TB

TECHNICAL BRIEF

RADIANODE

TB-2026-001

Measuring Energy in Cellular IoT Devices

Why bench instrumentation and network emulation must be combined

---

Version 1.0
August 2026

Radianode Ltd Registered in England and Wales Company No. 17359155

PUBLICATION CLASSIFICATION

Classification

Document classification	Technical Brief
Document ID	TB-2026-001
Purpose	Measurement methodology
Contains	No measured results
Version	1.0
Publication date	August 2026
Status	Released
Issued by	Radianode Ltd

Series position

Radianode publications are organized into three series, each answering one question. A Technical Brief states how an engineer should think about a problem. It does not report measurement.

SERIES	ANSWERS	THIS PUBLICATION
TB	How should an engineer think about this problem?	TB-2026-001, this document
ERS	What did we actually measure?	ERS-2026-001 released; further entries in preparation
CS	What happened when this was applied?	Engineering Case Studies

Related publications

REFERENCE	RELATIONSHIP
ERS-2026-001	INT8 post-training quantization of a sequence-to-point energy disaggregation model. Supplies the quantized model used as the compute workload.
ERS-2026-002	In preparation. Applies this methodology to coincident-load supply behaviour.
ERS-2026-003	In preparation. Applies this methodology to the local-inference versus offload crossover.

Revision history

VERSION	DATE	CHANGE
1.0	August 2026	Initial release.

ENGINEERING SUMMARY

Summary

Engineering decision	Whether a battery-powered cellular IoT device should perform inference locally and transmit a result, or transmit raw data and infer elsewhere.
Problem	The decision is made at architecture stage from datasheet figures and inference benchmarks. Neither accounts for link condition, repetition, per-event overhead, or the five orders of magnitude separating sleep floors from transmit bursts.
Method	A conducted RF chain and a controlled RAN emulator make coverage and power-saving configuration settable variables. Dual-path current measurement spans sub-microamp to amp, because 8-bit quantization on the dynamic path cannot represent a microamp sleep floor. Serial GPIO phase encoding attributes energy to individual event phases.
Outcome	The crossover interval between local inference and offloading becomes measurable, as does its movement across commanded link conditions, its sensitivity to model precision, and coincident-load supply droop.
Limitations	No real propagation, no operator-specific network behaviour, no cross-vendor generalization from a single module, and no substitute for field trials.

SCOPE

Scope and assumptions

Scope

This brief defines the measurement capabilities required to resolve the local-inference versus offload decision for battery-powered cellular IoT devices operating on Cat-M1 and NB-IoT. It covers measurement architecture, attribution method, and the boundaries of valid conclusions.

Out of scope

- Measured results. No measurement data is reported here. Measured results are published through the Engineering Readiness Series.
- Bench procedures, wiring detail and calibration sequences.
- Model architecture, training and quantization method.
- Radio access technologies other than Cat-M1 and NB-IoT.

Assumption register

ID	ASSUMPTION
A1	The module supply rail can be electrically separated from the host board rail, allowing module-only current to be measured.
A2	Device firmware can be instrumented to signal phase boundaries on a single GPIO line without materially altering the energy being measured.
A3	Power-saving configurations commanded by the emulator fall within the range applied by deployed networks.
A4	Attenuation in the conducted chain is stable across a measurement session and verifiable at the device connector.
A5	The reporting interval and payload size under test are representative of the target application class.
A6	Every phase boundary of interest is distinguishable from a single encoded line. Where two phases must be separated within one acquisition and cannot be, they are reported as a combined interval.

THE ENGINEERING DECISION

The decision this brief addresses

ENGINEERING DECISION

Should computation be performed locally or offloaded for battery-powered cellular IoT devices?

This brief establishes the measurement methodology needed to answer that question. It does not answer it.

A connected device that senses something must decide what to do with the data. It can process the data locally and transmit a small result, or transmit the raw data and process it elsewhere. The first option spends energy on computation. The second spends energy on radio.

For a mains-powered gateway this rarely matters. For a battery-powered or energy-harvesting node with a ten-year service expectation, it determines whether the product is viable.

The decision is usually made at architecture stage, before hardware exists, using energy figures taken from module datasheets and inference benchmarks. This brief sets out why those figures are insufficient, and what a measurement environment has to provide in order to resolve the decision properly.

PROBLEM

Why current practice is insufficient

Module datasheets publish current consumption per operating mode. These figures are accurate, and they are not the numbers the decision needs.

Transmit current assumes an unstated link condition. A cellular module's transmit power is set by the network through closed-loop power control. The same module transmitting the same payload consumes substantially different energy in good and poor coverage, because the network commands it to transmit harder. A single transmit current figure describes one point on that range without stating which.

Repetition changes the accounting entirely. NB-IoT and Cat-M1 extend coverage by repeating transmissions. In deep coverage a single uplink message may be repeated many times over. The energy cost of a message is therefore not a property of the message. It is a property of the message and the link condition together.

Per-event energy is dominated by overhead, not payload. For small payloads, the energy spent establishing an RRC connection, waiting through the release timer and returning to sleep can exceed the energy spent transmitting the data. Two strategies with very different payload sizes may therefore show similar per-event energy, and figures indexed on data rate will not reveal this.

Sleep floors are the largest term and the hardest to observe. For a device reporting once an hour, sleep dominates the energy budget. The currents involved span five orders of magnitude, and a measurement taken at the wrong point on the board cannot see the low end at all.

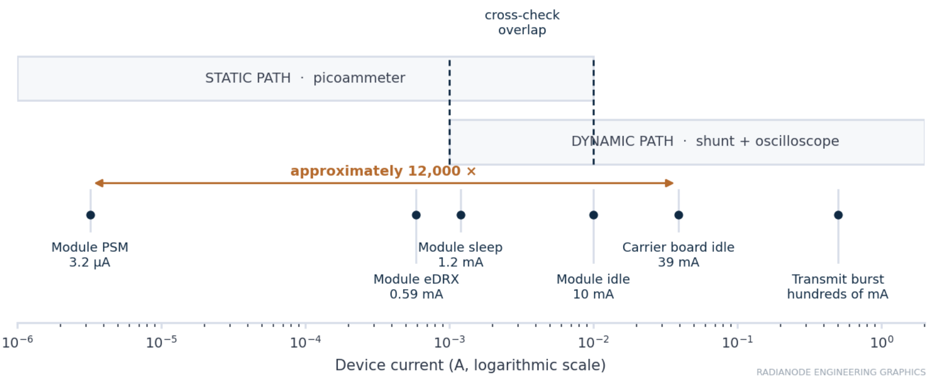

FIGURE 1

Dynamic Range of a Single Reporting Event

Device current spans roughly five orders of magnitude within a single reporting event. Values shown are vendor-published figures for a representative Cat-M1 and NB-IoT module and its carrier board. The marked span, from the module PSM floor to whole-board idle, is approximately 12,000 to 1, which is why the point on the board at which current is measured matters as much as the instrument. No single instrument covers the full range; the two measurement paths must overlap so they can be cross-checked against a common load.

This is not a theoretical concern. A three-month field study of thirty NB-IoT nodes deployed across more than twelve hundred locations found energy consumption imbalanced by a ratio of up to 75:1 between nodes running the same workload, attributed to differing coverage levels, long-tail power profiles and excessive control message repetition. A design decision taken from a single datasheet figure is being taken against a quantity that varies by nearly two orders of magnitude in service.

The consequence is visible in the analytical literature. Optimization studies of compute-versus-offload decisions commonly assume fixed transmit and receive power constants, on the order of a few hundred milliwatts each, applied uniformly across all link conditions. The resulting boundaries are internally consistent and untested against measurement.

Two measurement communities, neither sufficient alone

Cellular energy metrology measures module current against a base station emulator with a power analyser, producing battery lifetime estimates for NB-IoT and Cat-M1 devices. The network dimension is controlled and the instrumentation is sound. There is no compute workload. The device transmits; it does not think.

Edge inference energy metrology measures inference energy on microcontrollers and accelerators using shunt-based phase isolation and GPIO triggers, producing per-inference energy for quantized and unquantized models. The compute dimension is rigorous. There is no network. The device thinks; it does not transmit.

The trade-off between them has, to date, been addressed analytically rather than by measurement.

Why live-network measurement does not close the gap

Measuring a real device on a commercial network produces valid observations and unrepeatable ones. The measuring party controls neither the propagation environment, nor the cell load, nor the power-saving timers, which are set by the operator. A measurement taken on Tuesday afternoon cannot be reproduced on Wednesday morning, and cannot be reproduced at all by a third party elsewhere. Coverage is reported as an indicator such as RSRP, which is an outcome of the environment rather than a setting, so it cannot be commanded to a chosen value.

> ENGINEERING PRINCIPLE
>
> Coverage must be an independent variable, not an observed outcome.

METHOD

What a sufficient measurement environment requires

Four capabilities, together.

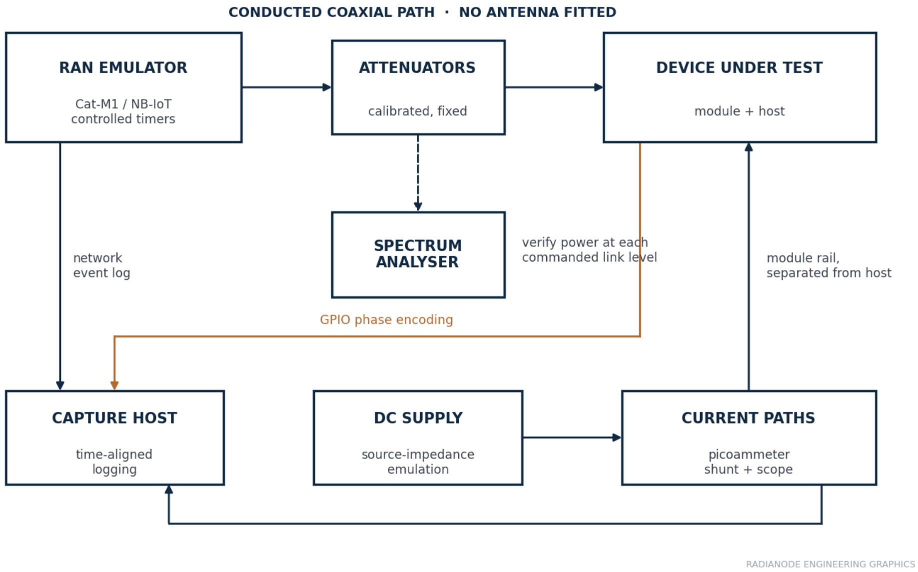

FIGURE 2
Measurement Environment

The RF path is conducted throughout, with no antenna fitted at any point. Attenuation sets the commanded link condition and is verified at the device connector. The module rail is powered and measured separately from the host rail. A single GPIO line encodes the active phase and is captured alongside the current waveform.

1. A conducted RF chain

The device under test is connected to the network emulator by coaxial cable through calibrated attenuators. No antenna is fitted at any point.

The compliance argument is straightforward: emulator configurations operate in licensed spectrum, and radiating would be unlawful and could deny service to the public.

The methodological argument matters more. In a conducted chain, link quality becomes a settable quantity. The engineer chooses the attenuation, verifies the resulting power at the device connector with a spectrum analyser, and records both. This is what makes the measurement repeatable by someone else.

2. A controlled radio access network

A licensed base station emulator supplying both Cat-M1 and NB-IoT allows the power-saving configuration to be set rather than inherited. PSM timers, eDRX cycle and paging time window, and connected-mode DRX all become controlled variables.

This matters because power-saving configuration is frequently where the architecture decision is actually settled. A strategy that wins under one PSM configuration can lose under another, and on a live network the engineer does not get to choose.

3. Dual-path current measurement

No single instrument spans a sub-microamp sleep floor and a transmit burst of hundreds of milliamps with adequate resolution at both ends. Two measurement paths are required, supported by a controlled supply, independent RF verification, and the network emulator that sets the link condition.

FUNCTION	INSTRUMENT	MEASURES	RANGE
Static current path	Keithley 2502 dual-channel picoammeter	PSM, eDRX and idle floors	sub-µA to low mA
Dynamic current path	Precision shunt with Keysight DSOX1102G oscilloscope	Wake, inference and transmit bursts	mA to A
Programmable supply	HP 6611C 40 W System Power Supply, 8 V, 5 A	Fixed rail voltage; source-impedance emulation and droop testing	0 to 8 V, 0 to 5 A
RF verification	Anritsu MS2712E Spectrum Master	Transmit power and occupied bandwidth per link level	9 kHz to 4 GHz
Cellular RAN	Amarisoft eNodeB/EPC software with PCIe SDR transceiver	Cat-M1 and NB-IoT with controlled power-saving configuration	Sub-6 GHz

The function column is normative. The instrument column records the implementation Radianode operates. Any instrument meeting the stated range and resolution is acceptable provided its uncertainty is declared.

The dual-path split is forced by quantization, not preference. The dynamic path digitizes at 8 bits, giving 256 levels across the full-scale setting. Scaled to capture a transmit burst of roughly 350 mA, one quantization step is about 1.4 mA. A module PSM floor of 3.2 µA is several hundred times smaller than a single step, so it is not merely noisy on the dynamic path, it is unrepresentable. The static path exists to measure what the dynamic path cannot encode at all.

Calibration status is declared in each publication that reports measurement.

Both current paths measure a common mid-range load before each session. Disagreement beyond the combined uncertainty budget halts the session. Uncertainty budgets for both paths are a required output, not an optional appendix.

> **ENGINEERING PRINCIPLE**
>
> The module supply rail must be separable from the host board. Where it is not, the split between compute and radio is an inference, not a measurement.

4. Phase-resolved attribution

Per-event energy is not useful as a single number. It has to be decomposed into wake and boot, pre-processing, inference, connection establishment, uplink transmission, downlink reception, connection release and tail, and return to sleep.

Decomposition is achieved by the device firmware signalling phase boundaries on a GPIO line, captured alongside the current waveform. Network-side event logs are aligned to the captures through a marker transaction at session start, with the alignment residual recorded.

The channel budget constrains how this is done. A two-channel instrument spends one channel on the shunt voltage, which leaves one channel for phase signalling. Parallel phase lines are therefore not available, and phase must be encoded serially on a single line, by pulse width or by edge sequence, with the external trigger input reserved for a coarse session marker. This bounds attribution granularity and is stated in any publication reporting results from the harness rather than left implicit.

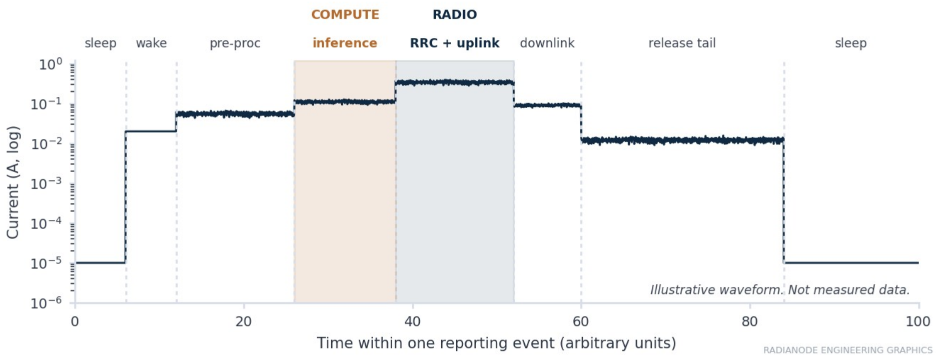

FIGURE 3

Phase Decomposition of a Reporting Event

Firmware signals phase boundaries on a single encoded GPIO line; network-side event logs are aligned to the capture through a marker transaction at session start, with the alignment residual recorded. Copper marks the compute phase and navy the radio phase. Waveform shown is illustrative and is not measured data.

Without this decomposition, a measurement can report that one strategy used more energy than another, but cannot say where the energy went, which is the part an engineer can act on.

OUTCOME

What becomes answerable

QUESTION	FROM DATASHEETS	FROM THIS ENVIRONMENT
What does the module consume in each mode?	Yes	Yes, verified independently
At what reporting interval does local processing beat offloading?	No	Yes, measured
How does that boundary move as coverage degrades?	No	Yes, across commanded link levels
Does model precision move the boundary by more than measurement uncertainty?	No	Yes
Does the power-saving configuration change which strategy wins?	No	Yes
Do transmit and compute bursts, occurring together, pull the supply below the module operating margin?	No	Yes, with emulated source impedance

The last question is worth isolating. A device that transmits while it computes draws both currents from the same supply at the same instant. On a constrained supply this produces a voltage droop that neither the module datasheet nor the inference benchmark predicts, because each characterizes its own load in isolation. The resulting failure is intermittent, load-dependent, and typically found in the field rather than on the bench.

LIMITATIONS

What this methodology does not answer

> LIMITATION
>
> The environment answers how a device behaves at a commanded link condition. It does not answer what link conditions the device will meet in service.

- **No real propagation.** Fading, mobility, interference and multipath are absent by design. This is the price of repeatability, and it is deliberate.
- **No operator-specific behaviour.** Deployed networks apply their own timer policies, scheduling and admission control. Emulated configurations approximate a range, not a specific operator.
- **No cross-vendor generalization.** Results from one module characterize that module. Claims across silicon require measurement across silicon.
- **No substitute for field trials.** The environment reduces the number of design questions a field trial must answer. It does not remove the trial.
- **Attribution depends on rail separation.** Where assumption A1 does not hold, compute and radio energy cannot be separated by measurement, and the conclusions weaken accordingly.
- **Phase granularity is bounded by the channel budget.** Serial encoding on one line cannot resolve phases that overlap in time. Overlapping phases are reported as a combined interval, not decomposed.

APPLYING THIS

Where this applies

The environment is most valuable before hardware is committed, when the architecture decision is still open and the cost of changing it is a design review rather than a product recall.

- **Selecting between local processing and offloading** for a device with a defined reporting interval and an expected coverage range.

- **Validating published energy figures** against your own duty cycle, particularly where a battery life claim underpins a commercial commitment.
- **Diagnosing intermittent field failures** in deployed devices where the failure correlates with transmission activity and no fault is reproducible on the bench.

This methodology provides a measurement framework for jointly characterizing local inference and cellular transmission energy under repeatable, controlled network conditions, enabling architecture decisions that cannot be made reliably from datasheet figures or isolated subsystem measurements alone.

REFERENCES

References

Standards

- 3GPP TS 36.331. Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC) protocol specification.
- 3GPP TS 36.213. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures. Uplink power control.
- 3GPP TS 24.301. Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS). Power saving mode, timers T3324 and T3412.
- 3GPP TS 36.101. Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment radio transmission and reception. Power classes and conducted test conditions.
- 3GPP TR 45.820. Cellular system support for ultra low complexity and low throughput Internet of Things. Source of the ten-year battery lifetime target at 5 Wh.
- JCGM 100:2008. Evaluation of measurement data. Guide to the expression of uncertainty in measurement.

Further reading

Cellular energy metrology against a base station emulator:

- A. Sørensen, H. Wang, M. J. Remy, N. Kjettrup, R. B. Sørensen, J. J. Nielsen, P. Popovski and G. C. Madueño. Modelling and experimental validation for battery lifetime estimation in NB-IoT and LTE-M. arXiv:2106.13286, 2021.

- P. Andres-Maldonado, M. Lauridsen, P. Ameigeiras and J. M. Lopez-Soler. Analytical modeling and experimental validation of NB-IoT device energy consumption. IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5691 to 5701, 2019.

Live-network measurement and its variance:

- Understanding power consumption of NB-IoT in the wild. Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom), 2020. doi:10.1145/3372224.3419212.
- F. Michelinakis, A. S. Al-Selwi, M. Capuzzo, A. Zanella, K. Mahmood and A. Elmokashfi. Dissecting energy consumption of NB-IoT devices empirically. IEEE Internet of Things Journal, vol. 8, no. 2, pp. 1224 to 1242, 2021.
- B. Martinez, F. Adelantado, A. Bartoli and X. Vilajosana. Exploring the performance boundaries of NB-IoT. IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5702 to 5712, 2019.

Edge inference energy metrology:

- C. Banbury et al. MLPerf Tiny benchmark. arXiv:2106.07597, 2021.
- Benchmarking energy and latency in TinyML: a novel method for resource-constrained AI. arXiv:2505.15622, 2025. Shunt-based measurement on the microcontroller core domain.

Related Radianode publications

REFERENCE	TITLE
ERS-2026-001	INT8 post-training quantization of a sequence-to-point energy disaggregation model.
ERS-2026-002	In preparation. Coincident-load supply behaviour.
ERS-2026-003	In preparation. Local-inference versus offload crossover.