International Journal of Multidisciplinary Research and Growth Evaluation.

Rethinking Energy, Reliability, and Latency Trade-offs in Green Communication for Next Generation IoT

William Asiedu [1*], Robert Quainoo [2]
[1] Iowa State University, USA
[2] Northeastern University, USA

* Corresponding Author: **William Asiedu**

Article Info

ISSN (online): 2582-7138
Volume: 05
Issue: 06
November-December 2024
Received: 13-10-2024
Accepted: 15-11-2024
Published: 17-12-2024
Page No: 1987-1994

Abstract
Energy efficiency, communication reliability, and end to end latency have become the defining performance axes of the next generation Internet of Things (IoT), yet they pull system design in conflicting directions. Mechanisms that buy reliability, such as retransmissions, redundancy, and conservative coding, tend to spend energy and time, while aggressive energy saving through duty cycling, transmit power reduction, or harvesting can erode both timeliness and dependability. This article revisits the conventional view that these objectives form a fixed hierarchy in which one is simply traded for another. Drawing on recent literature in green IoT, ultra reliable low latency communication, wireless energy transfer, edge computation offloading, and the emerging sustainability agenda of sixth generation (6G) networks, we argue that the three-way relationship is better understood as a configurable operating region whose shape is set by workload semantics, channel conditions, and the available physical and architectural levers. We review the metrics by which efficiency is measured, organize the dominant enabling mechanisms by their trade-off signatures, map the requirements of major application domains onto the trade-off space, and propose a unified conceptual framework that treats energy, reliability, and latency as jointly negotiated rather than ranked. We close by identifying open problems in cross layer modeling, measurement, and the standardization of sustainability as a testable constraint. The intended contribution is a clearer vocabulary for reasoning about when a given green technique helps and when it merely shifts cost from one axis to another.

DOI: https://doi.org/10.54660/.IJMRGE.2024.5.6.1987-1994

Keywords: green communication, Internet of Things, energy efficiency, ultra reliable low latency communication, energy harvesting, edge computing, 6G, sustainability

1. Introduction

The proliferation of connected devices has turned communication energy into a first order concern for the Internet of Things. As deployments scale toward tens of billions of endpoints, the aggregate power drawn by radios, the carbon footprint of the supporting infrastructure, and the operational cost of maintaining battery powered nodes have moved from background considerations to central design constraints. Research on the IoT has long recognized that scale changes the character of the problem, because design choices that are negligible for a single device become decisive when multiplied across a population (Stankovic, 2014). Surveys of green IoT consistently identify the radio interface as one of the largest and least forgiving sources of energy expenditure, because communication must continue even when computation can be deferred or compressed (Alsharif, Jahid, Kelechi, & Kannadasan, 2023; Alazzawi *et al*., 2025). At the same time, the applications that justify next generation IoT are increasingly intolerant of delay and failure. Industrial control, vehicular coordination, remote operation, and tactile interaction demand that packets arrive both quickly and almost certainly, a regime captured by the ultra-reliable low latency communication (URLLC) service class introduced in fifth generation systems and extended in the 6G vision (Bennis, Debbah,

& Poor, 2018; Durisi, Koch, & Popovski, 2016). Reliability targets on the order of one failure in a hundred thousand transmissions, paired with latencies near one millisecond, leave little slack for the very techniques that conserve energy. The result is a structural conflict between the sustainability agenda and the dependability agenda that the next generation of IoT is expected to satisfy at once.

This tension is usually framed as a set of pairwise trade-offs: energy against reliability, energy against latency, and reliability against latency. Such framing is useful but incomplete. It encourages designers to optimize one pair while holding the third axis implicitly fixed, which can hide the fact that a gain on one front frequently relocates cost onto the axis that was not in view. A retransmission policy tuned for reliability under a latency budget, for example, changes the energy profile in ways that depend on feedback delay and blocklength in a manner that is not monotonic (Avranas, Kountouris, & Ciblat, 2018). Earlier surveys of the energy efficiency trade-off in wireless green communication already cautioned that no single efficiency metric can be optimized in isolation, because the levers that raise one inevitably move others (Mahapatra, Nijsure, Kaddoum, Hassan, & Yuen, 2016).

The purpose of this article is to rethink the three objectives as a single negotiated operating region rather than a ranked list. We do not propose a new physical layer scheme. Instead we synthesize the evidence from several active research lines, classify the principal enabling mechanisms by how they reshape the energy, reliability, and latency surface, and offer a conceptual framework intended to make the consequences of each design choice explicit. The aim is a sharper account of when a green technique genuinely expands the feasible region and when it only slides the system along an existing frontier.

The contributions of this article are fourfold. First, it consolidates the metrics used to quantify energy, reliability, and latency and clarifies how they interrelate. Second, it organizes the principal green communication mechanisms by their trade-off signatures, distinguishing those that reposition the operating point from those that expand the feasible region. Third, it maps the requirements of major IoT application domains onto the trade-off space, showing that the correct operating point is a property of the workload rather than the network alone. Fourth, it proposes a unified framework and an associated assessment discipline for reasoning about green techniques without double counting their benefits.

The remainder of the paper is organized as follows. Section 2 sets the scope, traces the evolution of green communication, and describes the performance envelope of next generation deployments. Section 3 reviews the metrics of efficiency, reliability, and latency. Section 4 develops the trade-off space, first as pairwise tensions and then as a combined region. Section 5 examines the enabling mechanisms and their effects. Section 6 maps application domains onto the trade-off space. Section 7 discusses the implications, Section 8 sets out open challenges, and Section 9 concludes.

2. Background and Scope

2.1. The evolution of green communication

Green communication did not begin with the IoT. Its intellectual roots lie in the energy efficiency concerns of cellular networks, where the operational cost and environmental impact of base station power consumption motivated a body of work on energy efficient design and on the relationship between energy and spectral efficiency. A survey of energy efficiency trade-off mechanisms for wireless green communication catalogued the early techniques, including sleep modes, cell switching, and radio resource management, and emphasized that each technique trades energy against some other quantity rather than delivering it for free (Mahapatra *et al*., 2016). As the field matured, the emphasis broadened from the network operator's energy bill to a systemic view that includes device lifetime, embodied energy, and renewable integration.

The transition to the IoT shifted the locus of concern from a relatively small number of high power base stations to a vast number of low power endpoints. Investigations of energy saving practices for the IoT identified a recurring set of levers, including efficient hardware, protocol redesign, topology control, and harvesting, while noting that the sheer device count makes even small per device inefficiencies consequential at the system level (Arshad, Zahoor, Shah, Wahid, & Yu, 2017). The contemporary framing, consolidated in recent reviews, treats green IoT as a set of coordinated interventions spanning the device, the link, and the network, all aimed at delivering required service quality at minimum energy and environmental cost (Alsharif *et al*., 2023).

2.2. Green communication in the IoT context

Green communication refers to the design of networking systems that deliver required service quality while minimizing energy consumption and the associated environmental cost. In the IoT context this objective acquires specific texture. Endpoints are frequently battery powered or energy harvesting, computationally constrained, and deployed in large numbers, so per device savings multiply across the population and node lifetime becomes a system level metric rather than a device level convenience (Alsharif *et al*., 2023). Reviews of the field group the available techniques into several families: energy efficient machine to machine communication, low power wireless sensor networking, efficient identification and tagging, and efficient hardware at the circuit and microcontroller level (Alsharif *et al*., 2023). A parallel strand emphasizes energy harvesting and renewable integration as a way to relax the battery constraint altogether, while noting the security and reliability complications that intermittent supply introduces (Alazzawi *et al*., 2025).

Two observations follow. First, energy in IoT communication is not a single quantity but a budget distributed across sensing, processing, transmission, and idle listening, with the radio often dominating during active periods. Second, the most effective green techniques are rarely free with respect to the other performance axes, because they act precisely on the mechanisms, such as transmit power, active time, and redundancy, that also govern reliability and latency. This second observation is the seed of the present article: any honest account of a green technique must trace its effect across all three axes rather than reporting only the axis it was designed to improve.

2.3. A taxonomy of energy expenditure in IoT communication

To reason about trade-offs, it helps to disaggregate where energy is actually spent. In a typical battery powered node, energy is consumed in four broad phases: acquisition, when

sensors and front-end electronics gather data; computation, when the data are processed, compressed, or encoded; transmission, when the radio transmits and the power amplifier dominates; and idle listening, when the radio is awake but not actively transmitting, waiting for a channel or a command. Comprehensive surveys of energy efficient computing for the IoT note that the relative weight of these phases varies with the application, and that the largest savings often come from eliminating idle listening rather than from optimizing transmission alone (Alsharif, Kelechi, Jahid, Kannadasan, Singla, Gupta, & Geem, 2024).

This disaggregation matters for the trade-off argument because different mechanisms act on different phases, and a mechanism that saves energy in one phase may add latency or reliability risk in another. Duty cycling and wake up radio target idle listening, computation offloading targets the computation phase, and finite blocklength coding and power control target transmission. A coherent green design must therefore reason about the phase it is optimizing and the cross-phase consequences, rather than treating energy as a single fungible quantity.

2.4. The next generation IoT performance envelope

The performance envelope expected of next generation IoT is broader and more demanding than that of earlier sensor networks. Standardization activity for the 2030 horizon treats sustainability as a cross-cutting requirement rather than an afterthought, placing energy efficiency alongside reliability, latency, and connection density as a primary key performance indicator (ITU-R, 2023). This elevation matters because it forbids the historical practice of optimizing throughput or reliability first and accounting for energy later; energy becomes a constraint to be satisfied rather than a residual to be minimized after the fact.

Within this envelope, the URLLC service class is the sharpest expression of the reliability and latency demand. Achieving sub millisecond latency at very low error probability requires operation in the finite blocklength regime, where the classical assumption that coding rate can approach channel capacity no longer holds and short packets must be coded conservatively (Polyanskiy, Poor, & Verdu, 2010; Durisi *et al*., 2016). The penalty for reliability in this regime is paid in spectral efficiency and, indirectly, in energy, because more channel uses or higher power are needed to close the gap. Foundational treatments of energy efficient design for cellular and IoT systems make clear that energy efficiency cannot be maximized in isolation from these spectral and reliability constraints (Buzzi *et al*., 2016).

3. Metrics and the Meaning of Efficiency

Before the trade-offs can be analyzed, the quantities being traded must be defined, because much of the confusion in the literature stems from inconsistent metrics. This section reviews the standard measures along each axis and the relationships among them.

3.1. Energy and energy efficiency

The most common metric of communication energy efficiency is the number of bits reliably delivered per unit of energy, often expressed in bits per joule. This metric is attractive because it normalizes energy by useful work, but it is sensitive to how reliability is defined, since bits delivered in error do not count as useful work, and to the time horizon, since idle energy accrues even when no bits are sent. At the device level, the related metric of node lifetime, the expected time before a battery is exhausted, captures the practical consequence of energy efficiency for deployments that cannot be easily recharged (Arshad *et al*., 2017). Surveys of green computing for the IoT extend the accounting to the computation phases, arguing that a complete energy metric must include processing and idle energy and not only transmission energy (Alsharif *et al*., 2024).

3.2. Reliability and latency

Reliability is typically expressed as a target probability that a packet is delivered correctly within a deadline, or equivalently as a maximum tolerable error or outage probability. In the URLLC setting these targets are extreme, with error probabilities reaching one in a hundred thousand or lower, which is what makes the finite blocklength penalty so consequential (Durisi *et al*., 2016). Latency is the time from when a packet is generated to when it is correctly received, and for control applications the relevant quantity is often not the average but a high percentile or a hard deadline, because a single late packet can cause a control loop to fail. The pairing of a deadline with a reliability target is the defining feature of the service class and the reason the two axes cannot be considered independently (Bennis *et al*., 2018). Beyond the physical layer, reliability at the system scale also depends on capacity planning and performance management, since enterprise scale rollouts must sustain dependable service under growing and variable demand (Oteri & Edivri, 2024; Adelanwa *et al*., 2023).

3.3. The energy and spectral efficiency relationship

Energy efficiency and spectral efficiency are linked by a relationship that is central to any green design. Increasing the rate at which bits are sent over a fixed bandwidth generally requires more energy per unit time, and beyond a point the energy per bit rises sharply, so there is a regime in which pushing for higher spectral efficiency is energetically expensive. Foundational analyses of energy efficient techniques for next generation networks formalize this relationship and show that the energy efficient operating point is rarely the spectrally efficient one, which means that a network tuned for peak throughput is unlikely to be green and vice versa (Buzzi *et al*., 2016). When a reliability constraint is added through the finite blocklength penalty, the relationship tightens further, because the conservative coding needed for reliability lowers the achievable spectral efficiency and so raises the energy per delivered bit (Polyanskiy *et al*., 2010).

4. The Energy, Reliability, and Latency Trade-off Space

4.1. Pairwise tensions

The cleanest way to see the conflict is to hold one axis fixed and observe the other two. Consider first energy against reliability. For a fixed latency budget, raising reliability typically means lowering the coding rate, adding diversity, or increasing transmit power, all of which raise energy per delivered bit. The relationship is governed by the finite blocklength bounds, which quantify how much rate must be sacrificed to reach a target error probability with a given number of channel uses (Polyanskiy *et al*., 2010). The penalty is steepest exactly in the short packet regime that IoT control traffic occupies, so the energy cost of the last orders of magnitude of reliability is disproportionately large.

Consider next energy against latency. Hybrid automatic

repeat request (HARQ) illustrates the subtlety. Retransmissions with incremental redundancy can deliver reliability at lower average energy than a single conservative transmission, but only when the latency budget is generous enough to absorb feedback delay and additional rounds. When the budget is tight, the time consumed by feedback can exceed the benefit, and a single shot transmission becomes preferable on energy grounds (Avranas *et al*., 2018). The energy minimizing strategy therefore depends on where the latency constraint sits relative to the feedback delay, which is a configuration dependent rather than a universal answer.

Consider finally reliability against latency. Tightening the latency bound shrinks the blocklength available for coding, which raises the error probability for a fixed rate and power. Recovering the lost reliability requires either more power, more bandwidth, or additional diversity, each of which has an energy implication. Resource allocation studies in the short blocklength regime show that the feasible reliability is a steep function of the allowed delay, so small relaxations in latency can yield disproportionate reliability or energy gains (Sun, She, Yang, Quek, Li, & Vucetic, 2019; She, Yang, & Quek, 2018). This steepness is itself a design opportunity, because it means that a workload that can tolerate a slightly looser deadline may be served far more cheaply.

4.2. From pairs to a combined region

The three pairwise views share a structural feature: the axis held fixed in each view is itself a decision variable. This is why pairwise analysis can mislead. A policy that appears to improve the energy and reliability pair may do so only by quietly consuming the latency budget, leaving the system with no slack for the next disturbance. We therefore advocate treating the three objectives as defining a single feasible region in a three-dimensional space, where each physical or architectural lever moves the operating point and, more importantly, reshapes the boundary of what is feasible at all.

In this view the design question is twofold. The first part is positional: given a fixed set of mechanisms, where on the existing frontier should the system operate, which is a matter of weighting the three objectives according to workload semantics. The second part is structural: which mechanisms actually expand the frontier, delivering better reliability and latency at the same energy or lower energy at the same service quality. Much of the confusion in the literature comes from conflating these two questions, because a technique that only repositions the operating point is easily mistaken for one that improves the underlying capability. Formally, selecting an operating point within the feasible region is a constrained resource allocation problem, and related work on constraint satisfaction and on the approximation complexity of such allocation problems indicates that the underlying optimization is rarely trivial even when the region is well characterized (Ahmed, Odejobi, & Oshoba, 2019; Odejobi, Hammed, & Ahmed, 2019). This repositioning versus expansion distinction is taken up again in the discussion.

Table 1 summarizes how the principal mechanism families, examined in detail in Section 5, tend to act on the three axes. The entries describe the typical direction of effect rather than guaranteed outcomes, since the realized result depends on workload and channel.

Table 1: Trade-off signatures of principal green communication mechanisms.

Mechanism family	Energy effect	Reliability effect	Latency effect
Conservative coding (low rate, finite blocklength)	Increases energy per bit	Improves	Neutral to worse
HARQ with incremental redundancy	Lower average energy if budget allows	Improves	Adds feedback and rounds
Duty cycling and sleep scheduling	Reduces idle energy	Risk of missed events	Adds wake up delay
Wake up radio (on demand)	Cuts idle listening sharply	Maintains reactivity	Low wake up delay
Wireless energy transfer and harvesting (SWIPT)	Relaxes battery limit	Supply intermittency risk	Harvest time competes with data time
Edge computation offloading	Saves device energy	Depends on link and server	Can cut or add delay
AI driven adaptive control (6G)	Context dependent savings	Can improve via prediction	Can reduce via anticipation

5. Enabling Mechanisms and Their Trade-offs

5.1. Finite blocklength coding and retransmission control

The physical layer is where the three-way tension is most exactly characterized. In the finite blocklength regime the maximum reliable coding rate falls below the Shannon limit by an amount that grows as the blocklength shrinks and the reliability target tightens (Polyanskiy *et al*., 2010). For short packet IoT traffic this means that the energy needed to deliver a payload at a given reliability is a sensitive function of how many channels uses the latency budget permits. Treatments of short packet communication for the mission critical regime make this the central design fact, because the asymptotic intuitions that guide throughput-oriented design do not transfer to the short packet setting (Durisi *et al*., 2016).

Retransmission control sits on top of this. Incremental redundancy HARQ can be energy efficient because later rounds add only the redundancy that the channel realization actually requires, rather than provisioning for the worst case up front. The catch is temporal. Each round carries feedback delay, and once that delay consumes a large fraction of the latency budget the apparent energy advantage of HARQ reverses, favoring a single conservative transmission instead (Avranas *et al*., 2018). Closed form analyses of incremental redundancy in the short blocklength regime confirm that the optimal number of rounds, the blocklength per round, and the power per round must be jointly chosen, and that the optimum migrates as the latency constraint moves (Makki, Svensson, & Zorzi, 2014).

The practical lesson is that retransmission is not categorically green or expensive. It is green only within a window of latency budgets and feedback delays, and outside that window it taxes the very axis it was meant to spare. Resource allocation methods for the short blocklength regime address this by jointly configuring blocklength, power, and bandwidth to meet the reliability and latency targets at minimum resource cost, which is the physical layer

expression of the combined region argued for in Section 4 (Sun *et al*., 2019; She *et al*., 2018).

5.2. Duty cycling, sleep scheduling, and wake up radio

If idle listening dominates the energy budget, the most direct remedy is to keep the radio off as much as possible. Duty cycling switches the radio between sleep and active states on a schedule, which sharply reduces idle energy but introduces two costs. First, a node that is asleep when an event occurs cannot report it until it wakes, adding latency bounded by the sleep interval. Second, scheduled sleep risks missing events entirely if they fall outside the listening window, which is a reliability cost. The schedule therefore directly trades idle energy against latency and event capture, and tuning it requires knowledge of the traffic pattern.

Wake up radio addresses the limitation of fixed schedules by adding a very low power receiver whose only job is to listen for a wake-up signal and rouse the main radio on demand. This decouples reactivity from the duty cycle, so a node can sleep deeply yet respond quickly when addressed, which improves the idle energy and latency trade-off simultaneously rather than exchanging one for the other (Rup & Bajic, 2024). In the language of the framework, wake up radio is closer to a frontier expanding mechanism than a repositioning one, because it relaxes the otherwise tight coupling between sleep depth and responsiveness. The cost is added hardware and the energy of the always on wake-up receiver, which must be small enough not to erode the savings, and reviews of the technology emphasize that this balance is what determines its practical value for sustainable industrial IoT (Rup & Bajic, 2024).

5.3. Wireless energy transfer and energy harvesting

If the radio is the dominant energy sink, another response is to change where the energy comes from. Simultaneous wireless information and power transfer (SWIPT) and related wireless powered communication schemes deliver energy and information over the same radio resource, aiming to free endpoints from finite batteries (Varshney, 2008; Zhang & Ho, 2013). The foundational result is that there is an intrinsic rate to energy trade-off: a receiver cannot extract maximum information and maximum power from the same signal at once, so architectures must split the signal in time or power between decoding and harvesting (Varshney, 2008; Zhang & Ho, 2013).

Mapped onto our three axes, harvesting relaxes the energy constraint at the source but introduces two new tensions. First, harvest time competes with data time, so a node that spends longer charging has less of the latency budget left for transmission, coupling the energy supply directly to latency. Second, harvested supply is intermittent and channel dependent, which injects variability into reliability because a node may lack the energy to transmit at the required power when an event occurs (Alazzawi *et al*., 2025). Harvesting therefore does not remove the trade-off region; it relocates part of it from the energy axis to the latency and reliability axes, and the net benefit depends on whether the application can tolerate the resulting variability.

Surveys of green computing and harvesting for the IoT add that the choice of harvesting source, whether radio frequency, solar, thermal, or mechanical, changes the statistics of supply and therefore the shape of the induced trade-off, so harvesting is not a single technique but a family with distinct signatures (Alsharif *et al*., 2024; Alazzawi *et al*., 2025). Where renewable supply is integrated at the installation level, the design problem extends to load distribution and microgrid sizing, as studied for hybrid solar and hybrid solar diesel deployments in resource constrained settings, which shape the energy availability that downstream communication can assume (Sunday & Omoegun, 2019; Ijiga, Oladoye, Bamigwojo, & Ogboji, 2023). For applications with hard guarantees, the open problem is to bound service quality under stochastic supply, a point we return to in Section 8.

5.4. Edge and fog computation offloading

Not all energy in an IoT task is spent on communication. Computation can be offloaded from a constrained device to a nearby edge or fog server, trading transmission energy and latency for local processing energy. Whether offloading is favorable depends on the size of the task, the quality of the link, and the load on the server, which makes it a natural decision problem rather than a fixed policy. Comprehensive surveys of green computing for the IoT treat edge, fog, and cloud offloading as complementary tiers, each with a different energy and latency profile, and stress that the right tier depends on the task and the network state (Alsharif *et al*., 2024). The cloud and edge tiers carry their own energy and reliability considerations, from energy efficient virtual machine placement in data centers to the design of cloud native microservices in edge architectures, both of which influence the cost of hosting offloaded work (Ahmed & Odejobi, 2018; Idika *et al*., 2021).

Studies that pose offloading as a joint minimization of device energy and task latency show that the optimal decision varies with channel state and queue conditions, and that no static rule dominates across regimes (Ale, Zhang, Fang, Chen, Wu, & Li, 2021; Zhao, Wang, Xia, & Fan, 2020). Queue conditions at the server and high concurrency among many simultaneous requests are central to this decision, as queueing and performance evaluation models for throughput and concurrency make clear (Akinola, Adesanya, Okafor, & Dako, 2024; Odejobi & Ahmed, 2018). Because the decision space is large and the environment is nonstationary, learning based controllers have become a common tool. Reinforcement learning agents can adapt offloading and resource allocation to changing conditions, balancing the energy saved at the device against the latency incurred over the link and at the server (Zhao *et al*., 2020; Akhavan, Esmaeili, Yousefi, Sun, Zarkesh-Ha, Badnava, & Devetsikiotis, 2022).

The trade-off signature of offloading is thus conditional: it can simultaneously cut device energy and end to end latency when the link is good and the server is lightly loaded, but it can add latency and even net energy when those conditions fail. Treating offloading as unconditionally green is a category error of the kind Section 4.2 warns against, and the value of learning-based control is precisely that it lets the system recognize which regime it is in and act accordingly.

5.5. Architectural and AI driven levers in 6G

At the network level, the 6G research agenda introduces levers that operate above the link. Energy aware network management, selective powering down of underused infrastructure, and resource sharing aim to reduce the energy floor of the network without degrading the service offered to active users. The sustainability literature for next generation networks frames energy efficiency as a design principle that must be embedded across scenarios rather than optimized

after the fact, which aligns with the elevation of energy to a primary indicator in the 2030 framework (ITU-R, 2023; Buzzi *et al*., 2016).

The distinctive new lever is anticipation. When a controller can predict traffic, channel evolution, or event arrival, it can preposition resources so that reliability and latency targets are met with less reactive over provisioning, which is where much energy is wasted. In principle this expands the feasible region rather than merely repositioning the operating point, because better information reduces the margin that must be held against uncertainty. The promise is real but conditional on the quality of prediction and the energy cost of the intelligence itself, and the literature is still consolidating the conditions under which the net effect is positive. The same learning machinery that improves offloading decisions can be applied to scheduling, power control, and sleep management, which suggests that anticipation is a general-purpose lever rather than a point solution (Zhao *et al*., 2020; Akhavan *et al*., 2022). Architectures for machine learning enabled predictive energy management over IoT sensor networks illustrate this direction, using learned models to forecast demand and preposition energy and communication resources (Kumuyi, Uzoka, Akeju, & Ozowara, 2024). Related work on predictive capacity planning and on machine learning driven cloud resource scaling shows that the same forecasting machinery improves resource utilization in large multi stakeholder systems (Edivri & Oteri, 2022; Ahmed, Odejobi, & Oshoba, 2020).

6. Application Domains and Trade-off Weightings

The framework developed here insists that the correct operating point depends on what the traffic means, not on the network alone. This section illustrates the point by mapping representative IoT application domains onto the three axes. The mapping is qualitative and indicative; its purpose is to show that the same network serving different workloads should operate at different points, and that a green technique appropriate for one domain may be inappropriate for another.

Industrial automation and control sit at the demanding corner of the space. A missed or late command can halt a production line or damage equipment, so reliability and latency are weighted heavily and energy can be spent freely during active control. Here the finite blocklength penalty and tight latency budgets dominate, and techniques that add delay, such as deep duty cycling, are inappropriate for the control path even though they may suit auxiliary monitoring sensors on the same factory floor (Rup & Bajic, 2024; Sun *et al*., 2019).

Vehicular and tactile applications resemble industrial control in their intolerance of delay and failure but add mobility, which makes the channel more variable and raises the value of anticipation. Predictive resource management is especially attractive here because the cost of reactive over provisioning is high when the channel changes quickly (Akhavan *et al*., 2022).

Smart metering and environmental monitoring sit at the opposite corner. Reports are periodic, individually non critical, and tolerant of delay and occasional loss, so energy and node lifetime dominate the weighting. Aggressive duty cycling, harvesting, and data aggregation are well matched here, and the latency and reliability costs they incur are acceptable given the workload semantics (Arshad *et al*., 2017; Alazzawi *et al*., 2025).

Healthcare and assisted living occupy an intermediate and heterogeneous position. Routine telemetry tolerates delay and favors energy savings, while alarm conditions demand reliability and timeliness without compromise. This heterogeneity within a single deployment is the strongest argument for treating the operating point as workload dependent and adaptive rather than fixed, since no single static configuration serves both the telemetry and the alarm well.

Table 2 summarizes the indicative weightings. The point is not that these weightings are precise, but that they differ, and that a green design that ignores the difference will misallocate effort by applying a uniform policy to non-uniform needs.

Table 2: Indicative trade-off weightings across representative IoT application domains.

Application domain	Energy weight	Reliability weight	Latency weight
Industrial automation and control	Low	High	High
Vehicular and tactile	Low to medium	High	High
Smart metering and monitoring	High	Low to medium	Low
Healthcare telemetry (routine)	High	Medium	Low to medium
Healthcare alarms (critical)	Low	High	High

7. Discussion

Synthesizing the preceding sections, energy, reliability, and latency are best treated as three jointly negotiated quantities constrained by workload semantics and resource availability, rather than as a fixed ranking in which lower priority objectives are surrendered to higher priority ones. Three implications follow.

First, the achievable combinations of energy, reliability, and latency form a feasible region whose boundary is set by the active mechanisms, so design proceeds in two stages: choose mechanisms that shape a favorable boundary, then select an operating point on that boundary according to the application.

Second, mechanisms differ in kind. Some only move the operating point along an existing boundary, exchanging one axis for another, while others expand the boundary and improve several axes at once; conservative coding and naive duty cycling tend to reposition, whereas wake up radio, informed anticipation, and well matched offloading can expand, and an honest evaluation should state which effect it claims. Third, the correct operating point is a property of the traffic rather than the network alone, because an alarm that must never be missed weights reliability and latency heavily and can spend energy freely, while a periodic environmental report weights energy heavily and tolerates delay and occasional loss, so a design that ignores these semantics will misallocate effort.

These implications translate into a practical way of assessing any green technique. One first identifies the axis the technique is designed to improve and the energy phase in which it acts, whether acquisition, computation, transmission, or idle listening. One then traces the consequences onto the other two axes, since few interventions are confined to a single axis and a claimed energy saving that silently consumes the latency budget is a

repositioning rather than an improvement. Finally, one judges whether the technique repositions the operating point or expands the feasible region and checks the claim against the workload semantics described earlier rather than against a generic benchmark. Reasoning in this way prevents the common error of crediting a technique with a system level improvement when it has only moved cost out of view, and it clarifies why some techniques are nearly universal, such as wake up radio, while others are strictly conditional, such as offloading, whose benefit depends on link and server state. None of this replaces quantitative optimization; it frames the optimization correctly, so that all three axes and the workload weighting enter the objective from the start.

8. Open Challenges and Research Directions

Several gaps stand between the framework above and its routine application in design.

- **Joint modeling across layers:** Most analyses isolate one mechanism, for example HARQ at the link or offloading at the network. A practical green IoT design composes several mechanisms whose trade-off signatures interact, and there is no widely adopted model that captures energy, reliability, and latency jointly across the acquisition, computation, transmission, and idle phases and across the physical, link, and network layers.
- **Measurement and benchmarking:** Repositioning and expansion are easy to confuse precisely because reported gains often fix the unexamined axis at a convenient value. The field would benefit from benchmarks that vary all three axes explicitly and report the full operating region rather than a single point, together with standardized energy accounting that includes idle and computation energy.
- **Energy cost of intelligence:** Anticipatory and learning based controllers promise frontier expansion, but the energy and latency cost of running the intelligence, including training, inference, and the communication of model updates, is frequently excluded from the accounting. Net effect studies that internalize this cost are needed before the promise can be relied upon (Akhavan *et al*., 2022).
- **Harvesting under hard guarantees:** Energy harvesting relaxes the supply constraint but injects variability that is difficult to reconcile with hard reliability and latency guarantees. Design methods that provide bounded service quality under stochastic supply, possibly by combining harvesting with small reserves and adaptive scheduling, remain an open problem (Alazzawi *et al*., 2025; Alsharif *et al*., 2024).
- **Workload aware adaptation:** Section 6 argues that the operating point should track workload semantics, yet most deployments apply a uniform policy. Mechanisms that classify traffic by its semantics in real time and adapt the energy, reliability, and latency weighting accordingly are an open and practically important direction.
- **Standardization of sustainability as a constraint:** With sustainability elevated to a primary indicator for the 2030 horizon, the open question is how to express energy as a hard constraint in resource allocation rather than a soft objective, so that conformance can be specified and tested (ITU-R, 2023).

9. Conclusion

Green communication for next generation IoT is often presented as a contest in which energy efficiency must be wrung from a system at the expense of reliability or speed. This article has argued that the more accurate and more useful picture is one of a jointly negotiated operating region whose shape, not merely whose operating point, is the object of design. The pairwise trade-offs that dominate the literature are real but partial, because each holds fixed an axis that is in fact a decision variable, and a gain claimed on one pair frequently reappears as a cost on the third.

By reviewing the metrics that define each axis, classifying mechanisms according to whether they reposition the operating point or expand the feasible region, mapping application domains onto the trade-off space, and tying the correct operating point to workload semantics, the framework offered here aims to make the consequences of green design choices explicit and to guard against double counting their benefits. The principal open problems are matters of cross layer modeling, honest measurement, workload aware adaptation, and the expression of sustainability as a testable constraint. Progress on these will determine whether the next generation of IoT can be both green and dependable rather than green at the expense of dependability.

References

1. Adelanwa A, Basnet A, Anene UN. Performance intelligence models for optimization and outcome measurement in large scale public services. Shodhshauryam, International Scientific Refereed Research Journal. 2023;6(1):319-354.
2. Ahmed KS, Odejobi OD. Resource allocation model for energy efficient virtual machine placement in data centers. IRE Journals. 2018;2(3):1-10.
3. Ahmed KS, Odejobi OD, Oshoba TO. Algorithmic model for constraint satisfaction in cloud network resource allocation. IRE Journals. 2019;2(12):516-532.
4. Ahmed KS, Odejobi OD, Oshoba TO. Predictive model for cloud resource scaling using machine learning techniques. Journal of Frontiers in Multidisciplinary Research. 2020;1(1):173-183.
5. Akhavan Z, Esmaeili M, Yousefi M, Sun X, Zarkesh-Ha P, Badnava B, *et al*. Deep reinforcement learning for online latency aware workload offloading in mobile edge computing. arXiv preprint arXiv:2209.05191. 2022.
6. Akinola AS, Adesanya OS, Okafor CM, Dako OF. Value chain automation in beverage logistics: Throughput, capacity, and cost avoidance via queueing models. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2024;10(4):1112-1132.
7. Alazzawi Q, *et al*. Green IoT: Energy efficiency, renewable integration, and security implications. IET Networks. 2025. doi:10.1049/ntw2.70003.
8. Ale L, Zhang N, Fang X, Chen X, Wu S, Li L. Delay aware and energy efficient computation offloading in mobile edge computing using deep reinforcement learning. arXiv preprint arXiv:2103.07811. 2021.
9. Alsharif MH, Jahid A, Kelechi AH, Kannadasan R. Green IoT: A review and future research directions. Symmetry. 2023;15(3):757. doi:10.3390/sym15030757.

10. Alsharif MH, Kelechi AH, Jahid A, Kannadasan R, Singla MK, Gupta J, *et al*. A comprehensive survey of energy efficient computing to enable sustainable massive IoT networks. Alexandria Engineering Journal. 2024;91:12-29. doi:10.1016/j.aej.2024.01.067.
11. Arshad R, Zahoor S, Shah MA, Wahid A, Yu H. Green IoT: An investigation on energy saving practices for 2020 and beyond. IEEE Access. 2017;5:15667-15681.
12. Avranas A, Kountouris M, Ciblat P. Energy latency tradeoff in ultra reliable low latency communication with retransmissions. IEEE Journal on Selected Areas in Communications. 2018;36(11):2475-2485.
13. Bennis M, Debbah M, Poor HV. Ultra reliable and low latency wireless communication: Tail, risk, and scale. Proceedings of the IEEE. 2018;106(5):1834-1853.
14. Buzzi S, I CL, Klein TE, Poor HV, Yang C, Zappone A. A survey of energy efficient techniques for 5G networks and challenges ahead. IEEE Journal on Selected Areas in Communications. 2016;34(4):697-709.
15. Durisi G, Koch T, Popovski P. Toward massive, ultrareliable, and low latency wireless communication with short packets. Proceedings of the IEEE. 2016;104(9):1711-1726.
16. Edivri J, Oteri O. Predictive capacity planning and resource utilization forecasting models for multi program and multi stakeholder IT portfolios. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2022;8(4):826-845.
17. Idika CN, Salami EO, Ijiga OM, Enyejo LA. Deep learning driven malware classification for cloud native microservices in edge computing architectures. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2021;7(4).
18. Ijiga OM, Oladoye SO, Bamigwojo OV, Ogboji AJ. Techno economic evaluation of hybrid solar diesel microgrids for underserved communities using simulation based load forecasting. Engineering Science and Technology Journal. 2023;4:1-28.
19. International Telecommunication Union, Radiocommunication Sector (ITU-R). Recommendation ITU-R M.2160: Framework and overall objectives of the future development of IMT for 2030 and beyond. Geneva: ITU; 2023.
20. Kumuyi O, Uzoka E, Akeju B, Ozowara DE. Architecture for machine learning enabled predictive energy management using IoT sensor networks. International Journal of Advanced Multidisciplinary Research and Studies. 2024;4(6):3138-3149.
21. Mahapatra R, Nijsure Y, Kaddoum G, Hassan NU, Yuen C. Energy efficiency tradeoff mechanism towards wireless green communication: A survey. IEEE Communications Surveys and Tutorials. 2016;18(1):686-705.
22. Makki B, Svensson T, Zorzi M. Finite block length analysis of the incremental redundancy HARQ. IEEE Wireless Communications Letters. 2014;3(5):529-532.
23. Odejobi OD, Ahmed KS. Performance evaluation model for multi tenant Microsoft 365 deployments under high concurrency. IRE Journals. 2018;1(11):92-107.
24. Odejobi OD, Hammed NI, Ahmed KS. Approximation complexity model for cloud based database optimization problems. IRE Journals. 2019;2(9):1-10.
25. Oteri O, Edivri J. Applied performance optimization frameworks for managing high demand enterprise technology programs and system rollouts. Gyanshauryam, International Scientific Refereed Research Journal. 2024;7(1):188-210.
26. Polyanskiy Y, Poor HV, Verdu S. Channel coding rate in the finite blocklength regime. IEEE Transactions on Information Theory. 2010;56(5):2307-2359.
27. Rup C, Bajic E. Green and sustainable industrial Internet of Things systems leveraging wake up radio to enable on demand IoT communication. Sustainability. 2024;16(3):1160. doi:10.3390/su16031160.
28. She C, Yang C, Quek TQS. Joint uplink and downlink resource configuration for ultra reliable and low latency communications. IEEE Transactions on Communications. 2018;66(5):2266-2280.
29. Stankovic JA. Research directions for the Internet of Things. IEEE Internet of Things Journal. 2014;1(1):3-9.
30. Sun C, She C, Yang C, Quek TQS, Li Y, Vucetic B. Optimizing resource allocation in the short blocklength regime for ultra reliable and low latency communications. IEEE Transactions on Wireless Communications. 2019;18(1):402-415.
31. Sunday EA, Omoegun GO. Optimizing electrical load distribution for hybrid solar installations in developing economies. International Journal of Scientific Research in Mechanical and Materials Engineering. 2019;3(6):27-47.
32. Varshney LR. Transporting information and energy simultaneously. In: Proceedings of the IEEE International Symposium on Information Theory (ISIT). Toronto, Canada; 2008. p. 1612-1616.
33. Zhang R, Ho CK. MIMO broadcasting for simultaneous wireless information and power transfer. IEEE Transactions on Wireless Communications. 2013;12(5):1989-2001.
34. Zhao R, Wang X, Xia J, Fan L. Deep reinforcement learning based mobile edge computing for intelligent Internet of Things. arXiv preprint arXiv:2008.00250. 2020.