



Elettra Sincrotrone Trieste



From TANGO Devices to LLM Tools

Q&A at [slido.com #3879430](https://www.slido.com/join/3879430)

The Elettra 2.0 control challenge

- ✓ 400 mA stored current at 2.4 GeV, user operations from 2027
- ✓ Photon brilliance: **>100×** the present Elettra ring
- ✓ Feedback correction rate: **up to 1 MHz** (today ~10 kHz)
- ✓ Control-network bandwidth: **100 Gbit** (today 1 Gbit)
- ✓ Thousands of heterogeneous TANGO devices: power supplies, BPMs, vacuum, RF, insertion devices, interlocks
- ✓ GUI- and script-centric operator paradigm does not scale to this regime

- ✓ Anthropic MCP (Nov 2024): an open standard for connecting LLM agents to typed external tools
- ✓ Each tool exposes a **name + JSON parameter schema + natural-language description**
- ✓ The agent reads the catalogue at reasoning time, selects a tool, fills typed arguments, and invokes it
- ✓ Native client support across major LLM provider SDKs
- ✓ **Key property — typed admissibility:** the catalogue *is* the action space; the agent cannot fabricate operations outside it

The TANGO MCP server

✓ **Architecture**

- ✓ Single-file Python service, built on FastMCP
- ✓ Streamable HTTP transport, default `:8000/mcp`
- ✓ Discovers TANGO devices at startup from any `TANGO_HOST`
- ✓ Read-only resources via `tango://` URIs
- ✓ Typed tools:
 - `list_devices`, `get_device_info`, `read_attributes`
 - `write_attributes`, `run_commands`, `monitor_and_adjust`
- ✓ Prompts: `control_device`, `analyze_beam_stability`

✓ **Safety & robustness**

- ✓ Optional Bearer-token authentication (multi-key, constant-time check)
- ✓ Device-name validation; numeric guards (NaN/Inf rejected)
- ✓ Thread-safe LRU device-proxy cache (max 256)
- ✓ Bounded closed-loop control: iteration & wall-clock caps, writes clamped to TANGO `min_value/max_value`
- ✓ Same code base → Digital Twin or real machine via one env var

Demo: operator–AI session on the Digital Twin

- ✓ Three turns on *sr/power_supply/psqf_s01.01* — 2026-02-13
- ✓ **Op** — *Please read the status of the power supply.*
 - **AI** — State: OFF | Status: Off | 12:43:31.
- ✓ **Op** — *Now turn on the same power supply.*
 - **AI** — Successfully turned on. State: ON | Status: On | 12:44:08.
- ✓ **Op** — *Now change its Current to 1.*
 - **AI** — Written: 1 | Confirmed read: 0.0 (ramping up) | Quality: ATTR_VALID | 12:45:00.
- ✓ **Observations**
- ✓ Plain-language intent → typed MCP tool calls
 - *read_attributes, run_commands, write_attributes*
- ✓ Agent recognises readback ≠ setpoint → infers ramping
- ✓ ATTR_VALID quality flag interpreted as device-server acceptance
- ✓ Multi-turn context preserved (“the same power supply” resolves correctly)

Status, open source, and roadmap

- ✓ **Today:** hardened MCP server, exercised against the Elettra 2.0 Digital Twin
- ✓ One env-var switch (*TANGO_HOST*) from Digital Twin to real machine
- ✓ **Open source:** github.com/abogani/tango-mcp — MIT licence, single file, unit tests, no infra to stand up
- ✓ **Near-term roadmap**
 - Parallel MCP tool calls for bulk BPM / power-supply reads
 - Growing skill library invoking Sequencer / Behavior-Tree devices (orbit correction, injection, anomaly diagnosis)
 - Formal access-control model mapping operator roles → permitted skills
 - **Level-1 supervised autonomy** on the real machine, init-2027

Known limitations / open problems

- ✓ **No accelerator topology:** devices exposed flat (wildcard `list_devices`); no lattice, ring position, or upstream/downstream relations — the TANGO DB does not model them
- ✓ **No global attribute/command catalogue:** `list_devices` returns names only; the full action space is discoverable only device-by-device via `get_device_info` (one proxy round-trip each)
- ✓ **Serious analysis is token-expensive:** results return as JSON in the LLM context; multi-device × multi-attribute × time-series analysis inflates the context window fast



Elettra
Sincrotrone
Trieste

Thank you!