International Network Generations Roadmap (INGR)

An IEEE 5G and Beyond Technology Roadmap

Energy Efficiency

1st Edition White Paper

FutureNetworks.ieee.org/Roadmap

International Network Generations Roadmap (INGR)

Chapters:

- Applications and Services
- Edge Automation Platform
- Hardware
- Massive MIMO
- Satellite
- Standardization Building Blocks
- Millimeter Wave and Signal Processing
- Security
- Testbed

White Papers:

Available 1st Quarter 2020

- Artificial Intelligence & Machine Learning
- Deployment
- Energy Efficiency
- Optics
- Systems Optimization

Download the entire document at

FutureNetworks.ieee.org/Roadmap

An exclusive benefit for subscribers to the

IEEE Future Networks Initiative.

Table of Contents

ABSTRACT

Key words:

Energy Efficiency, 5G Energy Gap, 5GEG, Power Value Chain, PVC, Power Cost Factor, PCF, Energy Harvesting, Sustainable Power, Embodied Energy, 5G Economic Gap (5GEcG), 5G Equality Gap (5GEqG), 5G Derate Factor (5GDF)

ENERGY EFFICIENCY

1. INTRODUCTION

This white paper gives a broad summary of what one can expect from the more in depth roadmap effort for this topic. It describes a high-level perspective and projection of the topic's technology status, in particular the challenges and gaps to be explored and reported in the 2020 edition of the IEEE INGR roadmap. The scope and stakeholders are summarized. Any expected linkages among the other INGR roadmap working groups are presented.

NOTE: This working group roadmap does not endorse any one solution, company, or research effort.

1.1. CHARTER

The Energy Efficiency (EE) Working Group (WG) is dedicated to ensuring awareness, resources, and proper linkages are captured and disseminated in a meaningful way to enable the most pragmatic (and therefore minimal) utilization of energy and associated carbon footprint for global communications networks (including mobile telephony and fixed IP networks).

1.2. VISION STATEMENT

The ultimate success of any new technology development is intrinsically tied to its energy requirements—be it a battery-operated device or a data center, energy is the "currency" that determines its business viability. Energy infrastructure is often a significant capital expenditure (CAPEX) driver and energy consumption is often the primary operating expenditure (OPEX) driver, increasingly displacing service and maintenance costs from edge to cloud to core. It is also fundamental to sustainability and without sustainability, systems are inherently inequitable. Its optimization will lead to a digital future rich in content and functionality for all to benefit. This vision is accomplished via inclusion in the IEEE Future Networks (FN) International Network Generations Roadmap (INGR) and the critical interactions with the many cross-functional stakeholder areas that are all inexorably dependent on the intricacies of energy architecture, distribution, and utilization.

1.3. SCOPE OF WORKING GROUP EFFORT

The EE WG is committed to education on energy-related issues/concerns/opportunities across all industry stakeholders and associated, extended ecosystems.

Ideally, all industry stakeholders will come to realize the importance of an obsessive focus on optimizing energy efficiency/utilization at every level (i.e., from component to system to network) as a critical area, as early in the development/deployment/standardization processes as possible to maximize positive results when deployed at all scales (i.e., from edge or small cell to the full network and utility levels).

There are currently utilized metrics that have brought important visibility to energy consumption, such as Power Usage Effectiveness (PUE) [1], but they are over-simplified and do not enable the level of granularity that is necessary to fully understand the trade-offs at system level and optimize efficiency across the ecosystem, all the way to the utility-scale. Whether the interest comes from technical, business, and/or sustainability motivations, new metrics such as the concepts and associated critical dependencies of the Power Value Chain (PVC), Power Cost Factor (PCF), and the 5G Energy Gap (5GEG) [2] must be internalized and applied appropriately. We must also provide a mechanism to seamlessly move between technical, economic, and socioeconomic analyses that will ultimately make or break the success of 5G deployments, which is why the additional concepts of the 5G Economic Gap (5GEcG), the 5G Equality Gap (5GEqG), and the 5G Derate Factor (5GDF) are introduced.

Deceivingly, we often disregard small amounts of energy that are consumed "at the edge of the network," without realizing that the farther a device or system is from the power plant and the closer it is to the edge, the higher the multiplication factor of its energy requirement.

Conversely, when it comes to data center energy consumption, there is often misinformation and/or lack of understanding in how to interpret piecemeal efficiencies versus other constraints and how it aggregates to global consumption. Today this total data center contribution is around 1% of global energy consumption [3] [4].

As many diverse stakeholders have an impact on the overall energy efficiency of the network, from energy generation all the way to its distributed consumption across the ecosystem, it is important to define the boundaries of each subsystem involved and how the interactions across boundaries impact the system health. In order to develop recommendations for the optimization of the system performance, it is essential that we develop a quantitative analysis that is global in nature and overcomes the current siloed approach, which publicizes the achievement of local minima, sometimes at the expense of overall performance. This objective requires the creation of metrics, Figure of Merits (FoMs) and Key Performance Indicators (KPIs) that bring commonality in energy characterization to a very diverse landscape and overcome the fragmentation of ownership, which leads to a lack of global optimization.

A roadmap format is an ideal way to accomplish the vision as it provides awareness, guidance, and tiered approaches for near- (~3 years), mid- (~5 years), and long-term (~10+ years) action.

This whitepaper shall serve as an introduction and segue to our ultimate goal of providing a detailed chapter as a novel contribution to the 2nd edition of the INGR (due to be released in late-2020 or early-2021).

The scope of this WG starts with the structure of this whitepaper and expands its content into the full roadmap chapter. This is best summarized by organizing the proposed INGR chapter table of contents, as follows –

1. Introduction
    a. Working Group Mission Statement
    b. Working Group Vision Statement
    c. Working Group Scope
    d. Stakeholders
        i. Key
        ii. Supporting
        iii. Ecosystems
    e. Linkages to Roadmap Content
2. Current State
    a. The Power Value Chain (PVC)
        i. Definition
        ii. Application in Telecommunications
        iii. The Edge Vs. The Core
    b. Network Energy Architecture
        i. What is the true cost of 1 W?

- ii. What is the true cost of 1 mW?
- iii. Global Telecommunications Energy Footprint
- iv. The "5G Energy Gap"
- v. The "5G Economic Gap"
- vi. The "5G Derate Factor"
- vii. The "5G Equality Gap"
- viii. Safety Concerns
- ix. Security Concerns

3. Path to the Future
   - a. Network-Level Energy Analysis
   - b. Key Metrics
     - i. Power Cost Factor (PCF)
   - c. Data Processing Architecture
   - d. Optimizing Energy Utilization
     - i. 3GPP Standard
     - ii. Component-Level
     - iii. System-Level
     - iv. Edge-Level
     - v. Base Station-Level
     - vi. Data Center-Level
     - vii. Network-Level
       1. Heterogeneous Networks (HetNets)
     - viii. Utility Grid-Level
   - e. Role of AI Deep Learning
   - f. Engineering Resources
   - g. Natural Resources / Sustainability
4. Case Studies / Successes
5. Conclusions and Recommendations
   - a. Summary of Conclusions
   - b. Working Group Recommendation
   - c. Follow-on Work
     - i. Embodied Energy
       1. Manufacturing
       2. Application
       3. End of Life
6. Contributors
7. References
8. Acronyms / Abbreviations
9. Appendix

The holistic view of EE touches so many aspects of global ecosystems that it is very difficult to capture all pertinent topics, even the ones of prime importance. Some topics are only touched upon within the scope of this document and this WG's activities, such as the identification, characterization, and assessment of embodied energy through the complete product lifecycle. Embodied energy is a topic not currently receiving sufficient industry attention and may be able to help explain a lot of discrepancies between model and actuals in economic analyses.

A deeper exploration of utility grid and energy generation impacts are beyond the scope of the WG at this time. These materials are assembled with this in mind and attempt to facilitate readers to seek more extensive coverage in other venues as appropriate. The relevant content and linkages here are provided as a conduit to understanding the larger issues at-hand as they relate to large-scale communications networks.

Security is another area having relevance to every aspect of the network, same as energy. Given the need to keep a finite scope around topics directly impacted by, or having a major impact to, energy efficiency, security topics fell outside of the scope for this edition.

1.4. STAKEHOLDERS

The complexity and diversity of the stakeholders is one of the key challenges to the implementation of a coordinated approach to system-wide energy optimization. It is therefore important that we understand their interactions and motivations, so that we can find a way to gain a shared perspective and achieve synergy of intents. This challenging and multifaceted nature of the stakeholders responsible for various aspects of network components (from a black-box perspective) leads to a siloing that inhibits collaborative efforts to bridge gaps.

The system-wide perspective must include a life-cycle assessment of the energy consumption. It is not only the network operation that contributes to it, but also the production and deployment of the hardware infrastructure. Even if new hardware solutions are developed to vastly reduce energy consumption during network operation, stakeholders might keep older equipment during its original intended life span. Hence, there can be long delays between when energy optimizing technology is developed and when it has an appreciable impact on the overall energy consumption.

Figure 1 exemplifies the elements and interactions occurring in the system, and it intentionally includes HW, SW and Operators, as all elements create a meaningful interaction.

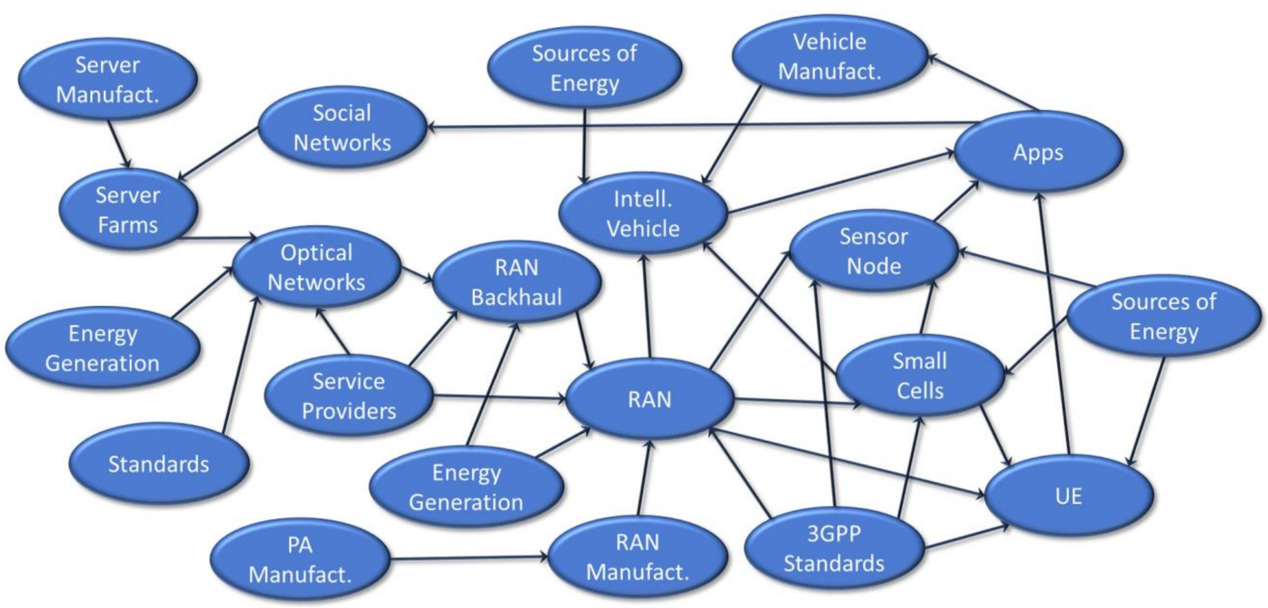

Figure 1 – The Future Networks Ecosystem, courtesy of IoTissimo

Each node in the ecosystem can be analyzed by assessing its optimization objectives, based on the inputs and outputs, as well as its constraints, which can be CAPEX, OPEX, Environmental Regulations, Public Policies, etc.

By viewing all interactions as the effects that outside forces exercise on the physical infrastructure, we can better identify constraints and opportunities. Figure 2 provides a high-level view of the approach that can be adopted to develop a methodology to address them.

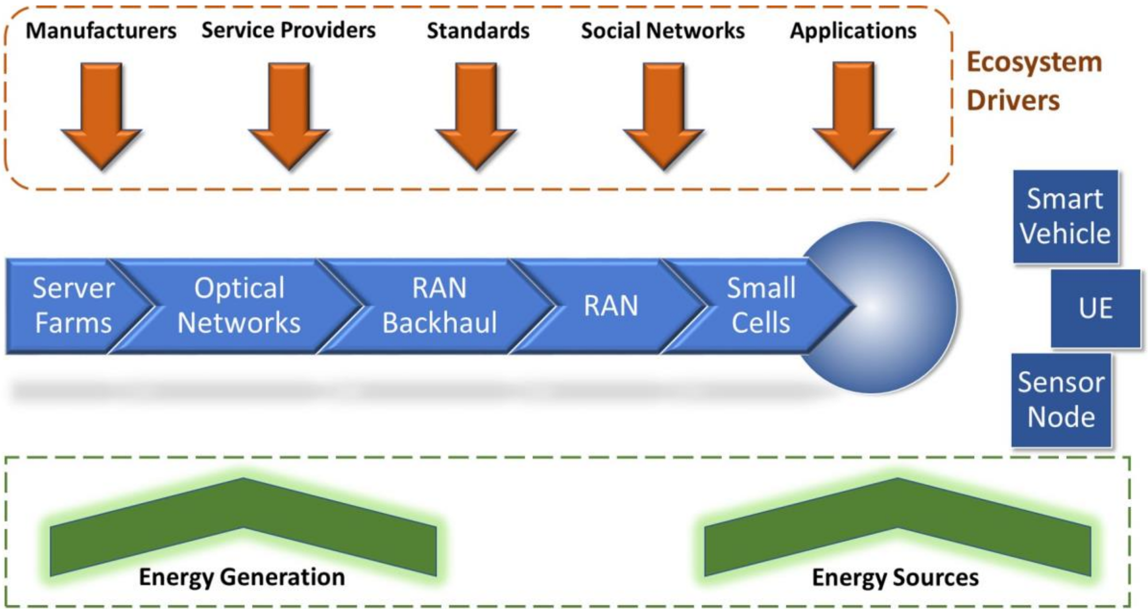

Figure 2 – Outside Forces Affecting the Infrastructure, courtesy of IoTissimo

As the viability of an application is dependent on its total cost, we identify Energy as a measure of such viability, as it affects both the cost of HW (CAPEX), due to power supply and heat mitigation requirement, as well as OPEX.

1.5. Linkages Between Shareholders

There is no stakeholder in the industry without a direct linkage and dependency on this EE WG. A critical focus on energy and power requirements, architecture, distribution, and utilization are essential to the success of any player in the ecosystem, no matter how big or small. Of course, this encompasses all the INGR WGs so it seems unnecessary to list them all out here for that purpose.

That being said, many different flavors of linkages take on different meanings in the multitude of context areas covered by topics in energy and power. For instance, even the concept of energy generation and distribution takes on two major meanings in both the utility- and system-level distributions. If one were to map out all the sources and loads in a city with the distribution lines connecting them, then it will look quite similar to a power distribution scheme within a single system or board. Luckily, many of the fundamental concepts for optimizing utilization and maximizing efficiency are just as applicable at the microwatt scale as they are at the megawatt scale. This also means many of the concepts tabled in this roadmap activity are broad-reaching and refer to technical and economic variables that are highly dynamic and leverageable in their application.

There are numerous aspects of energy management regulated by existing standards. Some relate to best practices in design, manufacturing, test, qualification; and many relate to safety/compliance. The interpretation and application of these many energy-related standards can be challenging because they may relate to a product, its manufacturing process, or both. Several examples and resources are captured in Table 1 below.

Given a case can be made for nearly every possible stakeholder to have a role in this effort (whether unbeknownst to them or not), we should be quick to note the need for pragmatism in how we identify these players and their opportunities for driving improvements in energy efficiency. There are plenty of "low-hanging fruit" opportunities to drive ROI of resources when partnering with organizations (i.e., industry, academic, municipalities, etc.) so that important, yet incremental changes can have a major impact on a global scale. One example of this may be to align with a group already dictating major standards for how networks are defined and another may be to drive design improvements into the radio equipment transporting all the network's data or the core data center hardware crunching all that data and porting it around the globe. Another way to approach is to gain the support of a very large user (i.e., large city digital transformation) encompassing all of the areas related to energy efficiency to serve as a case study and catalyst for others to follow.

Examples of industry organizations focused on energy efficiency are summarized in the table below:

Table 1 – Industry Groups/Resources/Standards with Energy Efficiency Focus and/or Content

ORGANIZATION	CONTACT	YEAR EST.	MAJOR CONTRIBUTION(S)
GreenTouch	https://s3-us-west-2.amazonaws.com/belllabs-microsite-greentouch/index.php@page=home.html	2010 (ended 2015)	• Network-wide Energy Efficiency Assessment • Energy Consumption Simulation Tools
EARTH (Energy Aware Radio and neTwork tecHnologies) Project	https://cordis.europa.eu/docs/projects/cnect/3/247733/080/deliverables/001-EARTHWP2D23v2.pdf	2010 (ended 2012)	• Network-wide Energy Efficiency Assessment • Energy Consumption Database & Simulation Tools
IEEE Future Networks Initiative (*Formerly 5G Initiative*)	https://futurenetworks.ieee.org/	2016	• INGR (this effort) • Network-wide Energy Efficiency Assessment • Tutorials, Workshops, Webinars, Podcasts, Articles, etc.
IEEE Sustainable ICT Initiative (*Formerly GreenICT Initiative*)	https://sustainableict.ieee.org/	2015	• Network-wide Energy Efficiency Assessment • Energy-Efficiency-Focused Standards • Tutorials, Workshops, Webinars, Podcasts, Articles, etc.
PSMA (Power Sources Manufacturers Association)	https://www.psma.com	1985	• Multiple Committees focused on Energy Efficiency • Energy-Efficiency-Focused Database • Numerous Energy-Related Events • Workshops, Webinars, Articles, etc.
IEEE PELS (Power Electronics Society)	https://www.ieee-pels.org/	1988	• Multiple Committees focused on Energy Efficiency • Numerous Energy-Related Events • Tutorials, Workshops, Webinars, Podcasts, Articles, etc.
3GPP™ (Third Generation Partnership Project)	https://www.3gpp.org/specifications	1998	• Mobile Broadband Standard Organization
ITU-T SG5 (International Telecommunication Union – Telecommunication Standardization Sector – Study Group 5)	https://www.itu.int/en/ITU-T/studygroups/2017-2020/05/Pages/default.aspx	1865	• Energy-Efficiency-Focused Standards • Tutorials, Workshops, Webinars, Podcasts, Articles, etc.
ETSI EE (European Telecommunications Standards Institute Environmental Engineering Technical Committee)	https://www.etsi.org/committee/ee	1988	• Energy-Efficiency-Focused Standards • Life Cycle Assessments

The following Cross-cut matrix shown in Figure 3 helps to clarify associations amongst the many working groups of the INGR. The EE WG touches nearly every key portion of the ecosystem.

	INGR Working Group Name	1. Massive MIMO	2. mmWave & Signal Processing	3. Hardware	4. Energy Efficiency	5. Edge Automation Platform	6. Satellite	7. Optics	8. Standardization Building Blocks	9. Testbed	10. System Optimization	11. Deployment	12. Applications & Services	13. Security	14. Connecting the Unconnected	15. Artificial Intelligence/ Machine Learning
		ACCESS				NETWORKS			SYS. & STANDARDS			ENABLERS & USERS				
	ACCESS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Massive MIMO		X	X					X	X						
2	mmWave and Signal Processing	X					X		X	X			X			
3	Hardware	X	X		X	X	X		X	X			X	X		
4	Energy Efficiency	X	X	X		X			X		X	X	X		X	X
	NETWORKS															
5	Edge Automation Platform	X								X			X	X	X	
6	Satellite	X	X	X	X	X		X	X				X	X		X
7	Optics*															
	SYSTEM AND STANDARDS															
8	Standardization Building Blocks	X	X	X	X	X	X	X		X	X	X	X	X	X	X
9	Testbed	X	X	X		X	X		X				X	X		
10	System Optimization*															
	ENABLERS AND USERS															
11	Deployment	X	X							X			X	X		
12	Applications and Services	X	X	X	X	X	X	X	X	X	X	X		X	X	X
13	Security	X	X			X			X	X	X		X		X	X
14	Connecting the Unconnected				X	X	X		X	X	X	?	X	X		
15	Artificial Intelligence/Machine Learning*															

Figure 3 – The INGR WG Cross-cut Matrix, courtesy of IEEE Future Networks

2. CURRENT STATE

Tracking EE in the context of this roadmap is a multifaceted and daunting task because of how much is covered by the very wide umbrella of the topic. Energy must be considered in so many different ways such as a power source, a sustainable resource, a commodity, a waste product (in terms of heat to mitigate), a cost-benefits analysis, and sadly a mere means to an end at times. If we really expand our horizons, then energy can be considered an end to global poverty and disparity if there is ample, safe supply to all.

We therefore introduce here the following concepts:

1. **Power Value Chain (PVC)** is a systematic representation, which describes the energy flow across all the distribution/conversion steps between source and load, that ties together the siloed stakeholders.
2. The **5G Energy Gap (5GEG)** is a hypothetical representation of the disparity between available energy (i.e., sources) and demand (i.e., loads) of the [mostly] "micro-power" devices representing the majority of "things" in the highly-scalable edge space of the network, based on proposed 5G use cases.
3. The **5G Economic Gap (5GEcG)** is a hypothetical representation of the disparity between available power a system can deliver and the increasing load demands on its outputs, which means a power-limited system and/or network component will not be able to utilize all its designed potential and therefore be inhibited from delivering on the calculated economics of the payback period.

4. The **5G Equality Gap (5GEqG)** is a hypothetical representation of the socioeconomic disparity between those that will be able to adapt their infrastructure and end use cases to unanticipated underperformance due to energy-limited (5GEG) and/or economically-limited (5GEcG) factors, and those that will not have the resources to be flexible enough to do so.

In order to have an appreciation for the risk potential of the 5GEG, it helps to do a simple analysis to understand the true cost of power using the point of consumption within the PVC as a frame of reference. Please see an analysis on the true cost of power consumed at the edge and following it all the way back up the PVC as follows:

- What is the **TRUE** cost of **1 mW** received at the edge?

  (All ranges best to worst case.)

  - **1-2 W** = transmitted by antenna of base station/access point
  - **17-50 W** = input of base station/access point
  - **8-15 %** = lost in transmission from power plant
  - **SUMMARY:**

    1 mW EQUIVALENT OF RECEIVED DATA AT THE EDGE REQUIRES 18,000 - 60,000 TIMES THAT POWER GENERATED (or 18 - 60 W) AT THE POWER PLANT.

This would not be such a big challenge if the increase in the number of receiver devices had no significant influence on the total power consumption, but that is no longer the case with 5G.

Until 4G, we lived in the era that Marconi created a century ago: that of a "broadcasting" technology, where the cell transmits uniformly in space. Adding more receivers does have an effect: as the number grows, the requirements on signal-to-noise ratio (SNR) of the tower electronics grows, and so does the power associated with Mixers, analog-to-digital (A/D) converters, etc. Yet, the growth of energy consumption with the number of receivers is relatively contained.

As we enter the 5G era, we no longer achieve higher capacity by improving the signal encoding to get closer to the theoretical Shannon limit; rather, we try to increase the performance of the network along three different trajectories: growth of bandwidth, reduction of latency, and ability to service a number of devices, which is orders of magnitude larger than in the past. Thus, many issues arise, which determine a non-linear growth of the energy consumption.

Increasing the bandwidth in the RF section has the following consequences: (i) the need to operate into the mmWave to access large transmission bands causes an exponential increase of signal loss in the air (also dependent on atmospheric conditions); and (ii) the need to adopt beamforming to deliver enough power at a useful distance with a reasonable power consumption in the cell leads to massive MIMO implementations, which require a large number of elements as well as multiple antennas to cover the 3D horizon. These issues have an exponential compounding effect on power dissipation, due to the required densification of the nodes (cells), the frequency, complexity and performance requirement of the electronics, and the need to provide such performance in environmentally "hostile" environments.

The quest for low latency to enable mission critical services leads to data management and processing at a significantly higher rate than normally acceptable, thus stressing every aspect of the system: from protocol execution to speed of the computing electronics. Additionally, the "mission critical" aspect of these applications forces both redundancy and guaranteed QoS, which increase complexity and cost. Furthermore, achieving such low latency requires very large computing power to be available close to the Edge rather than concentrated in the Cloud, thus requiring an enormous growth of the infrastructure.

Particularly challenging is the example of self-driving cars, where such infrastructure would have to exist everywhere before such vehicles could operate in a consistent way.

Enabling billions (or trillions!) of low power edge devices means lowering their transmission power, so that they can be operated by battery and/or energy harvesting. New standards allow such lower power protocols, but that also implies a densification of the cells, in order to reliably communicate.

Finally, once we go past the RF interfaces, the requirements in the "digital" world also grow non-linearly, as managing the massive amount of channels and data pushes the boundaries of silicon and optics technologies. Interestingly, not only does peak power dramatically increase, but also "quiescent" power: as AI permeates all of these applications, and all new data is constantly integrated into models that help optimize each system's performance, all of this computational power steadily increases "in the background."

When considering all of the above in the context of exponentially scaling by many billions or even trillions of devices, the highly uneven balance between source and load becomes quite obvious. This is the genesis of the 5GEG. This concept brings focus to the extreme disparity in the commutation of power from source to load not only in the RF communication block(s) of the network, but also in the large overheard of the data processing/transfer/storage, and all direct and secondary energy utilization associated with it.

To again take a lesson from small scale systems that holds true as we grow from microscopic to continental scale, we need to account for both the static and dynamic costs and performance of energy delivery in systems. The modern, general purpose, industry standard, data center server and its CPU provide a lesson. As the transistors continued to scale down in size via Moore's Law, continuously lowering the cost per transistor, the cost and complexity of meeting the static and dynamic power requirements continued to increase. What was a single four rail output bulk power supply and a cable harness in the 1990s quickly became the vastly more complex distributed point of load conversion systems of today due to static and dynamic power increase simultaneous with the voltage decrease needed for the ever-shrinking transistors.

Eventually, though, that was not enough. We have reached the point where we are capped both by the ability to deliver power into all the transistors in a modern CPU and we are also capped by the ability to extract the waste heat out. As a result, active power management and charge rationing is required so we must derate the performance potential of CPU designs because we cannot afford to simultaneously power all of the devices we can fabricate in them and maintain performance and reliability. Further, as we increase the complexity of the now active and intelligent power management system we open up new attack surfaces to either sabotage or data extraction via side channel attacks, and of course, active management consumes power itself.

As we work to develop the tools to analyze the impact of all of the discontinuous and non-linear energy demand effects that will result from the technologies and use cases inherent in the 4G to 5G transition, the potential need to derate performance of a particular end-to-end 5G and backing IT infrastructure due to the 5GEG will become either a 5G Economic Gap or 5GEcG (the economic utility provided by 5G infrastructure will be attenuated because it is energy-limited) or it will become a 5G Equality Gap (5GEqG) because the new global standard won't be globally affordable. For jurisdictions where mobile operators are required to pay spectrum licenses upfront and have substantial penalties for not meeting roll-out timelines, a 5G Derate Factor (5GDF) may be required and could mean leaving channel capacity "on the table," greatly impacting payback period due to combined CAPEX and OPEX effects. The goal of INGR EE WG is to work towards meaningful analytic tools so that the fullest implications of energy on the diversity of potential 5G rollouts can be understood by industry and government.

In order to break this down into pragmatic subcategories, this whitepaper shall focus on energy in the following hierarchical context (though not all subcategories will be represented in detail in the scope of this document):

- Energy
    - Generation
        - Distribution
        - Stability
        - The Power Value Chain (PVC)
            - Power Cost Factor (PCF)
    - Utilization
        - Distribution Losses
        - Conversion Losses
        - System Utilization
        - Energy Storage
        - Networks
            - The 5G Energy Gap (5GEG)
    - Economics
        - The 5G Economic Gap (5GEcG)
        - The 5G Equality Gap (5GEqG)
        - The 5G Derate Factor (5GDF)
    - Mitigation
        - Efficiency Efforts
        - Eliminating Non-Rechargeable Energy Storage
    - Sustainability
        - Embodied Energy
        - Cradle-to-Grave Product Lifecycle

The current state of critical business challenges and the technologies associated with them are identified later in this whitepaper in the Roadmap Timeline Chart in Table 2, which builds upon the current state and projects how they will be addressed, adapted, and mitigated over a 10-year timeframe. By reviewing this chart one can quickly see how diverse the Energy Efficiency stakeholder landscape is and how easily the scope can expand beyond what is reasonable to address by any single group and/or effort.

3. Future State

This roadmap format is important for capturing the direction and opportunities for enhancing Energy Efficiency over a ten-year time period (through 2030) and breaking it down into reasonable expectations in the near- (3-year) and mid-term (5-year) time frames.

3.1. Network-Level Energy Analysis

In some respects (i.e., climate change), our efforts are on more of a scale measured in many decades or even centuries so we must keep all these different time scales in-mind, even as we make relatively near-term suggestions/decisions. Today, we are much closer to the Wild West than we are to a sustainable future with ubiquitous exploitation of intelligent energy management and end-to-end optimization for efficiency. Even if all non-renewable power plants would be replaced by clean renewable sources in a distant future, the energy consumption still dominates the operational cost.

Many large corporations today are making pledges to achieve carbon neutrality, which acknowledges fundamental changes in the way we perceive and utilize energy required for sustainability, but there is a far deeper penetration of EE concepts and execution required to drive truly systemic change on a global basis. As previously described, there are many owners and stakeholders engaged through all facets of the network and therefore cooperation and commitment is required at an unprecedented level to overcome the boundaries between the black boxes and achieve global optimization.

The Current State was outlined above as an awareness and technical risk factor (e.g., The 5GEG) to the 5G network currently being deployed. These technical, limiting risk factors were then outlined from a different perspective looking at these limitations from an economic purview (e.g., The 5GEcG) and even a socioeconomic awareness (e.g., The 5GEqG) to help harmonize these analyses from system/network architecting to cost benefit and payback period. In reality, none of these concepts (even with associated metrics) will yield an ideal model of an ACTUAL deployment. Therefore, it is proposed they be considered as more of a characterization of the discrepancy between a model and reality, which results in the need for a derating factor (e.g., The 5GDF) to set more realistic expectations of energy utilization and OPEX as we move into the future.

3.2. Key Metrics

Cellular networks are an interesting case study for roadmaps because they operate on multiple time scales concurrently. While it can typically take the better part of a decade to deploy a next-generation wireless technology, there are incremental opportunities in the months/years timeframe to drive improvement at the component-, device-, or system-levels. In the future, efficiency of protocols and firmware might even be continuously improved using machine learning.

In order to understand the overall impact of implementing such changes at the infrastructure level, it may be useful at this time to introduce some key metrics capable of describing the end-to-end energy impact of different network components.

We therefore introduce here the following metrics:

- **Power Cost Factor (PCF)** is a unitless number that represents the multiplication factor required to quantitatively assess the overall cost of energy utilization at any given point within the PVC.

- **5G Derate Factor (5GDF)** is a unitless coefficient (<1) representing a scaling factor for the application of technical and economic risk factors to the ideal 5G network deployment model that will reduce the optimal, maximum designed capabilities of a network due to energy-limited (5GEG) and/or economically-limited (5GEcG) and/or socioeconomically-limited (5GEqG) factors.

A graphic representation of this concept is provided in the example shown in Figure 4. The PVC component blocks in the top row represent constituents that are energy-limited (i.e., limited by energy production, but easily scalable). The blocks in the bottom row represent constituents that are power-limited (i.e., limited by engineered power envelope, thus driving the need to constrain the application or load when pushing the limits of the envelope).

The 5GEG will have more impact on the portion of the PVC closer to the core, while the 5GEcG and 5GEqG will have more impact on the portion closer to the edge. Whether some or all of these factors apply, they shall dictate the value of the 5GDF necessary to accurately characterize and model a real-world deployment and place bets on its payback period and societal impact.

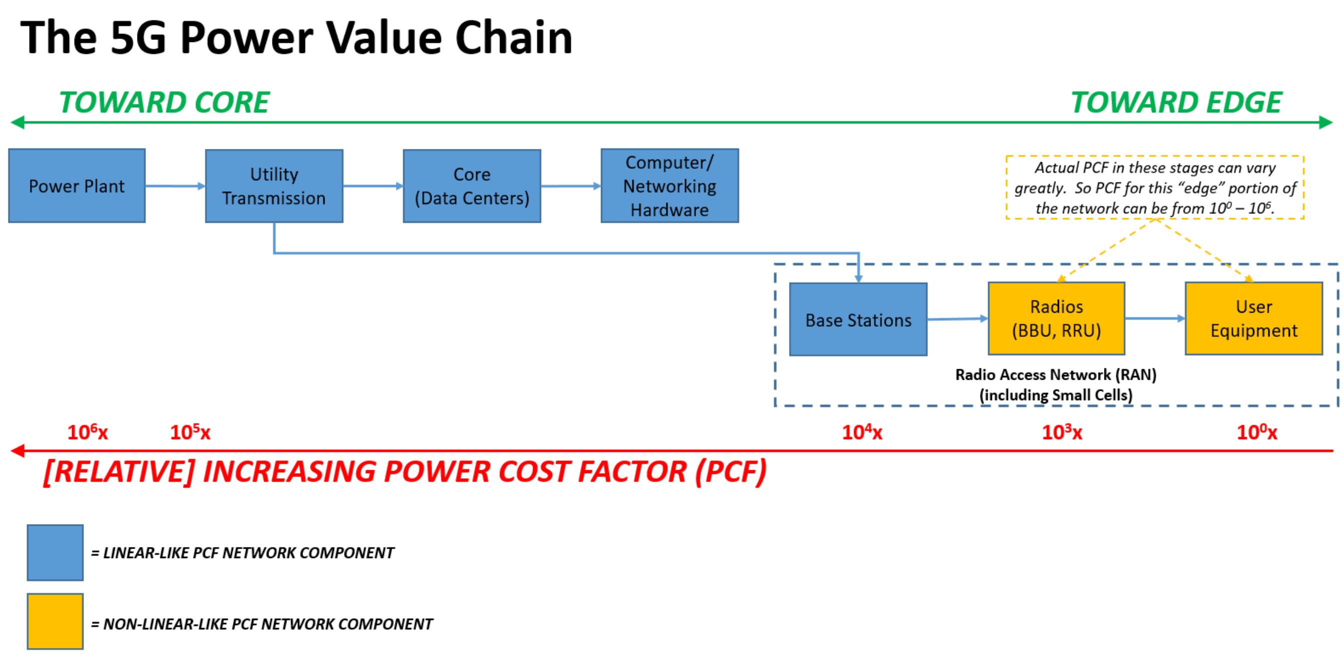

Figure 4 – The Power Value Chain (PVC) from Network Edge to Power Plant, courtesy of PowerRox

Given how quickly power scales with the number of systems and devices involved in end-to-end network deployment, it is not always obvious which devices present the greatest opportunity for efficiency improvement. For instance, one pays a much larger PCF as one operates downstream of the PVC. Hence, while a base station may carry a PCF closer to 10^1, a smartphone using that same base station may carry a PCF of around 10^5!

The PCF metric enables both technical and economic analyses by applying from the micro-level (i.e., single system development) to the macro-level (i.e., network-wide utilization, global energy optimization, etc.). More importantly, this simple metric quickly and easily translates into energy consumption and therefore energy costs, which allows it to transcend the lines between technologically- and economically-based analyses and harmonize what would otherwise be very difficult (perhaps even conflicting) attempts at modeling, characterizing, and predicting energy utilization.

To be clear, the proposed introduction of the PCF metric is not intended to be the cure-all solution for the many challenges identified in this whitepaper. PCF and 5GDF are merely a starting point that breaks down the barriers between network constituents and provides an oversimplified way of leveling the playing field between [what has traditionally been] complicated and heterogeneous analyses, which stood in the way of harmonizing these different analyses into a consistent, network-level model (or ideally PVC-level).

A key objective of this WG is to simplify some fairly complicated analyses represented by multidisciplinary stakeholders and enable them to all come together and speak a common language in the interest of enhanced, overall energy efficiency. It is understood that business and economic drivers (more than altruism or global sustainability) dominate the discussion so it is also important to propose solutions that allow entities to be differentiating and economically viable (from the standpoint of their specific place in the PVC), while concurrently energy-efficient and [ideally] compatible socioeconomically as well. Ultimately, it is expected these metrics and concepts will mutate into something entirely different or even lead to a series of metrics that are much more accurate and applicable to real-world modeling.

3.3. Data Processing Architecture

Following the trend of how one pays a higher cost for energy the further to the edge it is consumed, one must also recognize the cost and infrastructure benefits achieved by keeping as much processing as close to the point of raw data collection (i.e., the edge) as possible in order to minimize energy consumption. This philosophy has driven the emergence of technology focus areas such as Mobile Edge Computing (MEC) and Edge Buffering, which utilize various techniques to minimize the energy associated with the transport of large chunks of data by keeping as much (stored and/or raw data analytics) as possible processed closer to the point of origin. There are cases where it might be worth spending energy to move raw data away from the edge, in an effort to reduce the energy consumption elsewhere in the system. For example, joint interference rejection between base stations can be implemented to enable reduced transmit power over the radio links. It is nevertheless important to recall a key philosophy of optimal design architecture, which is "Just because you can does not mean you should."

The architecture adopted in the management of data collection, pre-processing, storage, and transmission is one of the fundamental areas impacting both local and global energy consumption. Each application has a different set of requirements and constraints: how much energy is available in the local battery to perform the task; what is the response time required; what are the available transmission capabilities; what is the local computing power, etc. In the past, the decisions have been mostly driven by hardware cost and battery power constraints, while in the future total energy cost will have a more substantial impact on the trade-off decisions, due to complexity and scale.

As an example, we can look at the "data management" of autonomous driving vehicles to gain a perspective on the architectural decisions that are required to achieve a viable solution. Autonomous driving represents a drastic departure from prior "dashboard" applications (e.g., entertainment, GPS, etc.), as it entails orders-of-magnitude increase in computing complexity to process all the inputs in real time and ensure the safe operation of the vehicle in all conditions.

As the mileage achieved with a battery charge depends in significant part on the energy utilized in the vehicle, a self-driving vehicle is therefore an excellent case study for analyzing the trade-offs between pre-processing, storing and transmitting data in order to achieve a viable total energy consumption. With exponentially rising data analytics capture and processing comes an exponential demand on the energy needed to achieve this at the edge and therefore has a growing and direct link and impact to EE at the overall network-level.

The job at hand is therefore that of reducing as much as possible the datasets and as early in the pipeline as possible, so that the problem is manageable according to both timing and energy (and cost!) budgets. We therefore accept significant inaccuracies, and “hybridize” the system (localized vs. centralized) to reach the best possible accuracy given the constraints.

Approaching energy utilization must be done at all levels and is difficult to relegate to purely a top-down or bottom-up approach. A key goal of this whitepaper is to raise awareness of the major motivations, drivers, and high-impact areas related to global energy efficiency, particularly within communication networks and everything they touch, which represents just about every aspect of modern life as we know it. With awareness comes a focus on optimizing energy utilization at all levels and utilizing a common language to ease and harmonize both technical and financial analyses of everything from system design to network/grid architecture.

From this brief whitepaper will come a full chapter in the 2nd Edition INGR as described above. This shall be the capstone deliverable for this WG in which the topic of EE receives appropriate attention for capturing critical needs and dependencies within the full framework of the roadmap.

4. Requirements and Technology Gaps

Here we shall summarize what is needed to achieve the future state(s) the working group will address in the IEEE roadmap effort. The focus is placed on the most critical items that will be of interest and can benefit from a focused effort to solve.

4.1. Roadmap Timeline Chart

Table 2 – Working Group Needs, Challenges, and Enablers and Potential Solutions

Name	*Current State* *(2020)*	*3 years* *(2023)*	*5 years* *(2025)*	*Future State* *10-years (2030)*
Need #1	#x Network Efficiency Improvement	##x Network Efficiency Improvement	###x Network Efficiency Improvement	####x Network Efficiency Improvement
Challenge(s) for Need 1	The 5G Energy Gap (5GEG)	The 5G Energy Gap (5GEG)	- Densification - mmWave, mMIMO Deployments	Ubiquitous HetNets of Small Cells
Possible Solution for Challenge	- Edge Buffering - Awareness & Education - Optimizing System Design for Power & Energy Utilization - 3GPP DTx Features	- Energy Harvesting (device-level) - Mobile Edge Computing (MEC) - Optimizing RAN for Power & Energy Utilization - Migration of Data Center Efficiencies	- Energy Harvesting (base station-level) - Standardized “5G Small Cell” with Interference Coordination - Greatly Improved PAE (perhaps	- Energy Harvesting (data center-level) - Radio Stripes

		from HPC/ Exascale to Enterprise Applications	requiring use of WBG devices) - Continue Migration of Data Center Efficiencies Toward Edge	
Need #2	Dynamic/Adaptive Base Stations	Dynamic/Adaptive Small Cells	Disaggregated Centralized Network	Cell-free Architectures
Challenge(s) for Need 2	- Complicated Control Plane - Departure from Norms	- Increased Inter-cell Interference - Need for Scalable Interference Management	- Major Network Architecture Paradigm Shift	
Possible Solution for Challenge	- Adjust Power for Real-time Traffic/Energy Demands - mMIMO with Spatial Interference Rejection	- Small Cell Situational Awareness - Coordinated Multipoint Methods	- AI-facilitated Dynamic Spectrum Allocation - User Centric No Cell (UCNC)	- Fully AI-driven, Real-time, Optimal Spectrum Utilization - Cell-free mMIMO Networks to Alleviate Interference
Need #3	Reduced Power Dissipation in Base Station Radios	Reduced Power Dissipation in UE Radios	Network-wide Energy Awareness	Regional/Global Energy Awareness
Challenge(s) for Need 3	- Requirements on Out-of-band Distortion Must Be Satisfied	- Requirements on Out-of-band Distortion Must Be Satisfied - Support for Many RF Bands	- Power/Energy Telemetry Data Acquisition - Power/Energy Data Analytics - Defining Energy-optimal Control Feedback Loop(s)	- Enabling/Deploying Energy-optimal Control Feedback Loop(s)
Possible Solution for Challenge	- Use mMIMO radios with many low-gain antennas with handset-grade hardware instead of few high-gain antennas.	- Dedicated Circuit Design With Reduced Distortion Margins - Cross-band Optimization For Energy Efficiency		
Need #4	Enhanced Economic Modeling	Network-Component-Based Economic Models	End-to-end Network Economic Models	Global Economic Models
Challenge(s) for Need 4	- Defining the 5G Economic Gap (5GEcG) - Defining the 5G Derate Factor (5GDF) - Strict Payback Period Targets Driving Socioeconomic Disparity	Characterizing individual components (specific yet compatible)	- Cooperation Between Network Stakeholders - Model Complexity - Model Validation	- Cooperation Between Global Stakeholders - Model Complexity - Model Validation - Socioeconomic Considerations
Possible Solution for Challenge	Applying the 5G Economic Gap Analysis (5GEcG)	- Network Component Energy Utilization Metric(s)	- Disaggregated Network Metric(s)	Demonstration of ability to optimize energy utilization from

	Applying the 5G Derate Factor (5GDF) Considering the 5G Equality Gap (5GEqG)	- Energy Efficiency Metric Standardization	- Demonstration of Energy & TCO Savings	micro to macro levels (bidirectionally) Validated Model Database

5. CONTRIBUTORS

Brian Zahnstecher, PowerRox (Co-chair)

Francesco Carobolante, IoTissimo (Co-chair)

Earl McCune, Eridan Communications

Magnus Olsson, Huawei

Rajesh Uppal, self

Anthony Magnan, Verizon

Kirk Bresniker, HPE

Emil Björnson, Linköping University

Anirban Bandyopadhyay, Global Foundries

David Su, Atmosic

Bruce Nordman, Lawrence Berkeley National Laboratory

Steve Allen, pSemi/Murata

Apurv Mathur, Nokia Bell Labs

Tom Lambalot, NewEdge Signal Solutions

6. REFERENCES

[1] Wikipedia contributors. "Power usage effectiveness." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 4 Mar. 2020. Web.

[2] B. Zahnstecher, "The 5G Energy Gap [Expert View]," in IEEE Power Electronics Magazine, vol. 6, no. 4, pp. 64-67, Dec. 2019.

[3] Masanet, E., Shehabi, A., Lei, N., Smith, S., and Koomey, J. 2020. "Recalibrating global data center energy-use estimates" In Science Magazine. February 28. https://science.sciencemag.org/content/367/6481/984.abstract

IEA (2019), "Tracking Buildings", IEA, Paris https://www.iea.org/reports/tracking-buildings

7. ACRONYMS/ABBREVIATIONS

Term	Definition
1G-4G	First Generation to Fourth Generation
3GPP	Third Generation Partnership Project
5G	Fifth Generation
5GDF	The 5G Derate Factor
5GEG	The 5G Energy Gap
5GEcG	The 5G Economic Gap
5GEqG	The 5G Equality Gap
ACK/NAK	Acknowledgment/negative acknowledgment
AI	Artificial intelligence
API	Application programming interface
B2B	Business to business
B2C	Business to consumer
BS	Base station
CAPEX	Capital expenditure
COTS	Commercial off-the-shelf
D2D	Device to device
DL	Downlink
DPA	Doherty Power Amplifier
DTx	Discontinuous transmission
EE	Energy Efficiency
eMBB	Enhanced mobile broadband
ETSI	European Telecommunications Standards Institute
FN	Future Networks
FoM	Figure-of-Merit
GHz	Gigahertz
GSMA	GSM (Groupe Speciale Mobile) Association
HIR	Heterogeneous Integration Roadmap
HPC	High Performance Computing
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
INGR	International Network Generations Roadmap
IP	Internet protocol
ISP	Internet service provider
KPI	Key performance indicator
LPA	Linear Power Amplifier
LTE	Long-term evolution
M2M	Machine to machine
MEC	Mobile Edge Computing
MIMO	Multiple input, multiple output
mMIMO	Massive MIMO
ML	Machine learning
mMTC	Massive machine-type communication
mmWave	Millimeter wave
MVNO	Mobile virtual network operators
NFV	Network function virtualization
NR	New radio
OPEX	Operational expenditure

20 Acronyms/abbreviations

PA	Power Amplifier
PAE	Power Amplifier Efficiency
PCF	Power Cost Factor
PHY	Physical layer
PoC	Proof-of-Concept
PVC	Power Value Chain
QoS	Quality of service
RAN	Radio access network
RSRP	Reference signal received power
SDN	Software defined network
UAV	Unmanned aerial vehicles
UCNC	User Centric No Cell
UE	User equipment
UL	Uplink
V2I	Vehicle to infrastructure
V2V	Vehicle to vehicle
WBG	Wide Bandgap
WG	Working group
WRC	World Radiocommunication Conferences

ANTI-TRUST STATEMENT

Generally speaking, most of the world prohibits agreements and certain other activities that unreasonably restrain trade. The IEEE Future Networks Initiative follows the Anti-trust and Competition policy set forth by the IEEE-SA. That policy can be found at https://standards.ieee.org/develop/policies/antitrust.pdf.

This page is intentionally left blank.

FutureNetworks.ieee.org/roadmap

IEEE INTERNATIONAL NETWORK GENERATIONS ROADMAP, 1ST EDITION