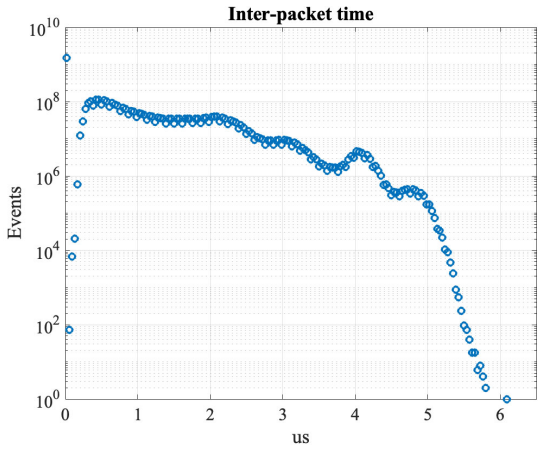

Figure 11: Distribution of the average BPM inter-packet time on the server (nominal inter-packet time is 864 ns).

overall latency (Fig. 11) but guarantees that no packets are lost, maintaining the integrity of the data stream.

End-to-End Latency

To evaluate the feedback loop's end-to-end delay, the server generates a UDP stream to the SFP input of the power supply, which drives an air-core coil with a square-waveform current. A current-to-voltage transducer on the power supply output modulates an RF signal fed to a BPM letting the server observe the PS output in baseband. Aligning data using the server's TSC clock, the temporal offset between the server setpoint (packet sent to PS) and BPM response (packet received from BPM) ranges from 40 µs to 50 µs. This offset corresponds to the time between the rising edge of black line and the rising edge of the blue-green line (Fig. 12).

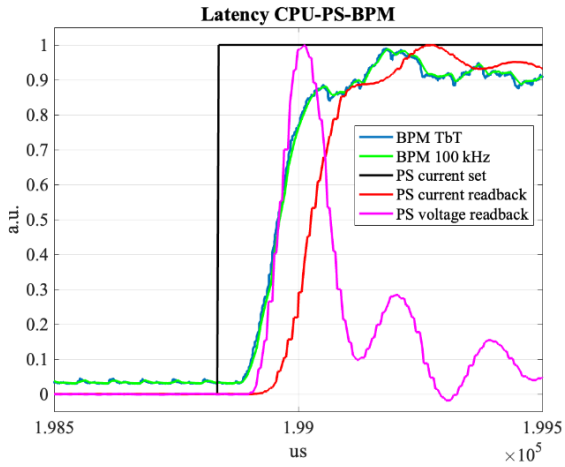

Figure 12: End-to-end system delay.

This 10 µs peak-to-peak jitter (not shown in the plot) arises from the 100 kHz power supply regulation loop, desynchronized from the setting period, plus additional latency from the server to prevent memory contention. A further 1 µs, needed for a 264×168 inverse matrix multiplication, completes the total end-to-end latency.

FUTURE PERSPECTIVES

With the platform nearly validated and its high core-to-core data exchange capabilities, application development requires an inter-core control system for direct read/write access to machine parameters, mimicking a standard control system.

On top of this lightweight layer, services such as configurable Multi Input-Multi Output SVD-based feedbacks for fast orbit control, but not limited to, can be executed, along with additional services for running high-frequency logical sequences on in-memory machine parameters.

The system will also support multiple instances of the Digital Twin developed in C from the MATLAB Accelerator Toolbox porting (C-AT), which implements the full *atlinopt6* and *ohmienvelope* functions and replicates the machine together with all the feedback systems that govern it. In addition to reproducing the behaviour of the real feedbacks, including theoretical response matrices and simulation of crosstalk or mismatches, direct in-memory access to machine data reduces latency when retrieving simulator outputs. Thanks to the lightweight design, multiple instances of the C-AT Digital Twin can run concurrently synchronized to the machine, using different corrective parameters or lattice configurations, without overloading the control system.

A potential weakness of this centralized system is the risk of failure. However, using DPDK at the user-space level ensures that a fault only terminates the affected process. A backup server, configured identically to the master with mirrored network traffic, remains synchronized and minimizes swap-over time.

CONCLUSION

Elettra 2.0 will bypass front-end computers, sending all time-sensitive data via Ethernet to a pair of redundant servers that handle fast machine control and more complex schemes than current vertical solutions. HPC hardware and a software-defined approach will allow easy future upgrades, making this system competitive in terms of cost and performance.

REFERENCES

[1] S. Gallenmüller, F. Wiedner, J. Naab, and G. Carle, "How Low Can You Go? A Limbo Dance for Low-Latency Network Functions", *J. Network Syst. Manage*, vol. 31, no. 1, Art. no. 20, 2023. doi:10.1007/s10922-022-09710-3

[2] L. Anastasio, A. Bogani, M. Cappelli, S. D. Gennaro, G. Gaio and M. Lonza, "Integration of DPDK for Real-Time Communication in the Elettra Synchrotron Orbit Feedback Control System: Jitter and Latency Optimization," in *IEEE Trans. Ind. Inf.*, vol. 21, no. 7, pp. 5104-5114, Jul 2025. doi: 10.1109/TII.2025.3547016

[3] G. Brajnik *et al.*, "First Experiences with the New Pilot-Tone-Based eBPM System in Elettra Storage Ring", *in Proc. IBIC'24*, Beijing, China, Sep. 2024, pp. 122–125. doi: 10.18429/JACoW-IBIC2024-TUP31