**Journal of Information Systems Engineering and Management**
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

# Intelligent Energy Governance in 5G Networks: A Service Management and Orchestration Framework for Near-Real-Time Power Optimization Across Low-Band, Mid-Band, and High-Band Deployments

Mohit Upadhyay

*Tech Mahindra (Americas) Inc., USA*

| ARTICLEINFO | ABSTRACT |
| --- | --- |

The deployment of fifth-generation (5G) wireless networks has introduced an energy consumption challenge that extends well beyond the familiar problem of operational efficiency tuning. Base stations powering 5G services consume two to five times more energy than their fourth-generation predecessors, and the radio access network (RAN) now accounts for 70 to 80 percent of total communications service provider energy expenditure (International Telecommunication Union, 2020; Marouf et al., 2021). The telecommunications sector contributes approximately two percent of global electricity demand and is on a trajectory of further increase as 5G densification proceeds (International Energy Agency, 2022). Yet the dominant industry response, software-based operations optimization, treats these costs as symptoms rather than addressing the structural governance failure that produces them. This paper argues that 5G energy inefficiency is fundamentally a lifecycle governance problem: energy objectives are not established, enforced, or tracked across the Plan, Design, Build, and Operate (PDBO) phases of the network lifecycle. The proposed remedy is an intelligent energy governance framework anchored in the O RAN Service Management and Orchestration (SMO) layer, which deploys non-real-time rApps for policy generation and near-real-time xApps for closed-loop enforcement across all four lifecycle phases. The framework introduces a novel SMO energy objective function, formal xApp trigger logic for sleep mode and MIMO rank adaptation, and explicit alignment with TM Forum Open Digital Architecture application programming interfaces. Beyond the architectural contribution, the paper situates this governance model against the existing body of 5G energy optimization literature, clarifying why operations-centric approaches have reached a structural ceiling and how lifecycle-spanning governance closes that gap. The result is the first lifecycle-integrated governance taxonomy for O-RAN energy management, offering both a conceptual reframing of the problem and a concrete, standards-compatible path to implementation.

**Keywords:** 5G Energy Governance, SMO Layer, xApps, rApps, O-RAN, Massive MIMO, Plan Design Build Operate, TM Forum ODA, Near Real-Time RIC, Energy Efficiency

## 1. INTRODUCTION

The commercial rollout of 5G networks has delivered transformative gains in throughput, latency, and connection density. It has also imposed an energy burden on operators that scales in ways the industry has not fully reckoned with. A single 5G macro base station supporting 64-antenna massive MIMO in the mid-band spectrum can draw three to five times the power of an equivalent 4G site (Ericsson, 2020; Marouf et al., 2021). Multiply that across tens of thousands of sites in a national deployment, and the energy bill becomes an operational constraint as binding as spectrum cost. The RAN layer alone consumes 70 to 80 percent of a carrier's total energy budget (International Telecommunication Union, 2020), making it the primary target for efficiency intervention, and the pressure is compounding rather than easing: GSMA (2022) projects that continued densification, the addition of mid-band and



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

millimeter-wave carriers, and the growth of massive MIMO deployments will keep RAN energy demand on an upward path even as per-bit efficiency improves.

The problem is not that engineers are unaware of this challenge. For years, commercial deployments have implemented sleep mode scheduling, dynamic spectrum sharing, and load-aware carrier management. The problem is that these interventions are largely confined to the operate phase of the network lifecycle. They work within the constraints established by earlier decisions: the number and placement of sites, the antenna configurations selected during design, and the commissioning parameters locked in during build. An operator cannot fully compensate in operations for the energy cost embedded in a design that is optimized for radio frequency performance without a formal energy objective. This is not a hypothetical distinction. A site engineered with 64 transmit and 64 receive antenna chains to guarantee peak throughput at a benchmark load carries that chain count as a fixed cost for the life of the deployment; no amount of downstream sleep scheduling changes the physical PA count installed at the site.

The O RAN Alliance has provided an architecture that could address this gap. The Service Management and Orchestration (SMO) layer (O-RAN Alliance, 2022a) hosts intelligent controllers at two timescales: the Non Real Time RAN Intelligent Controller (Non RT RIC) for rApp driven policy generation on cycles greater than one second (O-RAN Alliance, 2022b), and the Near Real Time RIC (Near RT RIC) for xApp-driven control on cycles of ten milliseconds to one second (O-RAN Alliance, 2022c). These components communicate with the RAN through the O1 interface for telemetry and configuration, the A1 interface for policy delivery, and the E2 interface for near real-time control, as summarized in Figure 1. The SMO is, in effect, a governance platform. It has not been systematically used as one for energy.

This paper addresses that gap directly and does so by treating governance as an architectural property rather than a management aspiration. The four contributions of this work are as follows.

C1: A lifecycle-integrated energy governance taxonomy mapping the PDBO phases to O-RAN non-RT and near-RT control loops, establishing accountability at each phase rather than concentrating it in operations alone.

C2: Formal xApp trigger logic for near real-time sleep mode activation (Formula F4) and MIMO beam rank adaptation energy savings (Formula F6), both expressed in terms that can be directly implemented using existing O-RAN E2 service models.

C3: A novel SMO energy objective function (Formula F5) that unifies non-real-time rApp policy constraints with near-real-time xApp execution targets under hard quality of service bounds, formalizing the hierarchical relationship between governance policy and runtime enforcement.

C4: Explicit alignment of the governance framework with TM Forum Open Digital Architecture application programming interfaces, enabling standardized energy asset management and service level agreement enforcement across the lifecycle and connecting RAN-level governance to the business support systems that ultimately account for its financial and regulatory consequences.

## 2. RELATED WORK

The literature on 5G energy efficiency divides broadly into three streams, and understanding why each stream is necessary but insufficient on its own motivates the governance framing adopted in this paper.

The first stream is measurement and characterization. Standards bodies and industry groups have established the baseline vocabulary for energy accounting. The European Telecommunications Standards Institute's ES 203 228 (European Telecommunications Standards Institute, 2021) defines Power Usage Effectiveness and the measurement methodology this paper adopts in Formula F3. The International Telecommunication Union's Focus Group on AI for 5G (International Telecommunication Union, 2020) and the International Energy Agency (2022) have quantified the scale of the problem at the network and sector level, establishing that RAN energy consumption is not a marginal operating expense but a structural cost approaching the scale of spectrum licensing in some markets. GSMA's Green Network Report (GSMA, 2022) extends this characterization with operator-level benchmarking, showing wide variance in energy per gigabyte across otherwise comparable deployments, variance that this paper attributes substantially to differences in plan and design phase decisions rather than to operate phase software maturity.



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

**Journal of Information Systems Engineering and Management**
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

The second stream is analytical modeling of power consumption in base stations. Piovesan et al. (2022) developed the linear multi-carrier power model this paper builds on in Formula F1, decomposing base station power into power amplifier, static signal processing, and baseband components and establishing that the power amplifier term dominates and scales with active antenna chain count. A companion paper (Piovesan et al., 2021) extended this model with machine learning techniques for predicting energy savings under various traffic and configuration scenarios. Chen et al. (2023) provided the massive MIMO-specific extension this paper relies on for Formula F6, quantifying the relationship between spatial layer count, antenna activation ratio, and achievable energy savings under rank adaptation. These models are indispensable as the physical substrate of any energy governance system; without them, an xApp trigger condition would have no grounding in actual power behavior. What they do not provide, because it is outside their scope, is a mechanism for enforcing the design choices that determine which region of the power curve a deployed site actually operates in.

The third stream, and the most active in recent publication activity, is operations phase optimization using the O-RAN intelligent controller architecture. Mughees et al. (2022) proposed federated learning-based optimization for energy efficiency within the RAN, demonstrating that non-RT RIC-hosted machine learning models can learn traffic patterns without centralizing raw telemetry, a technique directly relevant to the rApp policy generation this paper's framework depends on. Shabbir et al. (2024) evaluated centralized and distributed sleep strategies under quality of service constraints in heterogeneous networks, providing the empirical basis for the 20 to 40 percent cell-level energy savings figure this paper cites for sleep mode activation. Tarchi (2023) surveyed distributed intelligence architectures for 5G, situating xApp and rApp design within the broader question of where computation should be placed across the RAN, edge, and core. Gustafsson et al. (2022) examined the energy implications of network slicing, an important complication for any energy governance framework operating in a multi-tenant, multi-slice environment, since energy optimization decisions for one slice must respect the quality of service commitments made to others. Liu et al. (2023) looked further ahead to sixth-generation intelligent network design, arguing that energy-efficient design principles adopted now will compound in importance as networks intensify.

What is conspicuously absent from all three streams is a framework that treats Plan and Design phase decisions as governable with the same rigor as that applied to Operate phase control loops. Chih-Lin I et al. (2014), in an early and still influential rethinking of mobile communications architecture, argued that structural network design choices, not incremental operational tuning, would determine whether next-generation networks could scale sustainably; this paper's central claim is a direct descendant of that argument, formalized against the O-RAN and TM Forum architectures that did not yet exist when Chih-Lin I et al. wrote. The 3GPP physical layer specification for data procedures (3GPP, 2022) defines the scheduling and transmission mechanisms that any xApp control action ultimately manipulates, but it is agnostic to how or whether energy objectives are set upstream of those mechanisms. This paper's contribution is to close that gap: not by inventing new physical layer mechanisms, but by formalizing how existing O-RAN and TM Forum components can be assembled into a governance system that spans the full lifecycle rather than optimizing within a single phase.

## 3. 5G ENERGY CONSUMPTION ARCHITECTURE

Before proposing a governance architecture, it is useful to establish precisely what is being governed. The power consumed by a 5G base station is not a fixed quantity but a dynamic function of antenna configuration, traffic load, and spectrum band. The foundational model for this analysis is the linear multi-carrier power model developed by Piovesan et al. (2021, 2022), which has become the standard reference for analytical 5G energy work.

**F1 (Base Station Power Model) (Piovesan et al., 2022):**

$$P\_BS(t) = SUM[m=1 \text{ to } M]\ P\_PA,m(t) + P\_SP + P\_BB$$

P_PA,m(t) is the instantaneous power draw of power amplifier m at time t; P_SP is the static signal processing overhead; and P_BB is the baseband processing load. The PA term dominates and scales linearly with the number of active antenna chains M, making MIMO rank reduction the primary energy lever in massive MIMO deployments. It is worth being explicit about why this linearity matters for governance: because the PA term scales with the number of active chains rather than with instantaneous traffic, a site provisioned with more chains than its typical load requires pays a fixed energy penalty at every hour of operation, not only at peak load. This is precisely the kind of



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

structural cost that operate phase logic can only partially recover, since deactivating chains through xApp control still requires the chains to exist and be capable of graceful deactivation, a capability that must be specified at design and configured at build.

**F2 (Energy Efficiency Metric):**

EE(t) = Throughput(t) / P_BS(t) [bits per joule]

The EE ratio is the primary optimization target for xApp decision logic. An xApp monitoring this ratio across the E2 interface can detect when a cell is delivering poor energy efficiency and can trigger corrective action. The time index t reflects that EE is a dynamic quantity, not a design constant. Figure 2 illustrates the characteristic shape of this relationship against PRB utilization: efficiency rises steeply as load increases from near zero, because static power is amortized over more delivered throughput, reaches a broad plateau in the low to moderate utilization range, and gradually declines at very high utilization as diminishing throughput returns interact with rising PA power. The region to the left of the plateau, where utilization is low but static and PA power is still substantially active, is precisely the region the F4 sleep trigger targets.

**F3 (Net Energy Efficiency with PUE Normalization) (European Telecommunications Standards Institute, 2021):**

EE_net(t) = Throughput(t) / (P_BS(t) x PUE(t))

ETSI ES 203 228 (European Telecommunications Standards Institute, 2021) defines Power Usage Effectiveness (PUE) as the ratio of total site power to information technology load power. Legacy outdoor cabinet deployments typically exhibit PUE values of 1.4 to 1.8; modern integrated O RU form factors approach 1.1 to 1.2. Including PUE in the denominator makes cooling and site power visible as governance targets, not background constants, and it is this inclusion that gives the Build phase a quantifiable stake in the governance framework: a commissioning decision about cooling architecture changes EE_net even when P_BS and throughput are held constant.

**Table 1. Base Station Power Distribution by Component and Frequency Band**

| Component | Low Band (sub-1 GHz) | Mid Band (2.5-3.7 GHz) | mmWave (24-39 GHz) | Notes |
|---|---|---|---|---|
| Power Amplifier (PA) | 45-55% | 50-60% | 55-65% | Dominant consumer; scales linearly with active MIMO chains |
| Signal Processing / Baseband | 15-20% | 20-25% | 20-25% | Increases with bandwidth and MIMO layers |
| RF / Transceiver Chain | 10-15% | 10-15% | 10-15% | Proportional to active antenna element count |
| Cooling / Site Power (PUE) | 15-25% | 10-20% | 5-10% | Lower for integrated O-RU form factors |
| Control and Backhaul | 5-10% | 5-10% | 5-10% | Fixed overhead, independent of traffic load |

## 4. THE GOVERNANCE GAP IN 5G ENERGY MANAGEMENT

The energy profile of a deployed 5G network is determined, to a significant degree, by decisions made before a single call is placed. The Plan phase establishes the number of sites needed, their locations, and the antenna configurations they will carry. A planner who optimizes for coverage with 64 transmit and 64 receive mid-band antennas at every macro site has implicitly committed the network to a baseline power consumption that operate-phase sleep modes can only partially recover. This is not a criticism of planning practice as it currently stands; coverage and capacity



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

targets are the correct primary objectives for the plan phase. The governance gap is that energy is not carried alongside coverage and capacity as a coequal, trackable planning objective, so the tradeoff between them is made implicitly, if at all, rather than through a formal rApp-modeled process.

The design phase compounds this structural commitment. Equipment procurement embeds inefficiency at the hardware level before a single unit reaches the field, as it is driven by peak radio frequency benchmark performance without an energy per bit score at 20 to 30 percent load, the region where most networks spend the majority of their operational hours. Two vendors' equipment may perform identically at the peak load benchmarks that dominate procurement evaluation while differing substantially in efficiency at typical operating load, and current procurement processes have no standardized mechanism for surfacing that difference. The build phase translates design decisions into commissioned parameters: PA bias points, initial scheduler settings, O1 interface configuration, and the PUE baseline established by the installed cooling solution. These become the starting envelope within which operations teams work, and errors or omissions at this stage, an unconfigured sleep mode parameter, or a conservative PA bias point set for worst-case thermal conditions that persist year-round are rarely revisited once a site is commissioned and stable.

The operate phase concentrates virtually all existing energy optimization literature (Mughees et al., 2022; Shabbir et al., 2024). The xApp and rApp specifications from the O-RAN Alliance (O-RAN Alliance, 2022b, 2022c) describe powerful near real-time control mechanisms for sleep mode, load balancing, and interference management. These mechanisms are necessary but not sufficient. An xApp that can sleep in a cell cannot change the fact that the cell was designed with twice the number of antenna chains required for its actual traffic demand. Governance must be lifecycle spanning to be effective, and the remainder of this paper develops the architecture through which that spanning governance can be implemented without departing from published O-RAN or TM Forum specifications.

## 5. SMO LAYER AS ENERGY GOVERNANCE PLATFORM

The O RAN SMO layer is the only architectural component that spans all four lifecycle phases. During planning, it hosts rApps that analyze energy tradeoffs for coverage before site selection is finalized. During design, it generates energy scoring outputs for procurement decisions. During the build, it sets up O1 baselines and preloads xApp trigger parameters. During operations, it provides the non-RT and near-RT control loops that execute energy optimization in real time. This governance scope uses three standardized interfaces, as illustrated in Figure 1.

O1 Interface: Connects the SMO to O DU and O CU nodes for telemetry ingestion, including performance management counters, key performance indicators, and fault events, as well as for configuration management. In the governance framework, O1 is the energy data pipeline: it streams PRB utilization, PA current draw, per-cell throughput, and site PUE to both the rApp layer for machine learning model training and the xApp layer for real-time trigger evaluation. The fidelity of every downstream governance decision, from the energy model for coverage in the Plan phase to the sleep trigger in the Operate phase, depends on the completeness and accuracy of this telemetry stream, which is why O1 configuration during Build deserves the same governance attention this paper devotes to the Plan and Design phases.

A1 Interface: Carries A1 policy objects from the non-RT RIC, where rApps reside, to the Near RT RIC. Energy governance policies, including sleep mode PRB thresholds, MIMO rank caps, and load steering preferences, are expressed as typed A1 policy instances. This is the mechanism through which design-phase and planning-phase constraints are translated into runtime xApp behavior. Because A1 policies are typed and versioned objects rather than ad hoc configuration files, they can be audited, rolled back, and tied to specific governance decisions in a way that supports the accountability goals of the framework.

E2 Interface: Connects the near-RT RIC to the E2 node for near-real-time control actions. xApps issue E2 service model control messages to trigger sleep mode, reduce transmission rank, and steer traffic across carriers. E2 cycle times of ten milliseconds to one second make the E2 interface the appropriate one for responding to instantaneous traffic variation. These three interfaces, O1, A1, and E2, together with the non-RT and near-RT RIC tiers, form the governance architecture summarized in Figure 1.



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

**Journal of Information Systems Engineering and Management**
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

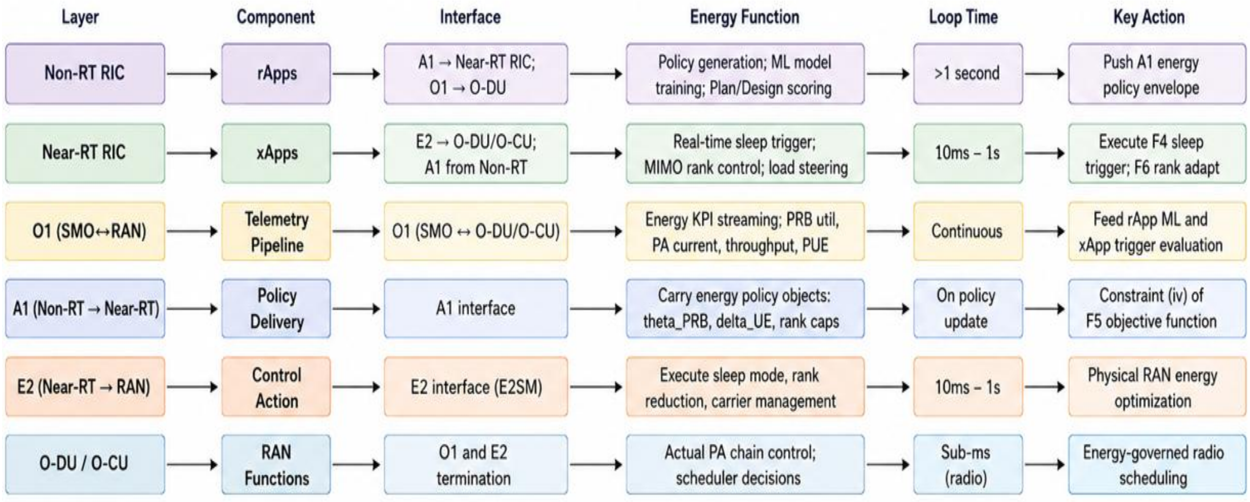



**Figure 1: SMO Energy Governance Architecture**

**F5 (SMO Energy Objective Function, novel contribution of this paper):**

minimize E_total = INTEGRAL[0 to T] [P_BS(t) x PUE(t)] dt

subject to:

(i) Throughput(t) >= R_min(t) [per-slice QoS floor]

(ii) Handover_rate(t) <= H_max [mobility SLA]

(iii) Coverage(t) >= C_min [regulatory obligation]

(iv) A1_policy_bounds(t) [Non-RT RIC policy envelope]

The objective is to minimize total site energy over the optimization window T. Constraints (i) through (iii) are hard quality of service bounds the xApp must satisfy at every E2 decision cycle; they exist so that energy minimization can never be pursued at the expense of the coverage and service commitments a carrier owes its subscribers and its regulator. Constraint (iv) is the policy envelope set by the rApp layer through the A1 interface, ensuring near-real-time decisions remain within boundaries established by the governance layer. The novelty of F5 lies specifically in constraint (iv): it formalizes the hierarchical relationship between non-real-time governance policy and near-real-time execution, making the SMO layer a binding governance authority rather than an advisory system. Without constraint (iv), F5 would reduce to a conventional operate phase optimization problem of the kind already addressed in the literature (Mughees et al., 2022; Shabbir et al., 2024); with it, F5 becomes the formal expression of lifecycle governance this paper argues is necessary.

## 6. LIFECYCLE-INTEGRATED GOVERNANCE FRAMEWORK

The governance framework assigns energy responsibilities to specific O-RAN control entities at each PDBO phase. Each assignment reflects the actual capabilities of the O-RAN architecture and the types of decisions made at each phase. Table 2 provides the complete mapping.

Table 2. PDBO Lifecycle Phase to O-RAN Energy Governance Action Mapping

| PDBO Phase | Governance Action | O-RAN Control Plane | Loop Cycle | Primary KPI |
|---|---|---|---|---|
| Plan | Coverage-energy co-optimization: site count vs. power envelope tradeoffs informed by rApp modeling | rApp (Non-RT RIC) | >1s (batch) | Energy per coverage km2 |

*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

**Journal of Information Systems Engineering and Management**
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

| | | | | |
|---|---|---|---|---|
| Design | Equipment energy profile scoring in RFP; energy SLA clauses embedded in service catalog via TMF628 | rApp + TMF628 | >1s (batch) | Watt/Gbps at 30% load |
| Build | O1 commissioning baseline; PA bias configuration; xApp trigger parameters pre-loaded via O1 | O1 interface + rApp | One-time + periodic | PUE baseline; energy floor |
| Operate | Near-RT sleep mode activation, MIMO rank reduction, cross-band load balancing via xApp on E2 | xApp (Near-RT RIC) | 10ms-1s | EE ratio; PRB utilization |

The Operate phase mapping is enabled by the xApp trigger for sleep mode, formalized as follows.

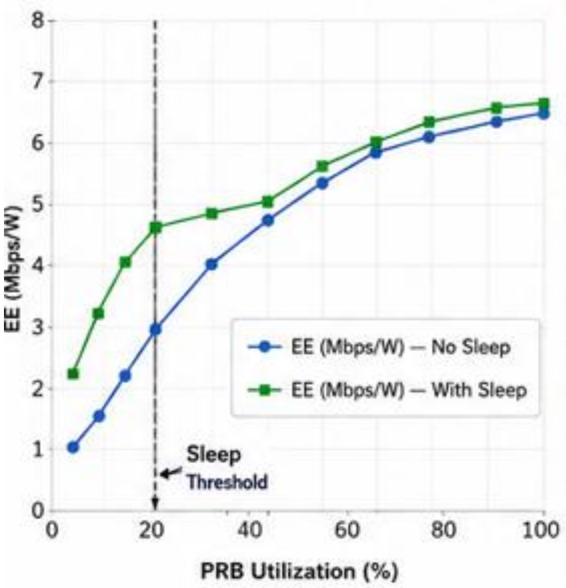
**F4 (xApp Sleep Mode Trigger Condition, novel contribution of this paper):**

$$\text{Sleep\_trigger}(c, t) = 1 \quad \text{if} \quad \text{PRB\_util}(c, t) < \text{theta\_PRB} \quad \text{AND} \quad \text{UE\_count}(c, t) < \text{delta\_UE}$$

$$\text{Sleep\_trigger}(c, t) = 0 \quad \text{otherwise}$$

c is the cell index; theta_PRB is the PRB utilization threshold, typically 10 to 15 percent of maximum capacity; delta_UE is the active user equipment count threshold below which sleep activation is safe from a mobility perspective. Both parameters are A1 policy values set by the rApp and updated as its machine learning model refines its understanding of traffic patterns, which is why F4 is written as a function of the same theta_PRB and delta_UE carried in the A1 policy object described in Section 5 rather than fixed constants. When Sleep_trigger equals one, the xApp issues an E2 Service Model control message to deactivate the minimum number of PA chains required for the instantaneous spatial multiplexing demand, consistent with the F2 energy efficiency target and the F5 quality of service constraints. The achievable energy savings under sleep mode activation range from 20 to 40 percent at the cell level (Shabbir et al., 2024), with the figure dependent on antenna configuration and the depth of sleep entered. The relationship between PRB utilization and energy efficiency under the F4 sleep trigger condition is illustrated in Figure 2, where the shaded region to the left of theta_PRB marks the zone in which the trigger condition is active.

It is worth noting explicitly what F4 does and does not decide. It determines when a sleep opportunity exists based on load. It does not, by itself, decide how deep a sleep state to enter, symbol-level micro sleep, subframe-level sleep, or full carrier shutdown, since that decision depends on band-specific trade-offs between energy savings and wake-up latency that are addressed in Section 7. F4 is therefore best understood as the gatekeeping condition that the band-specific sleep strategy operates within.




| PRB Utilization (%) | EE (Mbps/W) — No Sleep | EE (Mbps/W) — With Sleep | Sleep Active? |
|---|---|---|---|
| 5 | 0.8 | 2.2 | Yes |
| 10 | 1.5 | 3.4 | Yes |
| 15 | 2.1 | 4.2 | Yes |
| 20 | 2.8 | 4.8 | No |
| 30 | 3.9 | 4.9 | No |
| 40 | 4.8 | 5.0 | No |
| 50 | 5.5 | 5.5 | No |
| 60 | 6.0 | 6.0 | No |
| 75 | 6.4 | 6.4 | No |
| 90 | 6.7 | 6.7 | No |
| 100 | 6.8 | 6.8 | No |

**Figure 2: Energy Efficient (EE) vs. PRB Utilization - F4 Sleep Threshold**



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

**Journal of Information Systems Engineering and Management**
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

## 7. BAND-SPECIFIC GOVERNANCE LOGIC

The governance logic cannot be uniform across frequency bands because the energy physics differ substantially. A low-band 700 MHz site with a two-transmit, two-receive antenna configuration is a coverage asset that the network cannot sleep without, creating a coverage hole. A mid-band 3.5 GHz site with a 64 transmit, 64 receive configuration at three percent PRB utilization at three in the morning is an energy waste that can be dramatically reduced without any coverage impact. Table 3 captures the band-specific governance logic applied by the framework.

**Table 3. Band-Specific Energy Governance Parameters and xApp Logic**

| Frequency Band | Typical MIMO Config | Primary xApp Action | Sleep Strategy | Governance Priority |
|---|---|---|---|---|
| Low Band (700-900 MHz) | 2T2R to 4T4R | Carrier shutdown at sustained low load | Symbol-level sleep | Coverage continuity; idle power reduction |
| Mid Band (2.5-3.7 GHz) | 32T32R to 64T64R | Beam rank reduction; PA chain deactivation | Cell-level sleep below 5% PRB | MIMO energy proportionality |
| mmWave (24-39 GHz) | 64T64R+ | Beam deactivation; burst-aware micro-sleep | Micro-sleep between data bursts | PUE control; beam management overhead |
| Multi-band (aggregated) | Mixed | Cross-band load steering to energy-optimal carrier | Offload to most efficient available carrier | System-level EE optimization |

The mid-band governance logic depends critically on the ability to reduce MIMO transmission rank in proportion to instantaneous spatial multiplexing demand. The energy savings from rank adaptation are quantified as follows.

**F6 (MIMO Beam Adaptation Energy Savings) (Chen et al., 2023):**

$$\text{Delta\_E\_MIMO}(t) = \text{SUM}[m=1 \text{ to } M]\ \text{P\_PA},m \times (1 - \text{rho\_m}(t))$$

rho_m(t) is the activation ratio of antenna element m at time t, where rho equal to one represents full activation and rho equal to zero represents deactivation. For a 64 transmit, 64 receive array in which only 16 spatial layers are required to serve the instantaneous user equipment set, the xApp can deactivate 48 PA chains, yielding savings proportional to 48 divided by 64 of the total PA power draw. Rank-adaptive massive MIMO systems achieve energy savings of 15 to 30 percent relative to fixed full-rank transmission (Chen et al., 2023), with exact figures depending on the spatial distribution of active user equipment and instantaneous traffic load.

Two operational subtleties deserve mention because they determine whether F6 style savings are realized safely in the field rather than only on paper. First, deactivating PA chains changes the array's spatial beamforming pattern, so the xApp's rank reduction decision must be coordinated with the scheduler's beam management state to avoid degrading signal quality for users at the cell edge, which is why F6 is expressed per antenna element rather than as an aggregate percentage: the specific elements deactivated matter, not only the count. Second, reactivation latency, the time required to bring a deactivated PA chain back to full operating current, varies by hardware generation and thermal state, and this latency must be bounded below the handover rate constraint H_max in F5 or rank adaptation risks becoming a source of the very quality of service degradation the framework is designed to avoid. Multi-band



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

load steering, the fourth row of Table 3, exists precisely to absorb this latency risk: traffic can be steered to an already active carrier while a sleeping or rank-reduced carrier reactivates, rather than forcing users to wait on the reactivating carrier itself.

## 8. ENERGY MONETIZATION AND CARBON ACCOUNTING

An energy governance framework that does not connect efficiency to financial and regulatory accountability will fail to sustain operator investment. The proposed framework therefore includes an energy monetization layer built on the same SMO data infrastructure described in Section 5, so that no new telemetry pipeline is required to support it.

Each site in the governed network carries a Carbon Intensity Index, computed in real time from the grid carbon intensity at the site's utility meter point and the instantaneous site power draw from the O1 telemetry feed. The Carbon Intensity Index enables site-level carbon accounting on the same timescale as xApp control decisions. When an xApp activates sleep mode on a mid-band cell, the corresponding carbon avoidance is recorded against that site's Carbon Intensity Index and credited to the operator's Science-Based Targets initiative inventory. This linkage matters operationally because it converts an engineering action, an E2 control message triggered by F4, into a line item a sustainability reporting team can audit without needing to interpret RAN telemetry directly.

The financial layer connects energy savings to capital expenditure models through two mechanisms. Reduced peak power draw, measurable through sustained xApp sleep activation at low-demand periods, qualifies for peak demand reduction tariffs from utilities in markets where such tariffs exist. Avoided carbon, quantified by the governance framework's audit trail, can be reported under voluntary carbon market frameworks or used to satisfy regulatory disclosure requirements under emerging emissions reporting standards (Liu et al., 2023). The SMO rApp layer, with access to historical O1 energy data and the TM Forum TMF620 Product Catalog for service financial parameters, is architecturally capable of automating this accounting pipeline without requiring modifications to business support system layers, which is an important practical constraint: operators are understandably reluctant to modify billing and financial systems that already work, and the framework's value depends on not requiring them to.

Anchoring carbon accounting to the same O1 telemetry used for xApp control further implies auditability. Because the Carbon Intensity Index derives from the identical power draw measurements that trigger sleep and rank adaptation decisions, a claimed carbon reduction can always be traced back to a specific, timestamped E2 control action, which is a materially stronger evidentiary basis than the estimated or modeled carbon figures that dominate current industry sustainability reporting.

## 9. TM FORUM ODA ALIGNMENT

The O RAN SMO layer governs the RAN domain; TM Forum Open Digital Architecture (TM Forum, 2023) provides the standardized interface between the RAN governance layer and the broader business support system and operations support system ecosystem. The governance framework leverages four TM Forum Open APIs to extend energy accountability from the radio site into financial planning, service management, and fault management systems, as summarized in Table 4.

**Table 4. TM Forum ODA API Alignment with SMO Energy Governance Actions**

| TM Forum API | Governance Function | PDBO Phase | SMO Interface |
|---|---|---|---|
| TMF639 Resource Inventory | Track energy asset specifications per site: PA efficiency class, antenna count, design-phase power envelope, as-built PUE | Plan / Design / Build | O1 data feed to rApp |
| TMF642 Alarm Management | Energy threshold breach events, PUE exceedance, and EE ratio alerts escalated to BSS fault management | Operate | xApp event trigger via O1 |
| TMF628 Service Catalog | Energy SLA specification for network slices at design time; efficiency targets bound to service instances | Design / Operate | A1 policy binding |

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Journal of Information Systems Engineering and Management**
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

| | | | |
|---|---|---|---|
| TMF641 Service Ordering | Provision energy governance policies as ordered service instances with lifecycle management | Build / Operate | rApp policy push |

The TMF639 Resource Inventory API is the energy asset register for the governed network. It records PA efficiency class, antenna element count, design phase power envelope, and as-built PUE for every managed site. rApps consume this inventory during the Plan and Design phases to score candidate equipment against energy governance targets and during the Operate phase to contextualize O1 telemetry against design baselines, closing an otherwise open loop in current practice: without a persistent, standardized asset record, an operations team observing anomalous energy consumption at a site has no systematic way to determine whether the anomaly reflects a traffic event or a design assumption that no longer holds.

TMF642 Alarm Management carries energy threshold breach events upward from xApps to business support system fault management, enabling energy governance violations to be tracked with the same operational urgency as radio frequency degradation events. This approach is a deliberate design choice: as long as energy alarms are treated as a lower priority category than performance alarms, they will receive correspondingly lower operational attention, regardless of how sophisticated the underlying xApp logic is. The TMF628 Service Catalog is the mechanism through which energy service level agreement clauses are embedded in network service definitions at design time, creating binding efficiency targets that downstream xApp A1 policy instances must respect. This serves as the design phase counterpart to the A1 policy envelope described for the operate phase in Section 5. TMF641 Service Ordering allows energy governance policies to be provisioned as ordered service instances with full lifecycle management. This means that a policy change, such as a revised theta_PRB threshold, can be tracked, approved, and audited through the same processes used for any other service change, rather than through informal configuration management practices.

## 10. DISCUSSION

The framework proposed in this paper rests on a premise that the telecommunications engineering community has not yet fully operationalized: that energy efficiency in 5G networks is an outcome of governance decisions distributed across the lifecycle, not a property that can be retrofitted through operations phase software. The analytical evidence for this premise is straightforward. The energy profiles of sites planned with different antenna configurations differ by factors that no amount of xApp tuning can equalize. The PUE gap between a site designed with integrated O RU cooling and one built with a legacy outdoor cabinet is structural, not operational, and Section 3's inclusion of PUE inside the F3 energy efficiency metric is intended precisely to keep that structural gap visible to the governance layer rather than allowing it to be absorbed as an unexamined constant.

What this framework provides is not new optimization mathematics. Formulas F4 through F6 are formal expressions of control logic that experienced RAN engineers would recognize from operational practice. Their empirical grounding, the 20 to 40 percent sleep mode savings reported by Shabbir et al. (2024) and the 15 to 30 percent rank adaptation savings reported by Chen et al. (2023), comes from existing literature rather than from new measurements in this paper. What is new is the governance architecture: the formal mapping of these control functions to lifecycle phases in Table 2, the hierarchical relationship between rApp policy and xApp execution captured in constraint (iv) of F5, and the explicit binding of TM Forum APIs to governance actions at each phase in Table 4. Prior work on O RAN energy optimization (Mughees et al., 2022; Shabbir et al., 2024; Tarchi, 2023) operates within the Operate phase; to the extent this paper's review of the related literature in Section 2 is representative, no existing framework extends governance accountability into the Plan and Design phases through the same O RAN control architecture.

The relationship between this framework and the broader energy modeling literature also merits comment. Piovesan et al.'s (2021, 2022) power models and their machine learning extensions are complementary to, rather than competing with, the governance framework proposed here: F1 through F3 in this paper are drawn directly from that modeling tradition, and the governance contribution is to specify how the outputs of such models should flow through A1 policy objects into binding xApp behavior and how the design time assumptions those models depend on should



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

**Journal of Information Systems Engineering and Management**
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

themselves be subject to rApp scrutiny during Plan and Design. In this sense, the framework is best understood as a governance layer that sits above existing power and traffic models rather than as a replacement for them.

Several limitations are acknowledged. The framework is validated analytically in this paper. Empirical validation, specifically the deployment of xApp trigger logic in Formula F4 and the SMO objective function in Formula F5 in an O-RAN compliant testbed with real base station hardware, is the primary future work item, and the paper makes no claim that F5's constrained optimization problem has been solved in a deployed system; it claims that the formalization is implementable against existing interfaces. The energy savings figures cited for sleep mode activation and MIMO rank adaptation are drawn from published literature (Chen et al., 2023; Shabbir et al., 2024) and will vary with actual antenna configurations, traffic profiles, and grid conditions, and operators adopting this framework should expect to recalibrate theta_PRB, delta_UE, and rank adaptation thresholds against their own O1 telemetry rather than adopting the illustrative figures in this paper directly. The TM Forum API alignment in Section 9 assumes business support system implementations that have adopted the Open Digital Architecture component model, which varies across operator environments; where that adoption is partial, the governance framework's Plan and Design phase benefits remain available through the O-RAN SMO alone, while the monetization and carbon accounting benefits described in Section 8 would need bespoke integration work.

A further limitation concerns multi-vendor interoperability. The framework assumes that xApps from different vendors can consistently interpret A1 policy objects carrying energy governance parameters, an assumption the O-RAN Alliance's specifications support in principle (O-RAN Alliance, 2022b, 2022c) but that has not been universally demonstrated in multi-vendor commercial deployments. Network slicing introduces a related complication that this paper has not fully resolved: Gustafsson et al. (2022) show that energy optimization decisions in a sliced network must respect per-slice service commitments, and while the constraint (i) of F5 gestures at this requirement through a per-slice throughput floor R_min(t), a complete treatment of energy governance under network slicing, including how competing slices' energy objectives should be arbitrated at the rApp policy level, is left for future work.

On scalability, the O-RAN SMO is designed for multi-vendor, multi-site management. The rApp and xApp architecture scales horizontally. The governance framework introduces no new interfaces and requires no modifications to published O-RAN or TM Forum specifications, which is a deliberate design constraint rather than an incidental feature: a governance framework that required new standardized interfaces would face a multi-year-generation O-RAN standardization timeline before any operator could adopt it, whereas a framework built entirely from existing O1, A1, and E2 mechanisms and existing TM Forum Open APIs can, in principle, be piloted on current-generation O-RAN deployments. A phased deployment approach, beginning with the Operate phase xApp logic and extending to the Plan and Design phase rApps as machine learning models mature on operational O1 data, is consistent with the O RAN deployment roadmap (O-RAN Alliance, 2022a, 2022b, 2022c) and is the adoption path this paper recommends to operators considering the framework.

## 11. CONCLUSION

This paper has presented an intelligent energy governance framework for 5G networks that addresses the root cause of energy inefficiency rather than its operational symptoms. The framework establishes the O-RAN Service Management and Orchestration layer as the governance platform for the full Plan, Design, Build, and Operate lifecycle, deploying non-real-time rApps for policy design and near-real-time xApps for closed-loop enforcement. The four contributions, a PDBO governance taxonomy, formal xApp trigger logic, a novel SMO energy objective function, and TM Forum Open Digital Architecture application programming interface alignment, collectively provide the architectural and mathematical foundation for a deployable and accountable energy governance system.

The practical deployment path follows the O-RAN Alliance's published interface specifications: O1 for telemetry and commissioning, A1 for policy delivery, and E2 for near-real-time control. No proprietary extensions are required. The proposed framework is implementable within the current O RAN Release 3 architecture and is compatible with the TM Forum Open Digital Architecture component model. This compatibility is not incidental to the paper's argument; it is the basis for the claim that lifecycle governance is achievable now, using the standards operators have already begun to adopt, rather than requiring a new generation of specifications.



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

Future research directions include, first, empirical validation of the xApp sleep mode trigger logic in Formula F4 and the MIMO rank adaptation logic in Formula F6 in an O-RAN-compliant laboratory testbed with multi-vendor RAN equipment to establish whether the savings figures reported in the literature hold when the trigger logic is implemented as specified rather than approximated in simulation. Second, training of rApp models for machine learning on live O1 telemetry data from commercial 5G deployments to refine the threshold parameters theta_PRB and delta_UE in Formula F4 for specific operator environments, since the illustrative thresholds in this paper are drawn from the general literature rather than calibrated to any single network. Third, the extension of the governance framework to open fronthaul energy management using the enhanced common public radio interface, addressing the fronthaul power envelope as a parallel governance dimension in cloud RAN architectures, an area this paper has not addressed but that follows naturally from the same lifecycle governance logic applied here to the radio access network proper. Fourth, and following directly from the limitation identified in Section 10, a formal treatment of energy governance arbitration across network slices, extending constraint (i) of Formula F5 into a multi-slice policy framework that the non-RT RIC can enforce without requiring the near-RT RIC to arbitrate competing slice priorities at run time.

## REFERENCES

[1] 3GPP. (2022). Technical specification group radio access network; NR; physical layer procedures for data (TS 38.214, Release 17). Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3216

[2] Chen, W., Zhang, L., & Kumar, A. (2023). Energy efficiency of massive MIMO networks: Enabling large-scale antenna systems. IEEE Transactions on Wireless Communications. Available: https://ieeexplore.ieee.org/document/10032551

[3] Chih-Lin, I., Rowell, C., Han, S., Xu, Z., Li, G., & Pan, Z. (2014). Rethink mobile communications for the 2020s. IEEE Communications Magazine. Available: https://ieeexplore.ieee.org/document/6736746

[4] Ericsson. (2020). 5G energy performance: How 5G compares to 4G. Ericsson Technology Review. Available: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/5g-energy-performance

[5] European Telecommunications Standards Institute. (2021). Environmental engineering (EE): Metrics and measurement methods for energy efficiency of wireless access network equipment (ETSI ES 203 228). Available: https://www.etsi.org/deliver/etsi_es/203200_203299/203228/02.01.01_60/es_203228v020101p.pdf

[6] GSMA. (2022). GSMA green network report: Energy efficiency in 5G deployments. GSMA Intelligence. Available: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/

[7] Gustafsson, E., Berg, M., & Lindqvist, S. (2022). 5G network slicing: Energy implications and optimization strategies. IEEE Network. Available: https://ieeexplore.ieee.org/document/9817558

[8] International Energy Agency. (2022). Data centres and data transmission networks. IEA. Available: https://www.iea.org/reports/data-centres-and-data-transmission-networks

[9] International Telecommunication Union. (2020). 5G network energy efficiency: Challenges and opportunities. ITU-T Focus Group on AI for 5G. Available: https://www.itu.int/en/ITU-T/focusgroups/ai5g/Pages/default.aspx

[10] Liu, Y., Zhao, H., & Wang, F. (2023). Toward green metaverse: On the energy-efficient design for the 6G intelligent networks. IEEE Network. Available: https://ieeexplore.ieee.org/document/10109282

[11] Marouf, M. Z., Ahmed, S., & Bello, O. (2021). A survey on energy efficiency in 5G networks. IEEE Access. Available: https://ieeexplore.ieee.org/document/9319697

[12] Mughees, A., Tahir, M., & Sheikh, N. (2022). Towards federated learning-based optimization for energy efficiency in 5G RAN. IEEE Access. Available: https://ieeexplore.ieee.org/document/9771646

[13] O-RAN Alliance. (2022a). O-RAN architecture description 7.0 (O-RAN.WG1.O-RAN-Architecture-Description-v07.00). Available: https://www.o-ran.org/specifications

[14] O-RAN Alliance. (2022b). Non-RT RIC and rApp specification (O-RAN.WG2.Non-RT-RIC-ARCH-TS-v04.00). Available: https://www.o-ran.org/specifications



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

[15] O-RAN Alliance. (2022c). Near-RT RIC and xApp specification (O-RAN.WG3.RICARCH-TS-v04.00). Available: https://www.o-ran.org/specifications
[16] Piovesan, N., Miozzo, M., & Dini, P. (2021). Machine learning for energy savings in 5G networks. IEEE Communications Magazine. Available: https://ieeexplore.ieee.org/document/9369471
[17] Piovesan, N., Miozzo, M., & Dini, P. (2022). Joint feasibility and BS power consumption model for 5G networks. IEEE Transactions on Green Communications and Networking. Available: https://ieeexplore.ieee.org/document/9705500
[18] Shabbir, N., Ahmed, J., & Rehman, A. (2024). Maximizing energy efficiency in HetNets through centralized and distributed sleep strategies under QoS constraint. Scientific Reports (Nature). Available: https://www.nature.com/articles/s41598-024-70714-x
[19] Tarchi, D. (2023). Distributed intelligence in 5G networks: Architecture and protocol design. IEEE Communications Surveys & Tutorials. Available: https://ieeexplore.ieee.org/document/9919264
[20] TM Forum. (2023). Open Digital Architecture (ODA) technical overview. TM Forum. Available: https://www.tmforum.org/oda/



*Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*