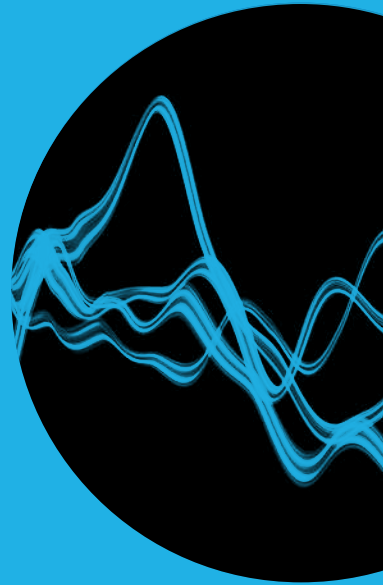


# Review

ERICSSON  
TECHNOLOGY



IMPROVING ENERGY  
PERFORMANCE  
IN 5G NETWORKS



ERICSSON

# Improving energy performance in 5G networks and beyond

Continued focus on energy performance in 5G and 6G development will be essential to enable new deployment scenarios with smaller and lighter telecom equipment, as well as minimizing the climate impact of mobile networks.

---

CECILIA ANDERSSON,  
JONAS BENGTTSSON,  
GREGER BYSTRÖM,  
PÅL FRENGER,  
YLVA JADING,  
MY NORDENSTRÖM

---

**As the 5G rollout continues and we look ahead to 6G, the exponential growth in processing is going to be one of the biggest challenges from an energy performance perspective.**

■ The telecom industry has a long history of prioritizing peak performance and high capacity. Until recently low energy consumption was typically perceived as a nice-to-have extra benefit rather than a crucial feature of state-of-the-art mobile network equipment. As awareness about the need to optimize energy performance in telecom networks continues to grow, several challenges must be addressed.

In 5G networks, digital processing in base stations can increase more than 300 times compared with early Long Term Evolution (LTE) products, primarily due to an increasing number of antenna branches, broader bandwidths and shorter transmission time intervals (TTIs). This increase is expected to become even larger in 6G. To handle this increase responsibly from an energy consumption perspective, it is essential that the future 6G standard is as lean as possible. Most importantly, the amount of mandatory and always-on signaling must be kept to a minimum, which will enable underlying components and subsystems to provide sufficiently high levels of load dependence in their energy consumption.

The percentage of energy consumption originating from compute and digital silicon will continue to rise as the New Radio (NR) rollout continues. The trends of increased bandwidth, more antenna ports and shorter TTIs are driving an exponential increase in processing needs on the digital frontend, which propagates into the radio unit, beamforming and layer one processing. Processing needs for layer two, packet processing and control functions are also increasing, but not as quickly.

### What is energy performance?

In existing mobile systems, where energy consumption has low dependence on the load and traffic growth is high, an energy-efficiency metric is not particularly useful. Even if no effort is made to reduce energy consumption, traffic growth will cause energy efficiency to improve from one year to the next.

The term energy performance broadens the focus beyond energy per bit to total energy consumption, highlighting the similarities between achieving high system performance and low energy consumption. Optimizing energy performance means minimizing the energy consumption for a set of performance requirements (user throughput, capacity, latency and so on). The concept of energy performance makes the trade-off between energy consumption and performance requirements transparent.

Energy performance is relevant in three separate contexts – economy, ecology and engineering – that have different stakeholders and use different terminology. In the economic context, optimizing

## THE TERM ENERGY PERFORMANCE BROADENS THE FOCUS BEYOND ENERGY PER BIT TO TOTAL ENERGY CONSUMPTION

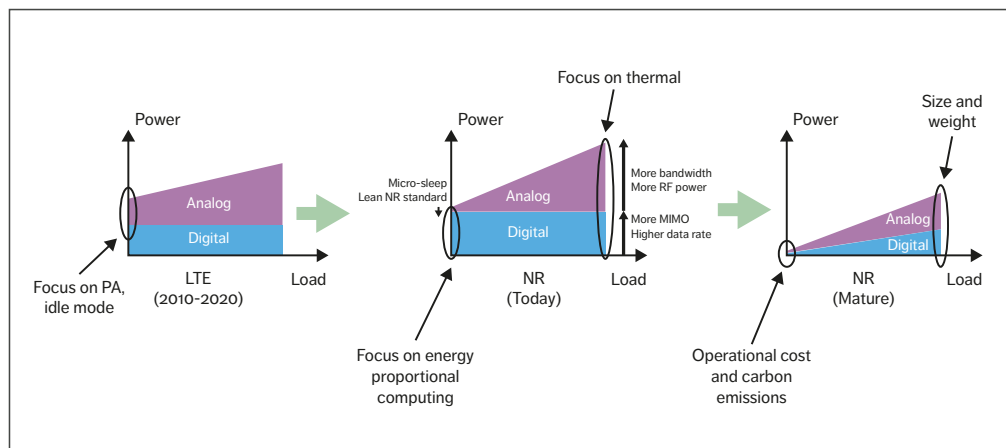
energy performance results not only in lower operating expenses for mobile network operators, but also in the potential for lower capital expenses due to the use of smaller and lighter equipment. This equipment enables new and simplified deployments as well as less-costly power supply and energy storage solutions. Low energy consumption is therefore of particular importance when discussing the costs for enhanced availability and enhanced national security.

In the ecological context, the value of optimizing energy performance is best demonstrated by the fact that about 94 percent of Ericsson's total carbon emissions impact is associated with the operation of our products [1].

The engineering context is arguably the most underestimated and underutilized aspect of energy performance. A sharp focus on reduced energy consumption in the product engineering phase will yield desirable reductions in product size and weight. Tighter energy performance product requirements drive technical, managerial and organizational innovation in a greater change process that leads to closer collaboration over multiple areas of technology.

### Terms and abbreviations

**AI** – Artificial Intelligence | **C-RAN** – Centralized RAN | **CSI** – Channel-State Information | **D-RAN** – Distributed RAN | **KPI** – Key Performance Indicator | **LTE** – Long Term Evolution | **MB** – Megabyte | **Mbps** – Megabits Per Second | **MHz** – Megahertz | **MIMO** – Multiple-Input, Multiple-Output | **ms** – Millisecond | **NR** – New Radio | **PA** – Power Amplifier | **PRB** – Physical Resource Block | **RAN** – Radio Access Network | **RAT** – Radio Access Technology | **RF** – Radio Frequency | **Tbit** – Terabit | **TTI** – Transmission Time Interval | **UE** – User Equipment | **W** – Watt



**Figure 1** The energy performance journey of mobile networks

Depending on the context, the impact of energy usage can be expressed in terms of money, carbon emissions, thermal design constraints and so on. But regardless of how the benefit is viewed, the solutions required to reduce energy consumption are very similar. It is important, however, to differentiate between targets whose purpose is to reduce average energy use (mostly relevant for operational costs, carbon emissions and the like) and those aimed at reducing peak energy use (mostly relevant for product dimensioning, thermal design, size and weight). Fortunately, though, effective solutions for energy reduction often reduce peak and average energy use simultaneously.

### The network energy performance journey

**Figure 1** illustrates the energy performance journey of mobile networks. The graph on the left shows the power consumption of a typical LTE base station in the period 2010-2020. The graph in the middle shows the power consumption of a typical New Radio (NR) base station today. The graph on the right shows the projected power consumption of a mature NR base station beyond 2025.

In LTE, the energy consumption of the radio access network (RAN) was dominated by base

stations that comprised around 80 percent of the RAN electricity use. Furthermore, within each base station, around 80 percent of the energy consumption was used in the power amplifiers (PAs) [2]. During this time the main energy performance goal was to reduce the idle-mode consumption of the PAs.

Network traffic was much lower in the early days of LTE. In a countrywide network, only 5 percent of the physical resource blocks (PRBs) were typically used for data traffic, while around 95 percent of the PRBs were empty. The high static energy consumption and low average traffic resulted in very low load dependency in the network, where the additional energy increase due to the traffic in the network was well below 2 percent of the total energy used [3].

In order to improve the load dependency, we introduced micro-sleep transmission, an energy-saving feature that deactivates and reactivates PAs in microseconds. Micro-sleep transmission is effective at all times when there are no transmissions from the base station. However, it could be even more effective if the LTE standard was not filled with mandatory and always-on reference signals that cannot be turned off, even when there is no traffic. The obvious solution to this was to redesign

## OUR RESEARCH SHOWS THAT IT IS POSSIBLE TO MOVE FROM TODAY'S SITUATION TOWARD A MORE ENERGY-LEAN TOMORROW

the physical layer so that only a minimum of signaling would be necessary when there is no data to transmit, but this required a new standard.

When 5G NR was developed, we ensured that its physical layer had an ultra-lean design [4]. This design makes features such as micro-sleep transmission much more efficient, and the effect on live networks is already significant [5]. Micro-sleep transmission has had a major impact over the course of the past decade by dramatically decreasing the energy consumption in the analog radio parts of base stations.

5G NR is also designed for massive MIMO (multiple-input multiple-output), and it supports wider bandwidths than 4G LTE. A typical LTE base station has two transmit and receive branches, 20MHz of spectrum, and the digital processing time in the base station is 1ms (corresponding to one TTI). Early NR products have 64 antenna branches, support 100MHz of spectrum and have a TTI of 0.5ms. This implies that digital processing in radio base stations needs to increase 320 times compared with early LTE products.

In today's NR products, the energy that digital components use can be as large, or larger, than the power used by the analog components (mainly PAs). The additional bandwidth and additional antennas were also introduced while keeping the radio frequency (RF) power spectral density constant at around 2W/MHz. While a typical 20MHz LTE base station could deliver 40W RF output power, a 100MHz NR base station can be capable of delivering 320W RF power.

More bandwidth and more RF power increased the peak power usage of analog components, while more antenna branches and more digital processing

increased the power consumption of digital components (see the middle part of Figure 1). This massive upscaling of capabilities has resulted in an increasing energy consumption trend in the industry. Fortunately, it is possible to alter this trend with new building practices. Our research shows that it is possible to move from today's situation (the middle part of Figure 1) toward a more energy-lean tomorrow (the right part of Figure 1) by taking the following actions:

1. Improve the load dependence of components (especially digital ones) by significantly improving sleep modes in terms of idle-mode energy usage, reactivation times and granularity. One example of a technique that is currently underutilized in many digital designs is dynamic voltage and frequency scaling (DVFS).
2. Relax the requirements for analog components dynamically (when possible).
3. Implement more power management functions in the network nodes (requires more load-dependent components to be effective).
4. Use artificial intelligence (AI) techniques to achieve network-level optimizations (requires more load-dependent network nodes to be effective).
5. Consider removing support for legacy technologies such as WCDMA (Wideband Code Division Multiple Access) and GSM (Global System for Mobile Communications).

This last point is particularly important, as supporting legacy technologies results in requirement creep and consumes development and testing resources, thereby reducing the resources available for the implementation of power-saving functions.

Successful execution of these actions would not only optimize energy performance; it would also enable operators to reduce total network energy consumption by adding additional capacity as traffic increases, as more capacity results in more idle mode operation [6].

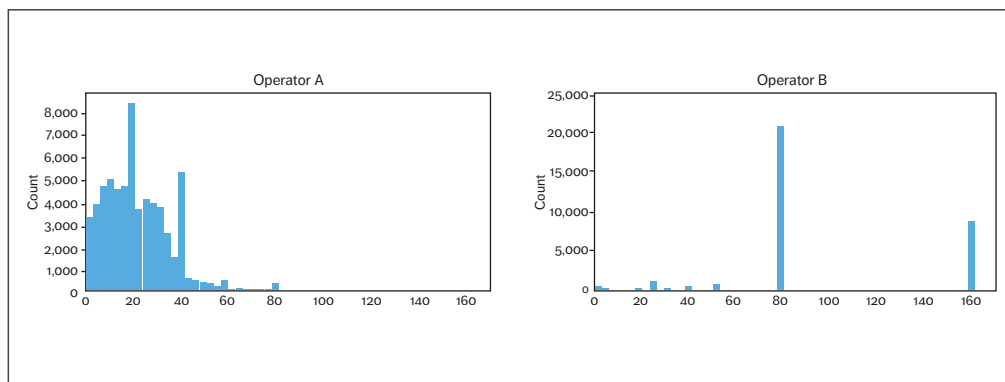


Figure 2 Histograms of configured maximum transmission power for two European operators

### Insights from field data

While some operators run their networks with the power settings of every base station fixed at the maximum level, others have more optimized and diverse configurations. Figure 2 provides histograms of configured maximum transmission power for two European operators. Operator A, on the left, optimized the power levels in its network, while Operator B, on the right, used default maximum power settings.

One study on this topic [7] indicates that the total network energy usage can be reduced by around 10 percent without compromising performance simply by tuning the output-power levels, as shown in Figure 2. There are similar differences in how existing energy-saving features such as MIMO sleep mode, booster carrier sleep, cell deep sleep and the low energy scheduling solution are activated (or not) in different networks. This indicates a clear need for improved tools to optimize the energy performance in existing networks.

### Meaningful reference points

Some energy-saving solutions are associated with a negative key performance indicator (KPI) impact. However, we would argue that this often depends on the reference KPI chosen to compare with, or perhaps even more importantly, what target KPI is expected to be fulfilled.

Figure 3 illustrates the total downlink traffic volume (purple) and the mean user throughput (blue) in a real network for one day. Due to variations in traffic volume, the user throughput differs by about a factor of two between the peak traffic hour and the minimum traffic hours. The question is, when evaluating the throughput obtained with an energy-saving feature turned on, should we relate that to the performance during peak hour or during the minimum traffic hour? The large KPI variations that we normally observe in the network are not always desirable, and an energy-saving feature might even be an attractive way to both obtain energy and cost savings, as well as more predictable throughputs in a network throughout the day.

It is not necessary to set the desirable KPI for an energy-saving feature to the minimum KPI obtained at peak hour. It can also be set to other targeted KPIs for a specific network, such as a KPI for which the network has been dimensioned. The network dimensioning KPI target can be regional and/or service dependent. In some areas such as city centers or indoor factories, the requirements may need to be higher, while in other areas they can be more relaxed. In the end, this should be an operator policy decision. One could, for example, allow a fraction of the performance above the minimum performance requirement to be used to reduce energy consumption.

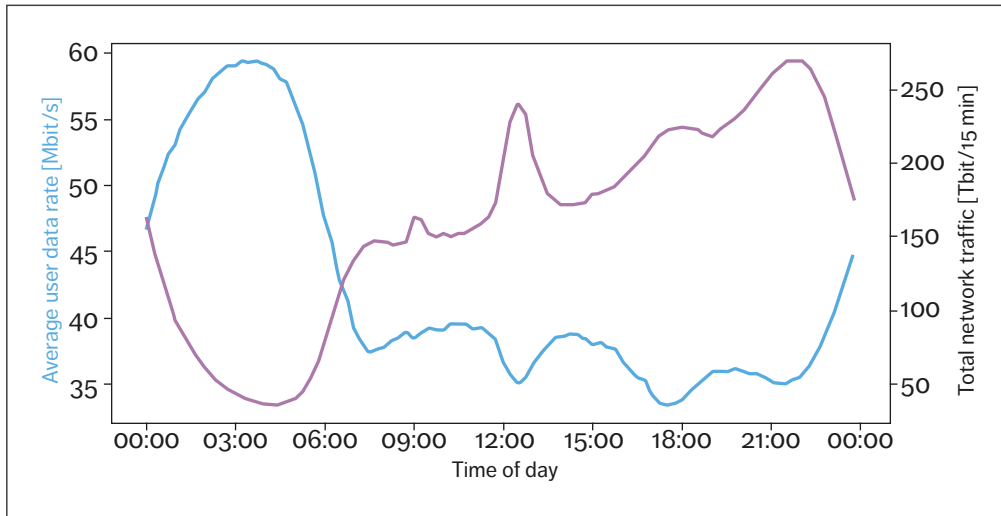


Figure 3 Variations in traffic volume (purple) and user throughput (blue) over a day in a real network

### Centralized versus distributed radio access networks

It is often argued that a centralized RAN (C-RAN) is inherently better than a distributed RAN (D-RAN) deployment for enabling low-network power-consumption solutions. As mobile networks are dimensioned for peak traffic, resources are always overprovisioned at some places in the network. As peak traffic occurs at different times of day in different locations of a network, a C-RAN can exploit pooling gains when processing resources can be shared in a common pool. In addition to the processing-pooling gain, C-RAN deployments can also utilize more efficient cooling, power-supply and energy-storage solutions. Based on this, considerable gains have been reported [8].

It is important to understand which parts of a network that can realistically be centralized in a C-RAN deployment and under what circumstances. C-RAN deployments are typically assumed to operate on off-the-shelf hardware, which opens up the possibility to add value with custom-made silicon that can deliver higher load dependence and lower fixed energy costs. Increased latency is a

consideration that will limit the centralization of the more time-critical and power-consuming lower-layer processing, such as the digital frontend. The analog (radio) parts must also remain distributed in a C-RAN deployment, as the users will still be distributed.

### Traffic: RAN for mobile broadband and beyond

There are two main approaches to reducing energy usage in RAN: rush to sleep, and rate adaptation. The rush-to-sleep approach aims to transmit the data as fast as possible in order to maximize the time in sleep mode. Rate adaptation instead aims to adapt the transmission rate to instantaneous requirements and thereby enable energy savings by under-clocking or deactivating some components while transmitting data.

When looking at traffic statistics in real networks, it is evident that both of these approaches are needed. About 95 percent of all data sessions are small (less than 1MB), and the one percent largest sessions contribute to almost three quarters of the total data volume. For small data sessions, rate adaptation is an effective way of reducing the energy

needs, as large bandwidth and many antenna ports are not required for most sessions. In contrast, the best way to handle large sessions from an energy performance perspective is by combining high data rates with effective mechanisms for the equipment to rush to sleep once the transmission is finalized.

Video traffic is currently estimated to account for 69 percent of all mobile data traffic, a share that is forecast to increase to 79 percent in 2027 [9]. To deliver a satisfactory user experience, a network must be able to play video content on demand, without delays or stalling. Video streaming uses buffering to smooth out throughput variability. As a result, video is a service that can easily be made compatible with power-saving features in the RAN.

The ability to avoid stalling requires rate adaptation on a relatively slow timescale (seconds). Ensuring a sufficiently short time-to-play for video services and time-to-content for web services requires an almost instantaneously available bitrate in the order of 20Mbps [10]. As acceptable time-to-content and time-to-play numbers can be in the order of one to four seconds, “almost instantaneously” needs to be significantly smaller than this (100ms, for example).

For web and video services, it is therefore sufficient for a base station to provide about 20Mbps to any user within 100ms if additional capacity to accommodate the new service can be made available within one second. This would result in around 1.5 seconds time-to-content, which is classified as excellent in the 2025 scenario in a recent Ericsson Mobility Report [9].

Critical machine-type communication and other low-latency services can also be made compatible with RAN-power saving features if the low-latency requirements are managed with care. Low-latency service requirements must be broken down into two additional categories: service-registration latency and protocol ramp-up latency. The activation of a low-latency service requires a quality of service guarantee from the RAN, which includes a setup procedure that can be allowed to take 100ms or more.

In addition, low-latency requirements need to be

contained regionally and per band. It is not necessary to provide countrywide support for ultra-low latency to enable factory automation in very limited areas. All the bands, cells and nodes that support low-latency services are not needed all the time. Even the most critical low-latency service does not require a network that is constantly in a high-alert state that prohibits the use of energy-saving rate adaptation or capacity adaptation in the RAN.

In short, it is possible to support all services today and in the future (as far as we know) while deactivating close to 100 percent of the excess capabilities in a base station if the following two conditions are met:

1. At least 20Mbps can be provided to an additional user within 100ms.
2. Any additional required capacity and rate can be made available within one second.

Any active but unutilized spare capacity beyond this is a waste of energy.

### Standardization and 6G

System design by standardization is the foundation that enables energy-efficient design of an entire network. But having a good standard is not enough – components and products must utilize the potential that the standard provides. Network management that includes the application of AI tools for load balancing and energy consumption minimization is also essential. To be effective, network management requires a good standard that enables low-energy and load-dependent operation, as well as products that utilize the energy savings that the standard enables.

NR is a lean standard that is already a powerful enabler of low network-energy usage. The most important thing from an energy performance perspective moving toward 6G is to maintain and extend the lean properties on which NR is based, such as enabling up to 160ms of transmission-free periods. For 6G the concept of lean design should be extended to better support network densification and even larger massive MIMO antenna arrays



without requiring excessive spatial repetition and beam sweeping of idle-mode signals regarding synchronization, system information and paging.

Today's NR specifications provide signaling support for idle-mode user equipment (UE) to see the full set of beams, bands and nodes that are configured and that can be made available in active mode. The need for this transparency in standards and products is questionable, given the energy cost of associated transmissions and shorter transmission-free periods, as well as reduced sleep possibilities.

In the discussions about next-generation technologies, there is a tendency to allow requirements from one area to propagate into other areas, or even into all areas. The support for extreme capabilities – such as extremely high data rates with corresponding extreme transmission bandwidths, extremely low and predictable latency and extreme reliability – is important to enable the wide range of use cases envisioned for 6G.

At the same time, such capabilities come with a cost in terms of network energy consumption. It is crucial that this cost is limited to the situations where and when the specific capabilities are required. One way of doing this is to prevent requirements for active mode from also applying to idle mode. We would argue that it is worth considering a stricter separation between active and idle mode in future networks.

In addition, efforts should be made to design functionality so that it is self-contained, refraining from the reuse of signals specified for one functionality to support other. This may sound counterintuitive, but experience shows that the associated dependencies between different functionalities often prevent desirable sleep-mode possibilities. A prime example of this is the cell-specific reference signals in LTE that are used for both active-mode demodulation of data and for idle mode UE cell search and mobility. This reuse results in a very high cost for transmitting signals in idle mode.

One of the improvements we would like to see in 6G is the ability to avoid the overhead cost of

obtaining channel-state information (CSI) when there is no data to transmit. A more preamble-based design of the physical radio links, where synchronization signals and reference signals for CSI acquisition are transmitted together with the data bursts, would make the cost of all supporting signals (synchronization, CSI acquisition and so on) explicit. More opportunistic scheduling of such supporting signals would become more natural as a result.

The availability of more shared and pooled infrastructure can contribute to further lowering total network energy consumption. Examples of this include multi-radio-access technology (RAT), multi-operator and/or multi-band operation. Although there already are several standardized solutions for operations in such scenarios, further enhancements are worth investigating.

With regard to the role of AI in energy performance improvements, the vast majority of AI functionality is expected to be related to implementation rather than specification. However, standardized AI-supporting functionality related to observability and control that targets energy optimization at network level would be helpful.

As the work on 6G progresses, it is important to be mindful of the risks associated with introducing new functionality into later releases of technology specifications. The potential of lean design can easily be compromised both in standards and implementation, depending on how new functionality and the associated signaling is added.

To summarize, 5G is already a very good standard that enables the implementation of low energy-consuming behavior in mobile systems. Lean design can be further optimized and enhanced in 6G by including additional domains (frequency bands, beams, nodes, RATs, slices and so on).

## Conclusion

Lean design is the foundation for improving radio access network energy performance. The lean design of New Radio (NR) standard was a major improvement compared with Long Term Evolution (LTE), enabling unprecedentedly low energy consumption in live 5G networks. It is of the utmost

importance that this progress is not compromised in future standardization and products.

Looking ahead, the main energy performance challenge will be scaling processing with traffic to meet the digital processing needs of high-performing networks with larger antenna arrays and bandwidths,

and shorter processing requirements. In the 6G timeframe, we are advocates for a further reduction in fixed idle-mode energy usage and peak power requirements – changes that will enable the use of smaller, lighter products and support novel deployment solutions in 6G networks.

## References

1. Ericsson, **Sustainability and Corporate Responsibility Report, 2021**, available at: <https://www.ericsson.com/en/about-us/sustainability-and-corporate-responsibility/sustainability-report>
2. EARTH Deliverable 2.3, **Energy efficiency analysis of the reference systems, areas of improvements and target breakdown**, December 2010, available at: <https://cordis.europa.eu/docs/projects/cnect/3/247733/080/deliverables/001-EARTHWP2D23v2.pdf>
3. IEEE Xplore, **2014 IEEE 79th Vehicular Technology Conference (VTC Spring), Assessment of Alternatives for Reducing Energy Consumption in Multi-RAT Scenarios**, 2015, Frenger, P; Ericson, M, available at: <https://ieeexplore.ieee.org/document/7022838>
4. Ericsson, **IMT-2020 self-evaluation: Radio Network Energy Performance**, 3GPP TSG-RAN WG1 #91, R1-1720954, Reno, US, November 27-December 1, 2017, available at: [https://www.3gpp.org/ftp/tsg\\_ran/WG1\\_RL1/TSGR1\\_91/Docs/R1-1720954.zip](https://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_91/Docs/R1-1720954.zip)
5. Ericsson blog, **5G energy consumption: what's the impact of 5G NR in real networks?**, October 8, 2021, Frenger, P; Jading, Y; Bengtsson, J, available at: <https://www.ericsson.com/en/blog/2021/10/5g-energy-consumption-impact-5g-nr>
6. Ericsson, **More capacity and less power: How 5G NR can reduce network energy consumption**, 2019, Frenger, P; Tano, R, available at: <https://www.ericsson.com/en/reports-and-papers/research-papers/how-5g-nr-can-reduce-network-energy-consumption>
7. Ericsson Technology Review, **Ensuring energy-efficient networks with artificial intelligence**, April 13, 2021, Vandikas, K; Hallberg, H; Ickin, S; Nyström, C; Sanders, E; Gorbатов, O; Eleftheriadis, L, available at: <https://www.ericsson.com/4972d5/assets/local/reports-papers/ericsson-technology-review/docs/2021/ensuring-energy-efficient-networks-with-ai.pdf>
8. IEEE Xplore, **Green Mobile Networks for 5G and Beyond**, 2019, Masoudi, M et al, available at: <https://ieeexplore.ieee.org/document/8786138>
9. Ericsson Mobility Report, **November 2021**, available at: <https://www.ericsson.com/4ad7e9/assets/local/reports-papers/mobility-report/documents/2021/ericsson-mobility-report-november-2021.pdf>
10. Ericsson blog, **Who cares about peak download speeds in 5G?**, January 18, 2022, Ludwig, R, available at: <https://www.ericsson.com/en/blog/2022/1/who-cares-about-peak-download-speeds-in-5g>

## THE AUTHORS



### Cecilia Andersson

◆ joined Ericsson in 2007 and currently works as a system designer specializing in RAN energy performance. She is a leading promoter of integrating and incorporating the network aspects of energy performance to reduce the total energy consumption of RANs. Andersson holds a Ph.D. in physics from Uppsala University, Sweden.



### Jonas Bengtsson

◆ is a principal developer specialized in RAN energy performance. One of the drivers behind making Ericsson modems the most

energy-efficient in the market 10 years ago, he is now the technical lead of a highly skilled energy performance team that has produced multiple innovations and proof of concepts in the energy performance area. Before joining Ericsson in 2002 he studied computer science at Lund University, Sweden.



### Greger Byström

◆ is a section manager for radio product development, with extensive experience working with functional systemization of several of the energy performance functions within Ericsson radio products. Before joining Ericsson in 2011, he studied engineering physics at Umeå University, Sweden.

### Pål Frenger

◆ joined Ericsson in 1999 and currently holds an



expert position in RAN energy performance at Ericsson Research. He has worked with radio-network energy performance for more than 10 years, has filed more than 300 patent applications and received the Ericsson Inventor of the Year Award in 2017. Frenger holds a Ph.D. in electrical engineering from Chalmers University of Technology, Gothenburg, Sweden.



### Ylva Jading

◆ joined Ericsson in 1999. She currently works as a senior specialist in network energy performance at Ericsson Research. She is

responsible for coordinating and leading the early energy performance activities in Development Unit Networks, as well as researching low energy aspects of future 6G networks. Jading holds a Ph.D. in physics from the Johannes Gutenberg University Mainz in Germany.

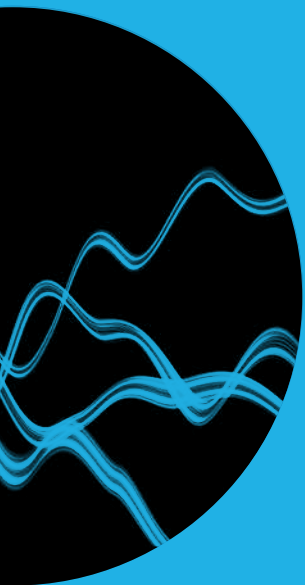


### My Nordenström

◆ is a system developer within Business Area Networks. She joined Ericsson in 2013, and in her current role she primarily focuses on reducing the energy consumption and improving the energy performance for Ericsson's RAN compute products. Nordenström holds an M.Sc. in design and product realization from KTH Royal Institute of Technology, Stockholm, Sweden.

## Further reading

- » **Ericsson blog, Breaking energy utilization with 5G, available at:**  
<https://www.ericsson.com/en/blog/5/2022/breaking-energy-utilization-with-5g>
- » **Ericsson blog, A holistic approach to address RAN energy efficiency, available at:**  
<https://www.ericsson.com/en/blog/2021/12/a-holistic-approach-to-address-ran-energy-efficiency>
- » **Ericsson, Smart, slim and sustainable 5G networks, available at:** <https://www.ericsson.com/en/ran/elevate-5g>
- » **Ericsson blog, Achieving sustainability with energy efficiency in 5G networks, available at:**  
<https://www.ericsson.com/en/blog/3/2021/1/achieving-sustainability-with-energy-efficiency-in-5g-networks>



ISSN 0014-0171  
284 23-3380 | Uen

© Ericsson AB 2022  
Ericsson  
SE-164 83 Stockholm, Sweden  
Phone: +46 10 719 0000