Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

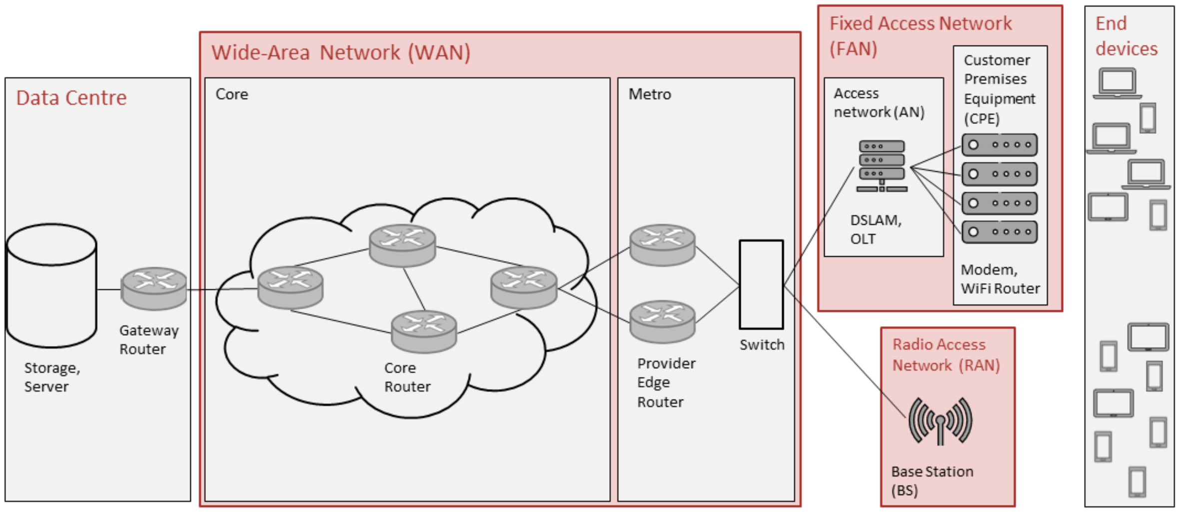
Federal Department of the Environment, Transport, Energy and Communications DETEC

Swiss Federal Office of Energy SFOE
Energy Research and Cleantech Division

Final report dated 28 June 2021

Investigating the Inconsistencies among Energy and Energy Intensity Estimates of the Internet

Metrics and Harmonising Values



Source: Vlad Coroama, 2021

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Date: 28 June 2021

Location: Bern

Publisher:
Swiss Federal Office of Energy SFOE
Energy Research and Cleantech
CH-3003 Bern
www.bfe.admin.ch

Co-financing:
ETH Zürich
Universitätsstr. 6
8092 Zürich
www.ethz.ch

Subsidy recipients:
ETH Zürich
Universitätsstr. 6
8092 Zürich
www.ethz.ch

Author:
Vlad Coroamă, ETH Zürich, vcoroama@ethz.ch

SFOE project coordinators:
Dr. Michael Moser, michael.moser@bfe.admin.ch
Roland Brüniger, roland.brueniger@brueniger.swiss

SFOE contract number: SI/502164-01

Zusammenfassung

Der Bericht präsentiert eine eingehende Analyse von Schätzungen der Energieintensität (EI) und des Gesamtenergieverbrauchs (E) des Internets. Die beiden Werte müssten über den gesamten Internetverkehr (Englisch: traffic, T) verbunden sein, da die Multiplikation von EI mit T zum Ergebnis E führen müsste. Diese Identität wird jedoch durch die heute vorhandenen Studien nicht erfüllt. Im Gegenteil, die mit den jeweiligen Methoden erzielten Ergebnisse liegen einen Faktor von 5 bis 26 auseinander. Die Studie identifiziert zwei wichtige Ursachen dieser Inkonsistenz: die unterschiedliche Behandlung von Zugangsnetzen (Englisch: access networks, ANs) und die Tendenzen hin zu Über- und Unterschätzungen, die der Top-Down- bzw. der Bottom-Up-Modellierung inhärent zugrunde liegen scheinen. Das WAN (Wide Area Network), die FANs (Fixed Access Networks) und die RANs (Radio Access Networks) einzeln betrachtend, zeigt die Studie, dass die grössten Diskrepanzen überraschenderweise nicht auf die relativ jungen und recht heterogenen RANs zurückzuführen sind, sondern auf dem schon recht etablierten WAN.

Die Studie identifiziert zudem drei Ebenen möglicher Harmonisierung von EI und E; in der Rangfolge von Strenge und Präzision, aber auch von zunehmenden Herausforderungen und Unsicherheiten, sind dies: i) eine pauschale Harmonisierung über alle Netztypen, ii) eine, die individuell innerhalb des WAN und innerhalb der Zugangsnetze harmonisiert, und iii) eine, die innerhalb jedes der drei Netztypen harmonisiert. Angesichts der Inkompatibilität innerhalb der bestehenden Literatur bezüglich der Energie- und Energieintensitätsschätzungen für das WAN konnte nur eine globale Harmonisierung erreicht werden. Infolge dieser gegenseitigen Validierung, bei der dennoch versucht wurde, so viel wie möglich auch innerhalb der einzelnen Netztypen zu harmonisieren, ergeben sich folgende Schätzwerte der heutigen Energie- und Energieintensitätswerte: der Energieverbrauch des WAN im Jahr 2020, E_{2020} (WAN) = 110 TWh, E_{2020} (FAN) = 130 TWh und E_{2020} (RAN) = 100 TWh; die Energieintensität des WAN im Jahr 2020, EI_{2020} (WAN) = 0,02 kWh/GB, EI_{2020} (FAN) = 0,07 kWh/GB und EI_{2020} (RAN) = 0,2 kWh/GB. Für kurzfristige Extrapolationen in die Zukunft erscheinen jährliche Energieintensitätsreduktionsfaktoren von 0,8 für WAN und RAN und 0,85 für das FAN sinnvoll und eine eher konservative Wahl.

Zukünftige Anstrengungen sind insbesondere erforderlich, um die Energiebewertungen des WAN zu harmonisieren. Konsistentere Ergebnisse können beispielsweise erzielt werden, indem entweder sowohl Top-Down- als auch Bottom-Up-Methoden eingesetzt werden, eine Hybridmethode angestrebt wird oder zumindest die Ergebnisse über die jeweils andere Methode validiert werden. Besonders wichtig und oft falsch interpretiert sind die Durchschnittswerte der Energieintensität des gesamten Internets. Aus Gründen der Konsistenz müssen diese i) den additiven Charakter der Intensitäten des WAN und der Zugangsnetze, ii) den alternativen Zugang über FAN oder RAN und iii) die (sich schnell ändernden) Anteile dieser beiden Arten von Zugangsnetzen berücksichtigen. Die Studie bietet Gleichungen, die sowohl zur Berechnung dieses Durchschnitts als auch zur Überprüfung der Konsistenz zwischen Schätzungen von EI und E eingesetzt werden können.

Résumé

Ce rapport analyse en profondeur les estimations de l'intensité énergétique (EI) et de la consommation énergétique globale (E) de l'Internet. Les deux devraient être connectés via le trafic Internet global (T), car la multiplication de EI par T devrait donner E. Cette identité, cependant, n'est pas satisfaite par les évaluations existantes aujourd'hui ; au contraire, les résultats obtenus par les deux méthodes sont séparés par un facteur de 5 à 26. L'étude identifie deux sources importantes de cette incohérence : le traitement différent des réseaux d'accès (AN) et les biais inhérents aux surestimations et sous-estimations qui semblent résider respectivement dans la modélisation descendante et ascendante. Considérant individuellement le réseau étendu (WAN), les réseaux d'accès fixe (FAN) et les réseaux d'accès radio (RAN), l'étude montre que, de manière assez surprenante, les écarts les plus importants ne proviennent pas des RAN assez récents et assez hétérogènes, mais du WAN bien établi.

L'étude identifie en outre trois niveaux d'harmonisation entre l'IE et l'E ; par ordre de rigueur et de précision, mais aussi de défis et d'incertitudes croissants, ce sont : i) une harmonisation globale pour tous les réseaux, ii) une harmonisation individuelle au sein du réseau étendu et des réseaux d'accès, et iii) une harmonisation au sein de chacun des trois types de réseaux. Étant donné l'incompatibilité des estimations de l'énergie et de l'intensité énergétique pour le WAN dans la littérature existante, seule une harmonisation globale a pu être réalisée. Suite à cette validation mutuelle, qui a également tenté d'harmoniser autant que possible les différents types de réseaux, une approximation correcte des valeurs actuelles de l'énergie et de l'intensité énergétique est la suivante : la consommation d'énergie du WAN en 2020, E_{2020} (WAN) = 110 TWh, E_{2020} (FAN) = 130 TWh et E_{2020} (RAN) = 100 TWh; l'intensité énergétique du WAN en 2020, EI_{2020} (WAN) = 0,02 kWh/GB, EI_{2020} (FAN) = 0,07 kWh/GB et EI_{2020} (RAN) = 0,2 kWh/GB. Pour les extrapolations à court terme dans le futur, les facteurs de réduction de l'intensité énergétique annuelle de 0,8 pour le WAN et le RAN, et de 0,85 pour le FAN, semblent significatifs et plutôt conservateurs.

Des efforts futurs sont nécessaires en particulier pour harmoniser les évaluations énergétiques du WAN. Des résultats plus cohérents peuvent être obtenus en déployant à la fois des méthodes descendantes et ascendantes, en visant une méthode hybride, ou au moins en validant les résultats via l'autre méthode. Les moyennes globales de l'intensité énergétique d'Internet sont particulièrement importantes et souvent mal interprétées. Par souci de cohérence, ils doivent prendre en compte i) la nature additive des intensités du WAN et des réseaux d'accès, ii) l'accès alternatif via FAN ou RAN, et iii) les parts (en évolution rapide) de ces deux types de réseaux d'accès. L'étude fournit des équations qui peuvent être déployées à la fois pour calculer cette moyenne et pour vérifier la cohérence entre les estimations de l'IE et d'E.

Summary

The report performs an in-depth analysis of estimates of the energy intensity (EI) and the overall energy (E) consumption of the Internet. The two values should be connected via the overall Internet traffic (T), as multiplying EI by T should yield E. This identity, however, is not satisfied by the existing assessments today; on the contrary, the results obtained via the two methods lie a factor of 5-26 apart. The study identifies two important sources for this inconsistency: the different treatment of access networks (ANs), and the inherent biases towards overestimates and underestimates that seem to reside within top-down and bottom-up modelling, respectively. Considering the wide-area network (WAN), fixed access networks (FANs), and radio access networks (RANs) individually, the study shows that, rather surprisingly, the largest discrepancies do not stem from the fairly recent and quite heterogeneous RANs, but from the well-established WAN.

The study further identifies three levels of harmonisation between EI and E; in order of rigour and precision, but also of increasing challenges and uncertainties, they are: i) a global one across all types of networks, ii) one that individually harmonises within the WAN and within the access networks, and iii) one harmonising within each of the three types of the networks. Given the incompatibility of energy and energy intensity estimates for the WAN within existing literature, only a global harmonisation could be achieved. Following from this mutual validation, which also tried to harmonise as much as possible within the individual types of networks, a fair approximation of today's energy and energy intensity values is as follows: the energy consumption of the WAN in 2020, E_{2020} (WAN) = 110 TWh, E_{2020} (FAN) = 130 TWh, and E_{2020} (RAN) = 100 TWh; the energy intensity of the WAN in 2020, EI_{2020} (WAN) = 0.02 kWh/GB, EI_{2020} (FAN) = 0.07 kWh/GB, and EI_{2020} (RAN) = 0.2 kWh/GB. For short-term extrapolations to the future, yearly energy intensity reduction factors of 0.8 for both WAN and RAN, and 0.85 for the FAN, appear meaningful and rather conservative.

Future efforts are needed in particular to harmonise energy assessments of the WAN. More consistent results can be achieved by either deploying both top-down and bottom-up methods, aiming for a hybrid method, or at least validating the results via the other method. Particularly important and often misinterpreted are the overall averages of the energy intensity of the Internet. For consistency, they need to account for i) the additive nature of the intensities of the WAN and the access networks, ii) the alternative access via FAN or RAN, and iii) the (rapidly changing) shares of these two types of access networks. The study provides equations that can be used both to compute this average, and to verify the consistency between estimates of EI and E.

Main findings

- The energy intensity (EI) and the energy (E) consumption of the Internet must be connected via the overall Internet traffic (T): EI [TWh/EB] * T [EB/year] = E [TWh/year].
- This must hold both for the entire Internet and for each network type individually: wide-area network (WAN), fixed-access network (FAN), and radio/mobile access network (RAN).
- While precise data for T exist, the existing estimates of EI and E do not satisfy this identity, yielding results that lie a factor of 5–26 apart.
- The main sources for these discrepancies lie in the inconsistent treatment of access networks and the inherent and converse biases of top-down and bottom-up modelling.
- Only values from the opposing ends of their respective spectra in the existing literature, i.e. high values for the intensities and low values for the energies, can fulfil the identity. Yet discrepancies remain, in particular for the wide-area network (WAN).
- Reasonable estimates for 2020 are: E (WAN) = 110 TWh, E (FAN) = 130 TWh, and E (RAN) = 100 TWh, EI (WAN) = 0.02 kWh/GB, EI (FAN) = 0.07 kWh/GB, and EI (RAN) = 0.2 kWh/GB.

Contents

Abbreviations

4E TCP	technology collaboration programme on energy efficient end-use equipment of the IEA
ADSL	asymmetric digital subscriber line
AN	access network
BS	base station (in RANs)
CDN	content delivery network
CPE	customer premises [networking] equipment (e.g., modems, routers)
DC	data centre
EB	exabyte (10^9 GB)
EDNA	electronic devices & networks annex of the “4E TCP” (see above) of the IEA
FAN	fixed access network
ICT	information and communication technology
IEA	international energy agency
IoT	Internet-of-things
IP	Internet protocol
ISP	Internet service provider
LAN	local-area network
LCA	life cycle assessment
LTE	long-term evolution (4G telecommunication network)
ML	machine learning
MPLS	multi-protocol label switching
NFV	network functions virtualisation
PB	petabyte (10^6 GB)
PON	passive optical network
PSTN	public switched telephone network
PUE	power usage effectiveness
RAN	radio access network
RBS	radio base station (see BS)
SDN	software-defined network
TB	terabyte (10^3 GB)
UPS	uninterruptible power supply
VDSL	very-high-bitrate digital subscriber line
WAN	wide-area network
XG	next generation
ZB	zettabyte (10^{12} GB)

1 Introduction

The broad use of digital information and communication technology (ICT), together with the profound changes in business operations and entire business models it triggers, is often called "digital transformation" – or "digitalisation" in short (Coroamă and Mattern 2019). Given its broad optimisation and substitution potential, digitalisation is increasingly envisioned as a miracle cure to tackle – or at least mitigate – the world's urgent environmental issues; in particular, it is seen as a possible key factor in reducing carbon emissions and resource consumption across various economic sectors.

At the same time, due to trends such as video streaming, cloud computing, Internet-of-Things (IoT), blockchain and machine learning (ML), there is also a growing concern with the own energy consumption of ICT in general, and the Internet in particular. Two traditions are established in the literature, presenting either estimates for the overall energy consumption of the Internet, or for its energy intensity (i.e. the amount of energy needed to transfer a certain amount of data). As will be shown below, these two values are not independent, but connected through the overall volume of Internet data traffic. However, current state-of-the-art estimates from the two traditions are incompatible, yielding results that differ substantially, often by more than an order of magnitude (i.e., a factor of 10). The resulting spread in results reaches from a relatively marginal problem to one of the most important sources for societal energy consumption and anthropogenic global warming.

This situation leads to various adverse consequences: overstated or ignored effects, misinformed press as well as general public and policy makers, counterproductive advocacies, misplaced or lacking policies. The current report addresses this issue by analysing the sources for the existing discrepancy and establishing principles to harmonise the two traditions.

1.1 Background information and current situation

Two types of studies have so far tried to shed light on the energy consumption of the Internet:

i. studies assessing the *energy intensity of the Internet (EI)* (i.e., the energy needed to transfer a certain amount of data), measured in either Joules per bit [J/bit] or kilowatt-hours per gigabyte [kWh/GB], and

ii. studies concerned with the yearly *overall energy consumption of the Internet (E)*, measured in terawatt-hours per year [TWh/year]:

The energy intensity of the Internet (EI) has been addressed in several studies, such as (Schien et al. 2013; Coroama et al. 2013). For fixed access networks, the current academic state-of-the-art indicates an average energy intensity of 0.007 kWh/GB for 2020 (Aslan et al. 2018). For the energy consumption of networks as a whole, on the other hand, there is some divergence in the academic literature, ranging from 325 TWh yearly (Malmodin and Lundén 2018) to 723 TWh (Andrae and Edler 2015).

These two values, EI and E, however, are necessarily connected via the *yearly Internet traffic (T)*, according to the identity in Equation (1): the overall energy consumption of networks (E) is the average energy intensity of the Internet (EI) multiplied by the overall Internet traffic (T),

$$E\left[\frac{TWh}{year}\right] = EI\left[\frac{TWh}{EB}\right] * T\left[\frac{EB}{year}\right] \qquad (1),$$

whereby the dimension terawatt-hours per exabyte [TWh/EB] is equivalent to the more usual [kWh/GB], as $1\,TWh = 10^9\,kWh$ and $1\,EB = 10^9\,GB$.

According to (Cisco 2020), the worldwide Internet data traffic for 2020 was around 254 EB/month, corresponding to 3048 EB (or 3.05 zettabytes, ZB) yearly. Multiplying this value by the 0.007 kWh/GB from (Aslan et al. 2018) would yield an annual energy consumption of only 21.3 TWh/year, one order of magnitude below the low end of current estimates for the overall energy consumption of the Internet from (Malmodin and Lundén 2018). Identity (1) is thus violated, indicating that at least one of the two numbers, E or EI, must be consistently wrong in the literature – under the assumption that the Internet traffic T is not off by an order of magnitude, which, given Cisco's mature methodology (Cisco 2018; 2019) and direct access to router data across the world, seems highly unlikely. To the best of our knowledge, this discrepancy is so far unaddressed in the literature.

1.2 Purpose of the study and structure of the report

In this context, the current study aims at analysing the sources for this inconsistency among the two traditions, and at finding ways towards their harmonisation.

Some of the questions it addresses are:

i) Where do the existing discrepancies among existing assessments of the energy intensity of the Internet (EI) and of the overall energy consumption (E) stem from?

ii) For which of the network types is this misfit particularly prominent among *wide-area networks* (WAN), *fixed access networks* (FANs), and *radio access networks* (RANs)?

iii) Is a harmonisation per network type worthy or necessary, or is a global fit sufficient?

iv) Given the heterogeneous access modes, is the concept of an average Internet energy intensity (irrespective of the access means) meaningful? If yes, what is a suitable metric?

v) For FANs, how must the customer-premises equipment (CPE, e.g., modems and routers) be accounted for: in the same way as the equipment of the core and metro network (energy per amount of data), or as power (i.e., energy per time), as (Coroama et al. 2015; Schien et al. 2015) suggest?

vi) For all questions above, which are the ones with a high degree of data unavailability or uncertainty, and how can it be mitigated?

The remainder of the report is organised as follows: Section 2 briefly addresses early discrepancies in the field, revealing their sources, and discussing why these cannot explain today's mismatches. Section 3 discusses the methods and assumptions deployed in this study, and in particular the system boundaries of the Internet, general assessment principles, and the evolution of Internet traffic. Section 4 provides an in-depth analysis of well-known academic and industry estimates of both the energy intensity and the overall energy consumption of the Internet. It also scrutinises the energy intensity of FANs, and in particular the modelling of CPE, the energy intensity of RANs, and proposes a combined metric for the two. Building on this analysis, Section 5 reveals sources for the existing discrepancies between EI and E assessments today, and makes recommendations on how top-down and bottom-up studies can validate each other. Section 6 summarises the results, and Section 7 discusses the largest uncertainties and derives the necessary next steps.

2 Background: Early discrepancies and their methodological sources

Studies on both the energy intensity of the Internet and its overall energy consumption were published starting in the early 2000s. Presumably the two earliest publications were (Gupta and Singh 2003) and (Koomey et al. 2004). Despite deploying the same method and building on the same data sources, these two early publications arrive to substantially different results, opening a long tradition of discrepancies that persists until today.

To compute the overall energy consumption of the Internet in the US, both (Gupta and Singh 2003) and (Koomey et al. 2004) start their analysis from a detailed inventory of computing and networking devices in the US (Roth, Goldstein, and Kleinman 2002). Using the number of devices in use and their average yearly energy consumption for the US in 2000, the report also estimates the total yearly energy consumption per category of devices. Taking this data as input, (Gupta and Singh 2003) compute the total energy consumption of networking devices in the US as **6.05 TWh/year** in 2000. Starting from the same data for the same year of reference, (Koomey et al. 2004) yield a result of **47.3 TWh/year** energy consumption for the Internet in the US.

2.1 Different system boundaries and understanding of 'the Internet'

Where does this substantial factor of almost 8 stem from? The difference is in what the authors consider to be part of 'the Internet'; in other words, and expressed in the language of environmental sciences, what they include within the *system boundaries* (Coroama and Hilty 2014) of the Internet. While (Gupta and Singh 2003) only include networking devices (core Internet routers 1.1 TWh, WAN switches 0.15 TWh, LAN switches 3.2 TWh, and hubs 1.6 TWh), (Koomey et al. 2004) also include all data centre (DCs) devices within the system boundaries – i..e, four types of servers (10.2 TWh), data storage (1.5 TWh) as well as the uninterruptible power supply (UPS, 5.8 TWh) devices needed for backup power. Adding all these additional DC devices and using the correct consumption figure for LAN switches from (Roth, Goldstein, and Kleinman 2002) – 3.3 TWh and not 3.2 TWh as wrongly indicated by (Gupta and Singh 2003) – they arrive to a total of 23.65 TWh for all devices considered within system boundaries. Ultimately, this number is doubled to the final result of 47.3 TWh yearly to account for the consumption of non-computing devices, in particular the cooling of the devices. The PUE is a measure for the efficiency of a room or building dedicated to the operation of devices, such as Internet routers. It is computed as the facility's total power (including lightning, air conditioning, and so on) divided by the power needed to run the ICT equipment only (Coroama et al. 2013), and assuming a PUE of 2.0 for the year 2000 was certainly no overstatement on part of the authors.

After presenting the total yearly energy consumption of 'networking devices' and the Internet in the US, respectively, both articles proceed to also compute the energy intensity of the Internet. In this second step, they continue using the same methods and data sources. Next to the energy consumption of devices from (Roth, Goldstein, and Kleinman 2002), both articles use a *top-down* methodology to compute the energy intensity of the Internet, dividing its overall consumption by the overall traffic. And both use the same source for the overall data traffic in the US, the 20-35 PB monthly in 2000 as presented by (Coffman and Odlyzko 2002). Using these data sources, a traffic figure towards the middle of the range (29 PB/month) and the simple top-down division, (Koomey et al. 2004) yield an energy intensity of the Internet of 136 kWh/GB for the year 2000. Meanwhile, (Gupta and Singh 2003) set the system boundaries of 'the Internet' even narrower than they had done with 'networking devices', and only consider the subset of devices that constitute the WAN, excluding LAN networking devices. They thus only divide 1.25 TWh/year by the yearly Internet traffic in order to compute the energy intensity of the Internet. As their computation mistakenly assumes the 20-35 PB from (Coffman and Odlyzko 2002) to represent the *yearly* (instead of the *monthly*) traffic, the result is flawed. Had they used the correct traffic value, however, the discrepancy to the result from (Koomey et al. 2004) would have been almost a factor of 40, corresponding to the ratio between 47.3 TWh and 1.25 TWh (divided by the same traffic, respectively).

2.2 Efficiency gains and reference year for the assessment

Another attribute with a strong impact on the result is the reference year for the assessment. This was most obvious around the early 2010s: While the interest in the topic was picking up, two of the best-known assessments could hardly have been more different. As analysed in (Coroama and Hilty 2014) and shown below in Figure 2, the estimates reached a spread by a factor of more than 20,000 – between the 136 kWh/GB by the already presented (Koomey et al. 2004) and the 0.0064 kWh/GB put forward by (Baliga et al. 2011).

Table 5-1: Y2000 AEC by Equipment Type, Total of 97 TW-h

Equipment Type	AEC, TW-h	% of Total AEC	Cumulative % of Total AEC[15]
Monitors	18.8	19%	19%
PCs – Desktop	17.4	18%	37%
Copiers	9.7	10%	47%
UPSs	5.8	6%	53%
Laser Printers	4.6	5%	58%
Server - Low-End	4.5	5%	63%
General Displays	3.4	4%	66%
Server – Workhorse	3.3	3%	69%
LAN Switches (C)	3.3	3%	73%
Facsimile Machines*	3.1	3%	76%
Cell Site Equipment (T)	2.3	2%	78%
Server - Mid-Range	2.0	2%	80%
Workstations	1.8	2%	82%
Transmission (Phone) (T)	1.8	2%	84%
Desktop Calculators*	1.7	2%	86%
Hubs (C)	1.6	2%	88%
Data Storage	1.5	2%	89%
POS Terminals*	1.5	2%	91%
Typewriters*	1.2	1%	92%
Routers (C)	1.1	1%	93%
Public Phone Network (T)	1.0	1%	94%
Private Branch Exchanges (T)	1.0	1%	95%
ATMs*	0.84	1%	96%
Scanners*	0.58	1%	97%
Inkjet Printers	0.56	1%	97%
Wireless Phones* (T)	0.49	1%	98%
PCs – Laptop	0.38	0%	98%
Impact Printers	0.37	0%	99%
Server – High-End	0.37	0%	99%
Supercomputers	0.37	0%	99%
VSATs*	0.23	0%	99%
Voice Mail Systems*	0.19	0%	99%
Line and Other Printers	0.15	0%	99%
WAN Switches (C)	0.15	0%	100%
Modems / RAS* (C)	0.06	0%	100%
CMTS* (C)	0.021	0%	100%

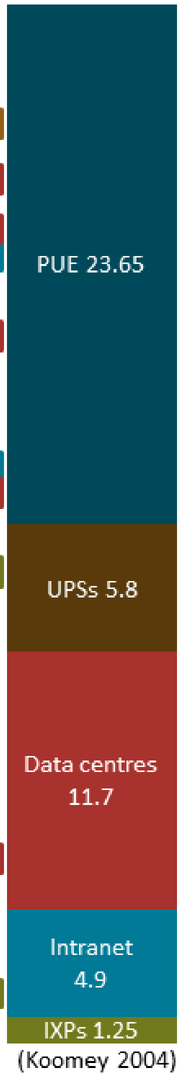


Figure 1 Summary Table 5-1 from (Roth, Goldstein, and Kleinman 2002), comprising types of devices together with their yearly network consumption in the US, used by both (Gupta and Singh 2003) and (Koomey et al. 2004) to compute the energy intensity of the Internet for the year 2000. Reprinted with the authors' permission. Own highlighting showing how distinct system boundaries for 'the Internet', together with the inclusion of PUE or not, can yield vastly different results.

(Baliga et al. 2011) was published by the former Centre for Energy-Efficient Telecommunications (CEET) at the University of Melbourne, led by Kerry Hinton, an institution which published an entire series of highly regarded articles on the topic. Its result is represented in the lower right corner of Figure 2. This substantial influence of the year of assessment comes to no surprise: The Internet communication infrastructure (both Internet routers and fibres) follow the efficiency gains as predicted by Moore's law. According to (Aslan et al. 2018), the energy intensity of the Internet has decreased on average by 30% per year. This corresponds to a halving over 2 years, and to a reduction by a factor of 30 over a decade.

The 11 years between (Koomey et al. 2004) – which refers to the year 2000 – and (Baliga et al. 2011) – which refers to 2011 – thus explain a difference of two orders of magnitude or, more precisely, a factor of 50. Together with another factor of 18 (Aslan et al. 2018) explained by the inclusion of DCs within the system boundary of (Koomey et al. 2004), these two factors explain a difference of about 1,000 between these two examples, or 3 out of the 4 orders of magnitude.

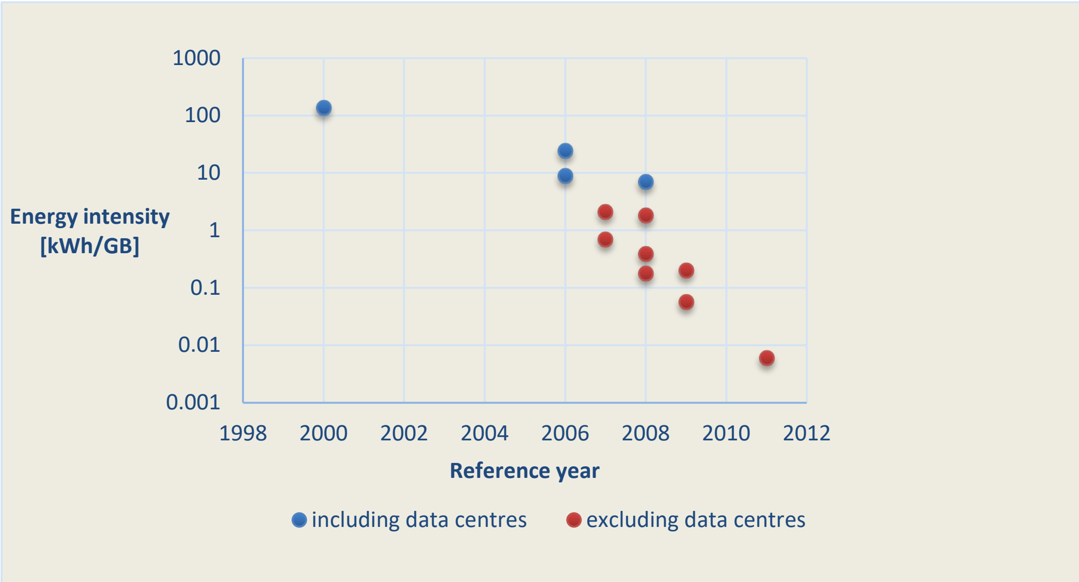


Figure 2 Important early assessments of the energy intensity of the Internet between 2000 and 2011, showing the influence of a) the system boundary – in particular, whether data centres are included (blue) or not (red) – and b) of the year of assessment (due to continuously increasing efficiency). Adapted from (Coroama and Hilty 2014).

2.3 Vanishing relevance of these factors today

Today, the two factors presented above have lost a good deal of their importance. The first one (inclusion of DCs) plays only a marginal role for two reasons: First, as argued in (Coroama and Hilty 2014), it has now become standard to account for DCs separately and not include them in calculations of the energy consumption or the energy intensity of the Internet (they need, though, of course be added on a per-case basis to energy computations for services that include both Internet and DCs, for example video streaming). Second, with the advent of hyperscale DCs, the lowering of the PUE to values of 1.12-1.3, and other improvements, efficiency gains in DCs have been even more impressive than for networks. Even if they were to be added, DCs would no longer dominate the result by an order of magnitude as they used to do, but on the contrary contribute to a much lesser extent than networks and, sometimes even less than the end-user device (Hintemann and Hinterholzer 2020).

The other factor, meanwhile (i.e. the reference year for the assessment), is still of central importance for the resulting energy intensity. It hardly represents a challenge, though: it is conceptually rather trivial, and pragmatically easily accounted for.

For different reasons, thus, these two sources of discrepancies are no longer main reasons for concern. There are, however, several other sources for the current discrepancies. They will be addressed in the reminder of this study. As will be argued below, albeit less prominent than the definition of 'the Internet' and the different year of assessment, these sources have always coexisted with the two which are nowadays well understood. One first hint has been presented already: The inclusion of DCs and a decade of difference explained a factor of 1,000 from the 20,000 between (Koomey et al. 2004) and (Baliga et al. 2011); a factor of 20, however, remains so far unaccounted for.

3 Methods

Given the considerations so far, we will deploy the following methodologies to define system boundaries, relate energy intensity to energy estimates, cope with the different reference years for existing assessments, and assess the yearly Internet traffic.

3.1 System boundaries and topics under scrutiny

In line with most of the literature, e.g. (Baliga et al. 2009; 2011; Coroama and Hilty 2014; Coroama et al. 2015; Schien et al. 2015; Wu, Ryan, and Smith 2019), we consider the following high-level topology of the Internet, as presented in Figure 3: data are usually transmitted from *data centres* (DCs) to *end devices* (sometimes also called end-user devices) via the network. The network consists of the *wide-area network* (WAN), which is “a network spanning a region, country or the globe […] comprising a diverse range of telecommunication networks” (Ryan, Smith, and Wu 2019). The literature sometimes distinguishes between core, metro and edge networks (Baliga et al. 2009; 2011) or core and metro networks (Coroama et al. 2015; Schien et al. 2015), but these distinctions are less relevant to the scope of our assessment, so we will use the simpler model of a WAN. The WAN consists of optical fibres (both land-based and undersea) and large Internet routers, building thus the data transmission highways of the Internet.

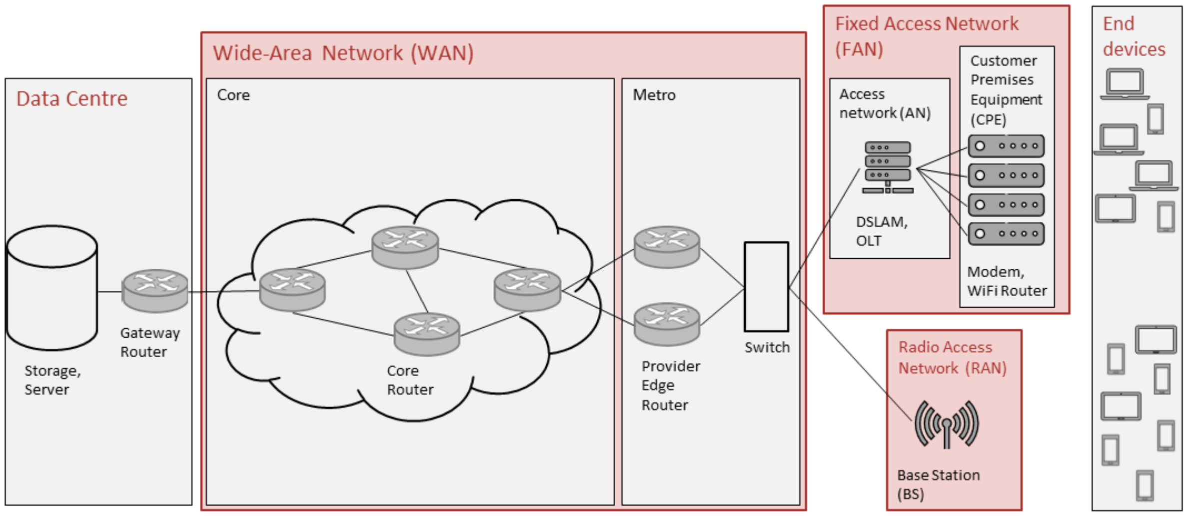


Figure 3 High-level topology of the Internet, distinguishing between data centres (DCs), end devices, and three distinct types of the network: the wide-area network (WAN), fixed access network (FAN), and radio access network (RAN).

The data reaches the end consumers (households or businesses) via *access networks* (ANs), which “connect each home to one of the edge switches” (Baliga et al. 2009). Traditionally, and up to around 2010, ANs have been exclusively fixed access networks (FANs), meaning that a cable (copper coaxial cables or increasingly optical fibres) reach into the users’ homes, although there might be a wireless WiFi component within the users’ homes. As shown in Figure 3, FANs consist of two main components:

- core Internet connection: a street-level component for several households, which, depending on the technology used can be e.g. a digital subscriber line access multiplexer (DSLAM cabinet) for DSL or an optical network unit (ONU) for optical networks, and
- end devices connection: the *customer premises equipment* (CPE), traditionally a modem and/or a router, today almost exclusively one integrated device with both functionalities. CPE

can also include repeaters or other specialised networking equipment. While CPE is physically located within the user's premises, it is logically part of the network (Coroama et al. 2015).

With the parallel emergence of 4G communication networks and the first smartphones in the second part of the 2000s, their rapid spreading during the 2010s, and the later emergence of further relevant technologies (e.g. tablets or 5G networks), mobile/radio access to the Internet became increasingly popular. The corresponding *radio access network* (RAN) corresponds, as sketched in Figure 3 and presented in detail in (ITU-T 2020) to

- core Internet connection: base stations (BSs) or radio base stations (RBSs), which contain several types of equipment such as radio controllers (RC), site equipment (e.g., air conditioning, rectifiers/batteries) as well as backhaul equipment required to connect the BS to the core network (ITU-T 2020), and
- end devices connection: as opposed to the architecture of FANs, in RANs there is no dedicated device to connect the end devices to the BS, as the wireless communication modules are incorporated into the end devices. Strictly speaking, their energy consumption would need to be detached from the energy consumption of the end devices and considered as part of the RANs energy consumption. As this energy, however, is small compared to the consumption of both the other components of end devices (in particular to their screens) and marginal as compared to the energy consumption of BSs, we will ignore the wireless communication modules of the end devices. For the purposes of this study, the RAN thus consists of the BSs.

As argued in (Coroama and Hilty 2014), for both conceptual and practical reasons, DCs and end devices should not be considered as part of the Internet – although their consumption is of course relevant, and for the assessment of individual services (such as video streaming), it must be added to the consumption of the network. This system boundary became mainstream and is in line with the current academic literature as well as recent publications within EDNA (Ryan, Smith, and Wu 2019; Wu, Ryan, and Smith 2019).

Of course, not all Internet data transmissions follow the topology from Figure 3. Internet-mediated human-to-human communication, for example, whether simple voice-over-IP (VoIP) or the increasingly popular video-based meetings (such as Zoom, Microsoft Teams, Google Meet, WebEx, or Skype) imply data transmission between end devices on both sides, although a data centre might be involved as well. This distinction would be relevant for a comparison of the energy intensity of different types of services (say, video calls versus video streaming), but – as we exclude both DCs and end devices – is immaterial to our analysis.

Furthermore, for energy efficiency and user experience reasons, data is increasingly mirrored as closely as possible to the users. This is in particular true for content that will foreseeably be downloaded by numerous users, such as recent software updates or the latest episode of a popular show. Such mirroring servers come in many forms, they can belong to the user's *Internet service provider* (ISP), be collocated on the ISPs premises by the content producer (such as Netflix or Amazon Prime), belong to commercial *content delivery networks* (CDNs) such as Akamai, whose business lies in the mirroring of content for both improved user experience and energy efficiency, or might belong to private CDNs of large content providers such as Google or Facebook. For simplicity and because they are outside system boundaries, such mirroring servers have not been represented in Figure 3. Their existence is nevertheless indirectly relevant to our analysis, as they make numerous Internet data transfers shorter and thus more energy efficient, and as a consequence also the *average* data transfer more energy efficient.

After these considerations, the method of this study can be more clearly stated: review the existing literature for estimates of both the overall energy consumption (E) and the energy intensity (EI) for WAN, FAN, and RAN, unravelling existing discrepancies and understanding their underlying sources. We call these

- E (WAN), E (FAN), and E (RAN), and
- EI (WAN), EI (FAN), and EI (RAN), respectively.

To this end, we performed an extensive literature research, documenting which network part is taken into consideration by each individual study. Relating studies of the Internet energy with studies of its energy intensity via the Internet traffic, we then studied existing inconsistencies and aimed at harmonising them.

Throughout the study we consider that the energy intensity of the network is independent of the data type transferred (Aslan et al. 2018). Different services (email, www, video streaming, etc) might induce different energy consumption in the end devices, but not along the network.

3.2 Assessment principles

Most bottom-up studies under scrutiny present their results as *energy intensity of data* (EI_D), i.e., as amount of energy needed to transfer an amount of data across the Internet. The energy intensity of data is usually expressed in [kWh/GB], although sometimes other units are used such as [TWh/EB] (which is actually equivalent), [J/bit], or [J/Mbit]. The necessary conversions were performed to express all values in [kWh/GB].

Some of the studies, however, present not the energy, but the *carbon intensity of data* (EI_D), i.e., the amount of greenhouse gases emitted to send an amount of data across the Internet. The (EI_D) is usually expressed in [g CO_2 / GB] or the equivalent [t CO_2 / PB]. To transform CI_D to EI_D, we use Equation 2 below

$$EI_D \left[\frac{kWh}{GB}\right] = {CI_D \left[\frac{g\ CO_2}{GB}\right]} \Big/ {CI_E \left[\frac{g\ CO_2}{kWh}\right]} \qquad (2)$$

where CI_E is the carbon intensity of electricity expressed in [g CO_2 / kWh].

The electricity intensity of data can thus be derived by dividing the carbon intensity of data by the carbon intensity of electricity. Of course, as a prerequisite for the correctness of Equation (2), the system boundaries for the intensities of data need to coincide with those of the carbon intensity of electricity. If the scope for data is global, for example, the average worldwide carbon intensity of electricity needs to be used. If the scope for data is a specific country, the carbon intensity of the electricity mix of that specific country shall be used. If the carbon intensity of data refers to a specific company which, for example, uses a large share of renewables, then the average carbon intensity of that specific mix needs to be taken into account.

We further assume that for a specific geography (the world, a continent, or a country) and for a specific time period (a day, a month, or an year), all traffic that is being transmitted over the WAN is then being delivered via either a FAN or a RAN. In other words, and focusing on entire years,

$$T_y(WAN) = T_y(FAN) + T_y(RAN) \qquad (3)$$

for any region, where $T_y(X)$ indicates all the data transmitted over the X type of network in year y.

More and more data are not delivered from far-away DCs, but from servers at, or close to, the edge network. They can be provider edge servers, collocated servers of content providers (such as Netflix or Google) or servers belonging to CDNs such as Akamai. This means that for an increasing number of Internet data flows, the WAN part of the transmission will be a rather short one, sometimes not even reaching the core network, but only metro and edge. As this does not represent the aim of this study, we do not model it explicitly; it is, however, implicitly represented in the ever decreasing EI (WAN). Equation (3) remains unaffected; while the core data flows do certainly not equal the total data flows of the ANs, the WAN consists not only of the core, but also of metro and edge networks.

Given Equation (3) and the definition of E (X) and EI (X), the energy and energy intensity of the Internet, respectively, can be computed as follows:

$$E_y(Internet) = E_y(WAN) + E_y(FAN) + E_y(RAN) \quad (4)$$

$$EI_y(Internet) = EI_y(WAN) + (1 - s_{RAN}) * EI_y(FAN) + s_{RAN} * EI_y(RAN) \quad (5)$$

where $E_y(X)$ and $EI_y(X)$ are the total energy and average energy intensity of X (i.e., the entire Internet, the WAN, all FANs, or all RANs) for the year Y, respectively, and s_{RAN} represents the share of RAN among all access networks for that same year (currently around 16% and steadily growing).

Equation (4) is a direct consequence of our definition of 'the Internet', as discussed in the previous section. The reason behind Equation (5), on the other hand, is that the average energy intensity of the Internet is the cumulative energy intensities of the WAN and of the AN. As, however, the AN is composed by either FAN or RAN, these need to be added corresponding to their respective relative weights, which is what Equation (5) does.

3.3 Period of assessment and harmonising the time perspective

The early studies (such as those presented in Section 2) are hardly comparable to today. The methods were just evolving back then, as the huge discrepancies already presented show. The topology of the Internet was also quite different, as were the technologies used at all stages (DCs, networks, end devices): for example, hyperscale DCs were just emerging, there was no RAN yet, nor smartphones or tablets. More desktop PCs than laptops were in use, and CRT screens were still mainstream. As important as these early studies were for the development of the field, we do no longer consider them as relevant (other than from a historic perspective) and will thus ignore them from now on.

Our main interest lies in the numbers of today (i.e., 2020). We will, however, also consider studies referring to the last decade, 2010-2019, for three reasons: First, many of the ubiquitously cited and sometimes also societally discussed studies are already a couple of years old and still very relevant. Second, trends are at least as important as today's numbers, and to understand trends some historic perspective is important (albeit not a too long one, as discussed above). Finally, the authors themselves often make predictions that reach to today and beyond.

While looking at the last decade of data, the crucial importance of the reference year (as discussed in Section 2.2) needs to be taken into account. In any instantiation of Equations (4) and (5) above, the year Y must be the same in all terms of the respective instantiation. Due to the need of using the same year in each such computation, we also rewrite Equation (1) from the introduction, introducing the same parameter Y to highlight this fact:

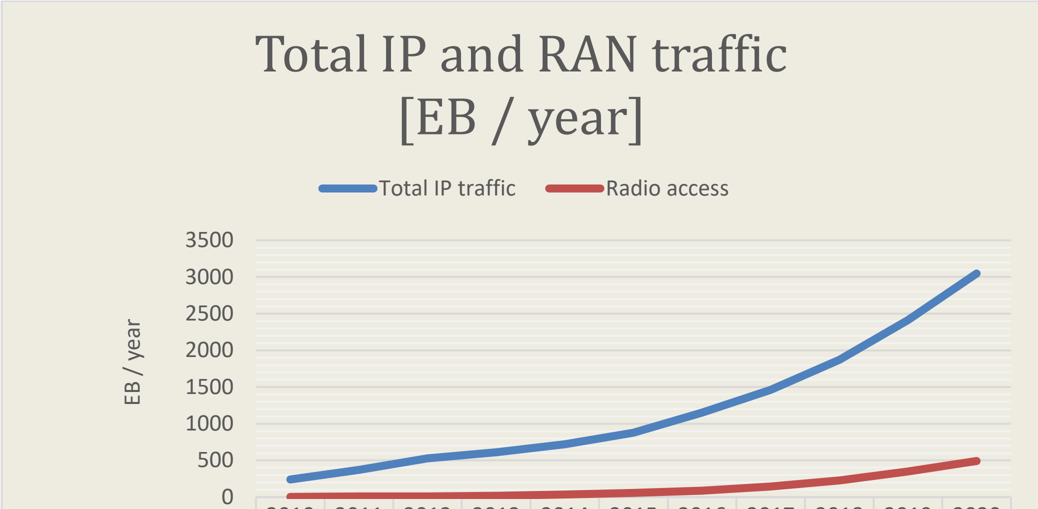
$$E_y\left[\frac{TWh}{year}\right] = EI_y\left[\frac{TWh}{EB}\right] * T_y\left[\frac{EB}{year}\right] \quad (6)$$

Most studies mention to which year they refer to. When this is not explicitly described, and given the usual times for data gathering, analysis, and academic publication process, we make the following assumption: papers published in the second half of one year (i.e., after the 1 July) refer to the previous year, papers published in the first half of the year (i.e., before 30 June) refer to two years before.

Extrapolations are typically performed by the authors themselves. For the rare instances when we perform them ourselves, we use the growth rate computed by the authors for a past period to extrapolate a couple of years into the future. We do not perform, nor accept from the literature, any extrapolation reaching more than 7-8 years into the future; the uncertainties become simply too large for any serious statement.

3.4 Internet traffic

For more than a decade, the single, outstanding, and very reliable source of data for the Internet traffic has been Cisco. Their reports were first called ‘Visual Networking Index’ (VNI), split later into ‘VNI: Global IP Traffic Forecast’ and ‘VNI: Global Mobile Data Traffic Forecast’, and more recently again reunited under the new name ‘Annual Internet Report’. From the plethora of data presented by these studies, we are mainly interested in the worldwide aggregates, both for the total IP traffic and the share of mobile/radio access.



	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Total IP traffic	240	372	528	612	720	876	1152	1464	1872	2412	3048
Radio access	2.4	7.2	10.8	19.2	33.6	56.4	88.8	144	228	348	492

Figure 4 Yearly traffic expressed in EB, as compiled from several Cisco reports. Both the total traffic (blue), and the share of traffic delivered via RAN (red), are shown.

For the years relevant to this study, 2010-2020 (see Section 3.3 above), we compiled data from several of these reports (Cisco 2013b; 2013a; 2015; 2016; 2018; 2019; 2020). The results are presented in Figure 4. Radio access is increasing at a much faster pace than fixed network access and WAN traffic: While Cisco foresaw an average annual growth rate of 26% for the entire IP traffic over the period 2017-2022 (Cisco 2018), it foresaw an average annual growth rate of 46% for the global RAN traffic over the same period (Cisco 2019). As a consequence, the share of radio access is continuously increasing, from just 1% in 2010 to over 16% today (see Figure 5).

Given its networking equipment placed around the world, Cisco is in a unique position to deliver reliable data on the worldwide data traffic. For more than a decade, its past forecasts have proven in hindsight to have been remarkably accurate. Given these facts, we take the traffic numbers put forward by Cisco as corresponding exactly to the reality. Hence, if there is an inconsistency, for example in Equation (6), it cannot be due to the traffic factor, and the reason must lie within the imprecision of either or both of the other factors.

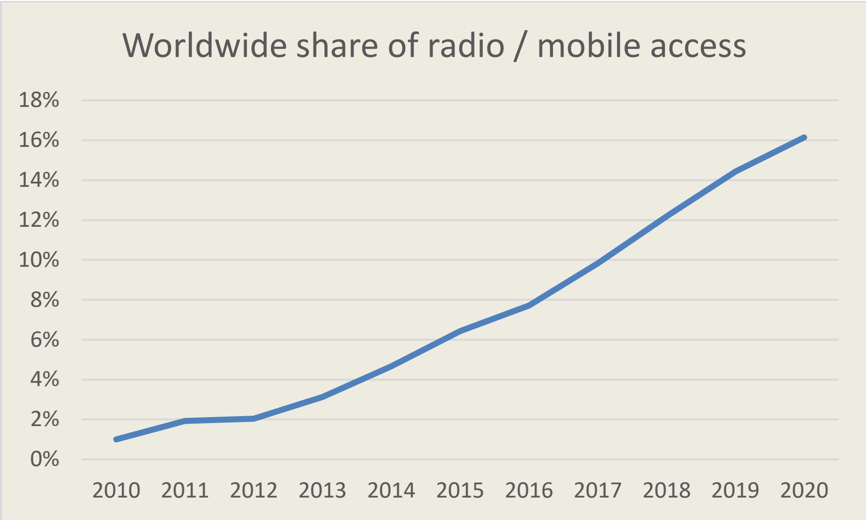


Figure 5 Share of radio/mobile access among the entire worldwide IP traffic, as compiled from different Cisco sources.

4 Analysis of existing estimates

In this section, we present a number of well-known academic estimates of both the energy intensity and the overall energy consumption of the Internet. They have been chosen and are analysed according to the criteria defined in Section 3. These academic estimates are complemented with recent numbers on the energy intensity of the RAN reported by telecom companies.

4.1 Estimates of the energy intensity (EI) of WAN, FAN, and RAN

We start by presenting estimates of the EI of WAN, FAN, and RAN. They are grouped by estimates for the last decade (2010 – 2019), current academic estimates, and current telecom industry estimates, respectively.

4.1.1 Estimates referring to the past decade (year of reference: 2010 – 2019)

(Williams and Tang 2012) estimate the use-phase power needed for the Internet transfer at the average broadband speed of 6.2 Mbit/s as 18.4 W (0.0184 kW). Dividing the latter 18.4 W by the former 6.2 Mbit/s yields an energy intensity of 2.97 J/Mbit, which is equivalent to **EI_{2010} (WAN) = 0.007 kWh/GB**. The authors, however, allocate the power consumption of all network components only for the duration of the data transfer, and assume a load of 100%, leaving thus both the idle energy unaccounted for, and dividing the power of the network components to an unrealistically high load. Both these aspects make the result even more of an underestimate than it would have been due to the general issues of bottom-up estimates presented in our study.

Another early bottom-up study analyses the impact of the location of content (CDN vs. distant server) on the energy intensity of Internet content delivery. It yields an average energy intensity of the WAN of **EI_{2010} (WAN) = 0.057 kWh/GB**, or 25 J/Mbit (Schien et al. 2012). As opposed to (Williams and Tang 2012), the authors do not idealise the network, but take into account the actual load of networking equipment and distributing it over the actual traffic. Moreover, they also account for “redundantly operated edge and core routers (hardware is duplicated to cope with failure)” (Schien et al. 2012).

(Malmodin et al. 2012) is insofar an atypical study, as although deploying a top-down methodology, it does not only present overall consumption results, but – by dividing through the yearly Swedish network traffic for 2010 (the year of assessment) – it also presents intensities (both energy and carbon) for the Internet in Sweden. Its result for the WAN is **EI_{2010} (WAN) = 0.08 kWh/GB**. The paper also makes a reference to the energy intensity of the AN by claiming that it is fourfold the size of the intensity of the WAN, while not distinguishing between FAN and RAN: "The use stage electricity consumption of the assessed IP core network was about one fourth of the energy consumption of the corresponding access networks (telephony, 2G and 3G mobile and fixed broadband)" (Malmodin et al. 2012). For 2010, there is no meaningful assumption that can be made retroactively as to the ratio of FAN to RAN. Additionally, the paper presents figures that seem to be in contradiction to the claim stated above: its Table 2 puts forward 19 GWh for the PSTN FAN in Sweden, 13 GWh for 2G RAN, and 8 GWh for 3G RAN. The total of 30 GWh is *3 times smaller* than the 126 GWh for the Swedish WAN the same table lists, and not *4 times larger*, as otherwise claimed throughout the paper. We thus consider only the WAN from (Malmodin et al. 2012) and ignore its analysis of the ANs.

(Schien et al. 2013) are to our knowledge the first, and still among the few, to present individual energy intensities for the WAN, FAN, and RAN alike, using the same system boundaries as we used here. Their estimates are:

- **EI_{2012} (WAN) = 0.038 kWh/GB**,
- **EI_{2012} (FAN) = 0.019 kWh/GB**, and
- **EI_{2012} (RAN) =0.293 kWh/GB**, RAN which was still 3G at the time.

For a top-down study of the energy intensity of the RAN, (Pihkola et al. 2018) divided the estimated total electricity consumption of Finnish RANs by the total data transferred through all its mobile networks. To this end, they used publicly available data on the total mobile data transmission volume, which is regularly reported by the Finnish Communications Regulatory Agency, and energy consumption data published in their sustainability reports by the three main mobile telecom operators in Finland, which together account for 99% of mobile subscriptions (Pihkola et al. 2018). The results show an impressive improvement of the RAN energy intensity by a factor of over 40 in just 7 years: from **EI_{2010} (RAN) = 12.34 kWh/GB** down to **EI_{2017} (RAN) = 0.30 kWh/GB**. The number for 2012 is also interesting: The top-down computation yields a result of **EI_{2012} (RAN) = 4.49 kWh/GB**, as opposed to the 0.293 GB/kWh for the same year assessed with a bottom-up methodology by (Schien et al. 2013).

4.1.2 Current estimates

As stated in the Introduction, the study by (Aslan et al. 2018) is the current state of the art of the academic understanding on the energy intensity of the WAN. Extrapolating exponential past trends (up to 2017) via least squares fit yields a value of around **EI_{2019} (WAN) = 0.01 kWh/GB** and **EI_{2020} (WAN) = 0.007 kWh/GB**. The study does not address the energy intensities of either FAN or RAN.

(Pihkola et al. 2018) also extrapolate their 2010-2017 results into the future, devising an energy intensity of just below **EI_{2020} (RAN) = 0.1 kWh/GB**.

(Wu, Ryan, and Smith 2019) is a comprehensive study looking at all three components of the network and have system boundaries compatible to the ones used in this study. They present both numbers for the overall consumption as well as for the energy intensity. For each of the three network categories, they define three types of deployed technologies: 'legacy/traditional', 'modern', and 'next generation' (XG), as follows:

- WAN
  - legacy: backhaul and metro network for public switched telephone network (PSTN)
  - modern: fibre optic, multi-protocol label switching (MPLS)
  - next gen.: software-defined networking (SDN), network functions virtualisation (NFV)
- FAN
  - legacy: asymmetric digital subscriber line (ADSL)
  - modern: very-high-bitrate digital subscriber line (VDSL), cable, passive optical network (PON)

- next gen.: XG PON
- RAN
    - legacy: 2G & 3G networks
    - modern: 4G networks
    - next gen.: 5G networks

The study deploys the following assumptions – most of them included in the study itself, some revealed through personal communication by the lead author Anson Wu (Wu 2021): Starting from current data, the yearly energy efficiency improvements of ANs are as follows: 10% for FAN XG, 5% for 4G (which is already quite mature), and 15% for 5G. Given the historic improvement of 22% for 4G based on (Andrae and Edler 2015), the 15% for 5G in the future are considered a conservative estimate. Although estimates consider 5G's final efficiency 10 times higher than that of 4G (Nokia 2020) – and thus its energy intensity 10 times lower – 5G is a sum of evolving technologies that didn't unfold their potential from the outset. The authors thus assume the same energy intensity for 4G and 5G RAN until (and including) 2018, moment from which they start to diverge. By 2030, the last year of the authors' forecast, the two would be indeed a factor of 10 apart (0.3 kWh/GB for 4G, 0.03 kWh/GB for 5G). For 2020, the intensities are 0.073 kWh/GB for 4G and 0.022 kWh for 5G, respectively.

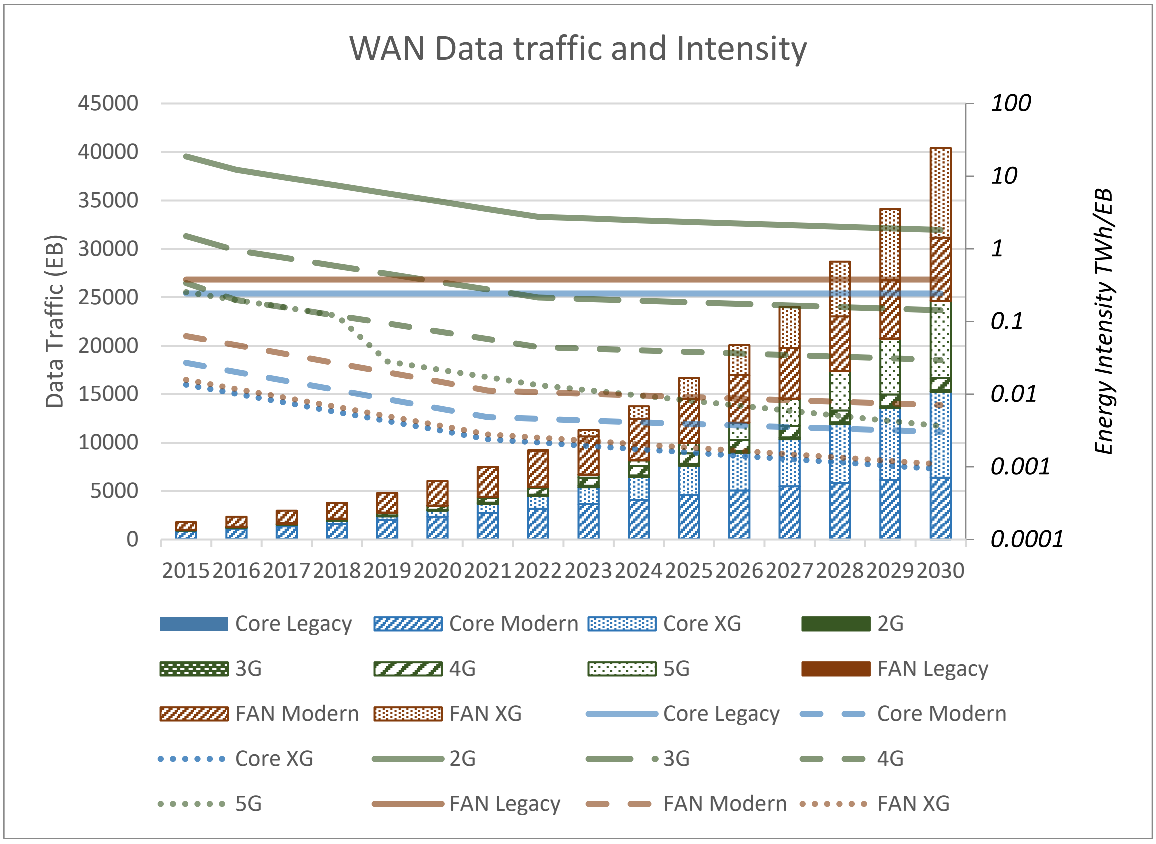


Figure 6 Overview Fig. 15 from (Wu, Ryan, and Smith 2019), presenting the energy intensities of the different types of networks (right vertical axis; lines in the graph, not bars). Reprinted with the authors' permission.

Figure 6 reproduces Fig. 15 from (Wu, Ryan, and Smith 2019), which graphically shows the aggregated results of the study. Via personal communication (Wu 2021), the exact numbers were received. The study presents a range of results depending on the various technologies still deployed in parallel, as listed in the bullet points above. As we are interested in comparison of single numbers, we proceeded as follows: For the energy intensities of WAN, FAN, and RAN, we chose the number in the middle (i.e., the *modern* technology as opposed to *legacy* or *next generation*) in the same way as done for (Andrae and Edler 2015) below, keeping in mind the different semantics of these numbers: While for (Andrae and Edler 2015) they represented the *expected case scenario* (as opposed to worst and best case), here they represent the estimates for the *modern technology* (as opposed to legacy and next generation technologies). With this assumption, the intensities for 2020 are as follows:

- **EI_{2020} (RAN) = 0.0064 kWh/GB**,
- **EI_{2020} (FAN) = 0.015 kWh/GB**, and
- **EI_{2020} (RAN) = 0.073 kWh/GB**.

Finally, (Bieser et al. 2020) analyse the footprint of Swisscom's mobile networks, developing two scenarios, one for today (i.e., 2020) and one for 2030. Most details on the methodology used are included in the study itself, some were revealed through personal communication by one of the study authors, Roland Hischier (Hischier 2021).

The goal of the study was to assess the direct effects of mobile networks on climate change in Switzerland. Given the focus of the study, FANs are not considered; RANs, however, are obviously at the core of the analysis as well as a share of the WAN that is dedicated to delivering the data accessed via RAN. In 2020, Swisscom's RAN consisted, next to the 2G-5G BSs, also of 33,000 km dedicated fibres. From the 'common backbone' (WAN, in our terminology), i.e., 70,000 km of core and metro fibres and the corresponding routers that Swisscom deploys throughout Switzerland, 10% were allocated to mobile networks, corresponding to the share of data accessed via RANs (Bieser et al. 2020).

A life-cycle assessment (LCA) was used, taking into account the lifecycle climate impact of the networks under study (mainly RAN and the corresponding share of the WAN, as discussed above). Both the overall emissions for Swisscom's RANs (plus share of WAN), and the carbon intensity within these system boundaries (i.e., kg CO_2eq / GB) were computed (Bieser et al. 2020). The authors followed a bottom-up approach, modelling an average BS site in close cooperation with the technology partners Swisscom and Ericsson (Hischier 2021). For 2020, the average energy consumption of a joint BS site for 2-4G technology was 17094 kWh/year and for 5G technology substantially lower at 5732 kWh/year (mainly due to the more efficient radio units). The resulting carbon intensities are 0.03 kg CO_2eq / GB for 2-4G RAN, and 1.01 CO_2eq / GB for 5G (Bieser et al. 2020).

This surprisingly bad value for 5G (despite its potential to be substantially more energy efficient than the older technologies) is due to the incipience of the technology, and in particular to little availability of 5G-compatible end devices but an already substantial rollout of 5G BSs; the energy (and carbon) footprint of the 5G RAN is allocated to relatively few devices and thus quite a low traffic. The 2030 scenario expects this to substantially change, 2G and 3G networks to be entirely faded out over this coming decade, and the carbon intensities of the remaining 4G and 5G technologies (5G being the dominating one by then) to have decreased to 0.008 kg CO_2eq / GB and 0.004 CO_2eq / GB, respectively.

Given the slightly different focus and system boundaries of this study, we use the following assumptions to transform its results and make them compatible to our analysis:

- As FANs were not at all covered, and the WAN only to the extent that it was relevant to the carbon emissions induced through RANs, we limit our interpretation of the study to the RANs.
- While for its research question (i.e., the carbon emissions induced through mobile networks), the system boundaries in (Bieser et al. 2020) stand to reason, our approach is different: we consider WAN, FAN, and RAN as three distinct parts of the network irrespective of the reason they came into existence, and analyse their respective energy intensities. To adapt the system boundaries, we would thus need to remove the 10% WAN considered within system boundaries by (Bieser et al. 2020). Their effect, however, is relatively marginal. They represent just 17.5% of the total fibre length (7,000 km among the total 40,000 km considered,

the other 33,000 km being dedicated RAN fibres and thus within our system boundary as well), and the fibres altogether have a lower consumption than the BSs. As the overall influence of this WAN share is below 5%, and there are more important uncertainties, we do not explicitly correct the system boundary here but do it together with the transition from LCA to use phase (see two bullets below).

- Due to the very low 5G usage, the results for 2020 are an outlier and are thus ignored. We consider the 2G-4G carbon intensity of 0.03 kg CO_2eq / GB as typical for 2020 (equivalent to 30 g CO_2eq / GB).
- The authors have taken an LCA approach, while all other studies consider only the use phase of networks. Fig. 5 on page 28 of the study reveals the individual shares of 6 categories: use and production each of core, transport and access networks. We eliminate the production of all these categories (41%) and the use phase of the core (1%, representing the share of the WAN attributed to data accessed over RAN, as discussed two bullets above), and remain with 58% which are relevant to our study (52% use-phase of the 'access network', i.e., the BSs, and 6% 'transport network', i.e., dedicated RAN fibres and part of the RAN in our understanding). From the 30 g CO_2eq / GB from (Bieser et al. 2020), a share of 58% (i.e., 17.5 g CO_2eq / GB) is thus relevant to our boundaries.

- Finally, we need to reverse-engineer from the carbon intensity of data to its energy intensity using Equation (2). Switzerland has a very low carbon intensity of its electricity production of only around 20 g CO_2eq / kWh, as the overwhelming majority of its electricity production is hydro or nuclear (BFE 2020). Due to electricity imports, however, the consumption mix has a substantially larger intensity of 149.5 g CO_2eq / kWh (BAFU 2018). Using this figure in Equation 2 together with the CI_D of 17.5 g CO_2eq / GB, as discussed above, yields an energy intensity of data (EI_D) of **EI_{2020} (RAN) = 0.117 kWh/GB**.

4.1.3 Current industry reports (RAN energy intensity)

Finally, we reviewed sustainability reports of leading European mobile telecom companies from Germany, France, the UK, Switzerland, Sweden and Finland. The only one found to directly devise the energy intensity of its RAN is the Finnish Elisa: according to the report, the intensity of its RAN decreased from **EI_{2019} (RAN) = 0.15 kWh/GB** to **EI_{2020} (RAN) = 0.12 kWh/GB** (Elisa 2020).

The German Telekom, meanwhile, does not report on the RAN individually, but on the average energy intensity of its entire networks: Dividing their overall 2020 energy consumption of 12,643 GWh by the entire IP data volume of 106 EB yields an energy intensity for (WAN + AN) of 119 kWh / TB (Telekom 2020), or 0.12 kWh / GB. This number includes the WAN, the RAN and the (negligible) street-level FAN, but not the CPE. To account for the CPE as well, we assume that the 49.1 million Telekom customers with a fixed line – 27.4 million ADSL and 21.7 million broadband (Telekom 2020) – have one WiFi router each, with a power drain of 8 W that are all continuously on. This implies an energy consumption of 252 MJ yearly each, equivalent to 70.08 kWh/year. For all 49.1 million customers, another 3,441 GWh need to be added to account for the CPE, leading to 16,084 GWh for all of WAN, FAN and RAN (including CPE) of Telekom. Dividing this number by Telekom's total 2020 traffic of 106 EB yields a slightly larger total energy intensity of 151 kWh / TB, or 0.15 kWh / GB.

No indication is given of how this is split between the WAN on one side and the access networks (FAN or RAN) on the other. We can make a reasonable estimate using a couple of assumptions based on the share of radio access as well as the ratios among the intensities of the individual networks presented in the literature, i.e.: 20% of access via RAN, 80% via FAN, the intensity of the RAN 10 times higher than that of the WAN and twice as high as the FAN intensity (including CPE). This yields a system of 3 equations with 3 variables, which can be unequivocally solved to

- **EI_{2020} (WAN) = 0.02 kWh/GB**,
- **EI_{2020} (FAN) = 0.1 kWh/GB**, and

- **EI_{2020} (RAN) = 0.2 kWh /GB**,

compatible to (albeit a bit larger than) the numbers presented by (Elisa 2020).

The Vodafone group devises the carbon intensity (expressed in t CO2eq / PB, which corresponds to g CO2eq / GB) of its RAN in the sustainability reports: It decreased from 926 g CO2eq / GB in 2017 to 577 in 2018 and 371 in 2019 (Vodafone 2019), and then further to 230 g CO2eq / GB in 2020 (Vodafone 2020). This number can be split between the carbon intensity of data outside the UK (242 g CO2eq / GB) and the one within the UK (120 g CO2eq / GB). To transform this latter figure to the energy intensity of data, we use it together with the 233 g CO2eq / kWh carbon intensity of electricity from UK's government conversion factors for company reporting of greenhouse gas emissions (BEIS 2020) in Equation (2), which yields an energy intensity of **EI_{2020} (RAN) = 0.51 kWh / GB**, a factor of 4 higher than the result of (Elisa 2020) and a factor of 2.5 higher than our estimate of Telekom's RAN intensity, based on the data from (Telekom 2020) and the assumptions mentioned above.

4.1.4 Summarising the reports on the energy intensity of the Internet

Table 1 below summarises the estimates on the energy intensity of WAN, FAN, and RAN presented above.

Table 1 Summary of existing estimates for EI (WAN), EI (FAN), and EI (RAN) for the last decade (2010 – 2019) and for today.

Period	Study	Year of ref.	EI [kWh/GB]			Type
			WAN	FAN	RAN	
Last decade 2010-2019	(Williams and Tang 2011)	2010	0.007			Bottom-up
	(Schien et al. 2012)	2010	0.057			Bottom-up
	(Malmodin et al. 2012)	2010	0.08			Top-down
	(Schien et al. 2013)	2012	0.038	0.019	0.293	Bottom-up
	(Pihkola et al. 2018)	2012			4.49	Top-down
	(Pihkola et al. 2018)	2017			0.3	Top-down
Current academic	(Aslan et al. 2018)	2020	0.007			Bottom-up
	(Pihkola et al. 2018)				0.12	Top-down
	(Wu et al. 2019)		0.0064	0.0150	0.0732	Unknown
	(Bieser et al. 2020)				0.117	Bottom-up
Current industry	(Elisa 2020)	2020			0.12	Top-down
	(Telekom 2020)		0.02	0.1	0.2	Top-down
	(Vodafone 2020)				0.51	Top-down

4.2 Estimates of the overall energy consumption (E) of WAN, FAN, and RAN

This section presents estimates of the worldwide overall energy consumption of WAN, FAN, and RAN. They are grouped by estimates for the last decade (2010 – 2019) and current estimates, respectively.

Several of the current numbers presented in Section 4.2.2 below are extrapolations from the earlier studies which will be presented first, in Section 4.2.1. Some of these extrapolations were put forward by the authors themselves. For (Van Heddeghem et al. 2014) and (Malmodin and Lundén 2018), due to their notoriety and the fact that they reach quite different results, we perform the extrapolations to 2020 ourselves, using in both cases the growth rate observed by the respective authors in the past to extrapolate several years into (their) future, i.e., to 2020.

4.2.1 Overall energy consumption of WAN, FAN, and RAN: data for 2010 – 2019

(Lanzisera, Nordman, and Brown 2012) are among the first top-down studies to organise their data in such a way that what we call WAN ('switching' and 'enterprise routers') and FAN ('enterprise WLAN', 'security appliances', and 'customer access equipment' and 'residential CPE') are clearly distinguishable. The study presents numbers for both the US and the world for 2007 and 2008, and extrapolations to 2009-2012. We use here the worldwide numbers for 2010: **24.2 TWh for WAN** and **33.6 TWh for FAN**. Given their still marginal role back then, RANs were not considered.

(Van Heddeghem et al. 2014) is a well-known top-down estimate of the development of the energy footprint of all the ICT equipment in the world between 2007-2012. The study considers three main categories: DCs, end devices (called 'personal computers', as desktop and laptop PCs together with their screens were the main category of end devices at that time), and networks. For this latter category, which is the one of relevance to our study, the paper distinguishes between three types of networks: 'telecom operator networks', 'office networks' and 'customer premises access equipment (CPAE)':

- The last category (CPAE), which we call CPE, is clearly part of the FAN.
- The first category, telecom operator networks, consists mainly of the WAN, but also includes the street-level part of the FAN.
- Office networks, whose importance has decreased since, would have been back in the day a mix between FAN and WAN.

Overall, the energy consumption of networks grew from 200 TWh / year in 2007 to 334 TWh / year in 2012. With 77% (257 TWh) of the total, telecom operators networks dominate this result, while office networks contribute with 7% and the CPE with 15%. Simplifying, we attribute the study's 'telecom operator networks' entirely to WAN, although this yields a slight overestimate as they also include the street-level part of the FAN. To compensate, we attribute the office networks entirely to the FAN, although they include both FAN and WAN devices. The errors induced through these two simplifying assumptions are much less than the influence of other uncertainties, in particular the extrapolation of these results from 2012 to the present, as presented below. These boundaries imply **E_{2012} (WAN) = 257 TWh** and **E_{2012} (FAN) = 77 TWh**. As by 2012 RANs only played a marginal role, they were not covered in the study.

(Andrae and Edler 2015) present data for 2010-2013, albeit it is not always clear which year precisely they refer to, and extrapolate all the way to 2030. Taking into account all three network components, their boundary between WAN and FAN is slightly different from the one we consider here, and they also use different terminology. By 'FAN', their study means what we (along with most of the literature) call both WAN and FAN together; while 'our' WAN is called 'fixed access wired' and the FAN 'fixed access Wi-Fi': "The core network, metro/edge network, content distribution network, wired access network, customer premises equipment and wireless local area networks, are regarded here as one primary entity called fixed access networks (FAN). FAN is divided into two parts, fixed access wired and fixed access Wi-Fi" (Andrae and Edler 2015).

Additionally, they only attribute the CPE to the FAN, including street-level AN (such as DSLAM cabinets or ONU units) to the WAN: "The core network, metro/edge network, content distribution network, wired access network, customer premises equipment and wireless local area networks, are regarded here as one primary entity called fixed access networks (FAN). FAN is divided into two parts, fixed access wired and fixed access Wi-Fi. Fixed access Wi-Fi is here regarded as customer premises equipment" (Andrae and Edler 2015). We would thus need to subtract the street-level AN from the WAN and add it to the FAN. As there is, however, no indication in the paper of the underlying assumptions or the sources for their numbers, and as presumably the CPE dominates the energy consumption within the FAN, we do not perform any actions, but consider the 'fixed access wired' from (Andrae and Edler 2015) as WAN, and the 'fixed access Wi-Fi' as FAN.

RANs are called 'wireless access networks' in (Andrae and Edler 2015), which distinguishes between five categories: 2G/3G voice traffic, 2G data, 3G data, 4G data, and 5G data. For our analysis, we add these numbers together for the RAN. The results for 2012 are as follows:

- **E_{2012} (WAN) = 196 TWh**
- **E_{2012} (FAN) = 51 TWh**
- **E_{2012} (RAN) = 197 TWh**

Finally, (Malmodin and Lundén 2018) present data for the years 2010 and 2015, with no future extrapolations. In section 4.2 'network operations', they report numbers on WAN (called 'fixed networks') and RAN (called 'mobile networks'). For **WANs**, there is a growth of 5% in these five years (from E_{2010} (WAN) = 100 TWh to **E_{2015} (WAN) = 105 TWh**) and for **RANs**, as expected, a much more substantial growth of 61% (from E_{2010} (RAN) = 85 TWh to **E_{2015} (RAN) = 137 TWh**). The street-level part of the FAN is presumably included in the WAN, but it is indistinguishable within the data. The other part of **FANs**, the CPE, is included with a larger category 'Telephones, Home Network Equipment, and Set-Top Boxes', but the authors do devise their energy consumption separately (Table 3 in the paper), and thus: **E_{2015} (FAN) = 83.5 TWh**.

4.2.2 Overall energy consumption of WAN, FAN, and RAN extrapolated to 2020

Taking the figures for 2012 by (Van Heddeghem et al. 2014), and using the study's own annual growth rates of 10.2% for WAN and 10.8% for CPE (i.e., FAN), we extrapolate the results to today. This would yield **E_{2020} (WAN) = 559 TWh** and **E_{2020} (FAN) = 175 TWh**, but of course the uncertainties over such a period are high.

(Andrae and Edler 2015) present estimates for the years 2010-2013 and extrapolations into the future up to 2030. For these extrapolations, they use three scenarios: best case, expected, and worst case. As the range between best and worst is quite large, we use the middle scenario, i.e., the expected case. For 2020, the results are as follows:

- **E_{2020} (WAN) = 439 TWh**
- **E_{2020} (FAN) = 185 TWh**
- **E_{2020} (RAN) = 99 TWh**

For (Malmodin and Lundén 2018), we also use the growth coefficients devised by the authors between 2010 and 2015 (5% for WAN, 61% for RAN) to extrapolate another 5 years into the future, arriving at **110 TWh for WANs** and **220 TWh for RANs** for 2020, respectively. In the supplementary material of the article, the authors make an interesting observation, that the "number of fixed broadband subscriptions and TV subscriptions can be used to estimate CPE (customer premises equipment, e.g. modems / gateways), STB's TV's (and TV peripherals)" (Malmodin and Lundén 2018). We thus use the same 5% extrapolation we used for WAN (and which is based on the authors' own extrapolation from 2010 to 2015) to extrapolate the energy consumption of CPE, which yields **88 TWh** for the CPE, and thus for the **FAN**, in 2020.

4.2.3 Summarising the reports on the overall energy consumption of the Internet

Table 2 summarises the estimates on the total energy consumption of WAN, FAN, and RAN presented above.

Table 2 Summary of existing estimates for the overall E (WAN), E (FAN), and E (RAN) for the last decade (2010 – 2019) and for today.

Period	Study	Year of ref.	E [TWh/year]			Type
			WAN	FAN	RAN	
Last decade 2010-2019	(Lanzisera et al. 2012)	2010	24.2	33.6		Top-down
	(Andrae and Edler 2015)	2012	196	51	197	Unclear
	(van Heddeghem et al. 2014)	2012	257	77		Top-down
	(Malmodin and Lunden 2018)	2015	105	83.5	137	Top-down
Current	(van Heddeghem et al. 2014)	2020	559	175		Top-down
	(Andrae and Edler 2015)	2020	439	185	99	Unclear
	(Malmodin and Lunden 2018)	2020	110	88	221	Top-down

5 Recent discrepancies and their sources

The state-of-the-art study on the energy intensity of the Internet, widely used and accepted in the academic world, is arguably the one put forward by (Aslan et al. 2018), which estimates 0.01 kWh/GB energy intensity for 2019 and **0.007 kWh/GB for 2020**. Multiplying its energy intensity value according to Equation (6) by the entire Internet traffic for the same year, 3048 EB (Cisco 2020), yields an energy consumption of 21.3 TWh. A naïve interpretation of the study would be to read this number as the overall energy consumption of the entire Internet in 2020.

In fact, the study explicitly excluded RAN from its analysis. And while it seems to be considering both WAN and FAN – "This study identifies representative estimates for the average electricity intensity of fixed-line Internet transmission networks over time and suggests criteria for making accurate estimates in the future" (Aslan et al. 2018) – it does actually not include CPE in its analysis. As CPE represent the lion's share (in terms of energy consumption) of the FAN, the estimate put forward by (Aslan et al. 2018) can thus be interpreted as the *current state-of-the-art understanding of the EI (WAN)*. However, even when applying Equation (6) only to the WAN, the result of 21.3 TWh is still a **factor of 5.2** smaller than the low end of the top-down estimates of the WAN overall energy consumption, the 110 TWh very conservatively extrapolated by us for (Malmodin and Lundén 2018).

Compared to the other estimates of E (WAN), the discrepancies are even larger:

- a **factor of 20.6** as compared to the 439 TWh from (Andrae and Edler 2015), and
- a **factor of 26.2** as compared to our extrapolation to 2020 of the data from (Van Heddeghem et al. 2014).

These large discrepancies show that the consideration of different system boundaries alone does not explain the difference among estimates of the energy intensity of the Internet and those assessing the its overall energy consumption. In fact, such mismatches leading to the violation of Equation (6) are persistent among the recent and current estimates of the EI and the E of the Internet. We discuss several examples in Section 5.1 and then address their sources in Section 5.2.

Table 3 Juxtaposition of different estimates for the energy intensity of the WAN, FAN, and RAN (columns 3-5), which multiplied with the respective traffic numbers (columns 6-8) according to Equation (7) should yield a total energy consumption of the Internet (column 9) comparable to the one obtained by adding estimates of the energy consumption of the WAN, FAN, and RAN (columns 10-12) together (column 13). In general, the identity postulated by Equation (7) is not achieved, the directly estimated overall Internet energy (column 13) being a couple of times larger than that based on the intensities (column 9). The ratio (column 13 / column 9) is presented in column 14. The table compares row-wise results of existing studies as follows:

* Rows 1–3 'WAN only': EI (WAN) from (Aslan et al. 2018), E (WAN) from (Malmodin and Lundén 2018), (Andrae and Edler 2015), and (Van Heddeghem et al. 2014), respectively. Own extrapolations for (Malmodin and Lundén 2018) and (Van Heddeghem et al. 2014), as described in the text.
* Row 4: all energy intensities (EIs) from (Schien et al. 2013), all energies (Es) from (Van Heddeghem et al. 2014).
* Row 5: all EIs from (Schien et al. 2013), all Es from (Andrae and Edler 2015).
* Row 6: as above, but with EI (RAN) substituted for the much larger top-down estimate from (Pihkola et al. 2018).
* Row 7: all EI values from (Wu, Ryan, and Smith 2019) – mid-range, 'modern' technologies (as opposed to 'legacy' or 'next generation'); E (WAN) and E (FAN) from (Van Heddeghem et al. 2014), E (RAN) from (Malmodin and Lundén 2018), extrapolated as described in the text.
* Row 8 'mainstream': EI (WAN) from (Aslan et al. 2018), EI (FAN) from (Wu, Ryan, and Smith 2019), EI (RAN) from (Pihkola et al. 2018; Bieser et al. 2020; Elisa 2020), all Es from (Andrae and Edler 2015)..
* Row 9 cherry-picking to achieve Equation (7): EI (WAN) from (Aslan et al. 2018), EI (FAN) own assumption as described in the text, EI (RAN) from (Telekom 2020); E (WAN) and E (FAN) from (Malmodin and Lundén 2018) very conservatively extrapolated from 2015 to 2020, E (RAN) from (Andrae and Edler 2015).
* Row 10: more cherry-picking to work towards also satisfying Equation (8): EI (WAN) and EI (RAN) per our computation based on (Telekom 2020), EI (FAN) own assumption as described in the text; E (WAN) from (Malmodin and Lundén 2018) extrapolated, E (FAN) average between (Malmodin and Lundén 2018) and (Van Heddeghem et al. 2014), E (RAN) from (Andrae and Edler 2015).

Row	Year	EI [TWh/EB]			T [EB / year]			E [TWh/year]	E [TWh/year]				Discrepancy Factor
		WAN	FAN	RAN	WAN	FAN	RAN	Internet	WAN	FAN	RAN	Internet	
1	2020, WAN only	0.007			3048			**21.3**	110			**110.3**	5.2
2		0.007			3048			**21.3**	439			**439.0**	20.6
3		0.007			3048			**21.3**	559			**559.0**	26.2
4	2012	0.038	0.019	0.293	528	517.2	10.8	**33.1**	257	77	0	**334.0**	10.1
5		0.038	0.019	0.293	528	517.2	10.8	**33.1**	196	51	197	**444.0**	13.4
6		0.038	0.019	4.49	528	517.2	10.8	**78.4**	196	51	197	**444.0**	5.7
7	2020	0.0064	0.0150	0.0732	3048	2556	492	**93.7**	559	175	221	**954.6**	10.2
8		0.007	0.0150	0.12	3048	2556	492	**118.6**	439	185	99	**723.0**	6.1
9		0.007	0.07	0.2	3048	2556	492	**298.7**	110	88	99	**296.9**	1.0
10		0.02	0.07	0.2	3048	2556	492	**338.3**	110	131.5	99	**340.8**	1.0

5.1 Inconsistencies between energy intensity and energy studies, and first harmonization attempts

In the few examples given in the beginning of Section 5, we have applied Equation (6) for the WAN and the year 2020. These examples are also depicted in rows 1–3 of Table 3: Column 3 shows the energy intensity of the WAN as computed by (Aslan et al. 2018), column 6 the worldwide Internet traffic for 2020 (Cisco 2020), and column 9 the multiplication of the two according to the right side of Equation (6). Columns 10 and 13 show both the results of E (WAN) studies: (Malmodin and Lundén 2018), (Andrae and Edler 2015), and (Van Heddeghem et al. 2014), respectively. According to Equation (6), columns 9 and 13 should coincide. In fact, the E (WAN) estimates are consistently above the result of EI (WAN) * T (WAN) with a factor depicted in the last column 14.

Generalising from the exclusive consideration of the WAN to all three parts of the network (WAN, FAN, and RAN), we need to use in the left side of Equation (6) the computation for the overall energy of the Internet from Equation (4), and on its right side the computation for the average energy intensity of the Internet, $EI_y(Internet)$, as developed in Equation (5). Remembering also that FAN-traffic plus RAN-traffic equal the WAN-traffic (Equation 3), yields the general Equation 7:

$$E_y(WAN) + E_y(FAN) + E_y(RAN) = EI_y(WAN) * T_y(WAN) + EI_y(FAN) * T_y(FAN) + EI_y(RAN) * T_y(RAN) \qquad (7)$$

The two sides of this equation are presented in Table 3:

- Columns 3–5 list EI (WAN), EI (FAN), and EI (RAN), respectively, as listed in Table 1. The units [kWh/GB] used in Table 1 and [TWh/EB] from Table 3 are hereby equivalent, as $1\ TWh = 10^9\ kWh$ and $1\ EB = 10^9\ GB$. We use the former in Table 1, as it is the usual and widely used unit for the energy intensity of data transfers, and the latter in Table 3 as well as in Equations (1), (6), and (7), to underline the correctness of the respective identities.
- Columns 6–8 present the traffic for the individual components and the corresponding year, T (WAN), T (FAN), and T (RAN), as indicated by the various Cisco studies.
- Columns 10–12 present E (WAN), E (FAN), and E (RAN), respectively, as estimated in the individual studies and listed in Table 2.
- Columns 9 and 13 compute the right and the left side of Equation (7), respectively. Inputs are the values from columns 3–8 and 10–12, respectively.
- Finally, column 14 computes the ration between column 13 and 9, i.e., the discrepancy factor between the left side of Equation (7) and its right side, where the two should actually have been equal.

In rows 4–5 we compare the intensity estimates by (Schien et al. 2013) to two well-known estimates of the Internet energy: (Van Heddeghem et al. 2014), which does not take RAN into consideration, and (Andrae and Edler 2015), respectively. As Table 3 shows, both yield an inconsistency of one order of magnitude. As the estimate for EI (RAN) from (Schien et al. 2013) seems quite low for the year 2012, in row 6 of Table 3 we substitute it by the top-down estimate for the EI (RAN) from (Pihkola et al. 2018), which is more than one order of magnitude higher. With this, the computation of the total RAN energy starting from its intensity also grows more than tenfold and reduces as a consequence the overall discrepancy between the two to less than half its former value (5.7 as compared to 13.4). The comparison of rows 5 and 6 raises some fundamental observations about bottom-up vs. top-down modelling, which will be addressed in Section 5.3 below.

For the year 2020 and all parts of the network, row 7 compares the mid-range energy intensities for WAN, FAN, and RAN from (Wu, Ryan, and Smith 2019) to some known estimates of the overall energy consumption: E (WAN) and E (FAN) from (Van Heddeghem et al. 2014) and E (RAN) from

(Malmodin and Lundén 2018); all extrapolated by us as described above. Using the mid-range energy intensities for WAN, FAN, and RAN from (Wu, Ryan, and Smith 2019) yields an overall energy consumption of just 93.7 TWh, which is one order of magnitude less than the (extrapolation of) the direct top-down assessments. Such extrapolations are by no means singular; the widely cited study (Belkhir and Elmeligi 2018), for example, uses precisely such extrapolations to reach similarly high values.

Row 8 of Table 3 presents what might be regarded as the state-of-the-art and most widely used understanding of both the energy intensities and the energy consumptions in use today:

- EI (WAN) = 0.007 kWh/GB from (Aslan et al. 2018) as argued above,
- EI (RAN) = 0.12 kWh GB, as unanimously reported by (Pihkola et al. 2018; Bieser et al. 2020; Elisa 2020),
- EI (FAN) = 0.015 kWh/GB as reported by (Wu, Ryan, and Smith 2019), and which corroborates to a back-of-the-envelope calculation of 800 million CPE in the world, running 24/7 and needing 7 W of power on average, transporting the 2556 EB of T (FAN) in 2020, which would yield an intensity of 0.019 kWh/GB.
- Finally, E (WAN), E (FAN), and E (RAN) from (Andrae and Edler 2015).

These ‘mainstream assumptions’ lead to a discrepancy by a factor of 6.1, showing a fundamental mismatch between the two types of estimates.

To arrive to relatively similar numbers and be able to satisfy to a large extent the identities as shown in Equations (6) and (7), some creative cherry-picking within the realistic ranges of the two types of studies is necessary. Table 3 shows this in its row 9, which deploys the following estimates:

- EI (WAN) = 0.007 kWh/GB from (Aslan et al. 2018), as argued earlier,
- EI (RAN) = 0.2 kWh/GB, as indicated by (Telekom 2020),
- EI (FAN) = 0.07 kWh/GB, 10 times worse than EI (WAN) but around 3 times better than EI (RAN),
- E (WAN) and E (FAN), from the most optimistic study by some margin (Malmodin and Lundén 2018), very conservatively extrapolated by us from 2015 to 2020 using just a 5% extrapolation factor, but E (RAN) not from the same study, but from (Andrae and Edler 2015), which devise E (RAN) = 99 TWh as opposed to our extrapolation from (Malmodin and Lundén 2018) that would have been 221 TWh, as shown in Table 2.

With these creative combination of rather pessimistic assumptions on one side, and the most optimistic on the other, the identity from Equation (7) can finally be almost perfectly fulfilled (299 vs. 297 TWh).

These estimates, whose exact values are presented in row 9 of Table 3, can be rounded as follows:

- EI_{2020} *(WAN) = 0.007 kWh/GB,*
- EI_{2020} *(FAN) = 0.07 kWh/GB,*
- EI_{2020} *(RAN) = 0.2 kWh/GB,*
- E_{2020} *(WAN) ~* E_{2020} *(FAN) ~* E_{2020} *(RAN) = 100 TWh*

All of these values – but for EI_{2020} *(FAN)*, which we computed as outlined above – are from the literature: EI_{2020} *(FAN)* from (Aslan et al. 2018) and very similar to (Wu, Ryan, and Smith 2019), EI_{2020} *(RAN)* from (Telekom 2020), which is (together with our assumptions) more pessimistic than other industry studies, E_{2020} *(WAN)* and E_{2020} *(FAN)* from (Malmodin and Lundén 2018), conservatively extrapolated from 2015 to 2020, and E_{2020} (RAN) from (Andrae and Edler 2015).

5.2 Harder constraints required for better harmonisation

The identity reflected in Equation (7) and considered so far was derived from the general identity in Equation (6), by expanding the overall energy of the Internet and its average energy intensity according to Equations (4) and (5), respectively. Equation (7) represents, however, just the simplest constraint connecting estimates of overall energies and energy intensities. There are two further steps of increasing rigour, increasing difficulty of compliance but also increasing uncertainties involved. They are presented in this subsection.

The next level of rigour places an additional constraint on the ratio between the overall energy of the WAN on one side and the energies of the ANs on the other; in other words between E_y (WAN) and E_y (FAN) + E_y (RAN). It starts from the observation from Equation (3) that the traffic over FAN and RAN equals the traffic over WAN. If we now consider the ratio

$$(E_y(FAN) + E_y(RAN)) / E_y(WAN)$$

and replace the energy in each part of the network according to

$E_y(X) = EI_y(X) * T_y(X)\ for\ X \in \{WAN, FAN, RAN\}$, this ratio becomes

$$(EI_y(FAN) * T_y(FAN) + EI_y(RAN) * T_y(RAN)) / EI_y(WAN) * T_y(WAN).$$

Further remembering that s_{RAN} represents the ratio of radio/mobile access among all access networks (i.e., together with Equation (3), $T_y(RAN) = s_{RAN} * T_y(WAN)$, and $T_y(FAN) = (1 - s_{RAN}) * T_y(WAN)$, we can rewrite this ratio as

$$(EI_y(FAN) * (1 - s_{RAN}) * T_y(WAN) + EI_y(RAN) * s_{RAN} * T_y(WAN)) / EI_y(WAN) * T_y(WAN),$$

yielding the following new identity

$$(E_y(FAN) + E_y(RAN)) / E_y(WAN) = ((1 - s_{RAN}) * EI_y(FAN) + s_{RAN} * EI_y(RAN)) / EI_y(WAN) \tag{8}$$

In other words, the ratio between the added energies of access networks and that of the wide-area network should equal the ratio between the energy intensities of the access networks (weighted by their respective share of data transfer) and the energy intensity of the WAN.

This additional constraint posed by the identity from Equation (8) questions the numbers put forward at the end of Section 5.1, which were chosen in such a way as to satisfy the fundamental identity from Equation (7). Using those numbers on both sides of Equation (8), would yield *a factor of about 2* on the left side of the identity (as the energies of the three parts of the network are considered roughly equal, around 100 TWh yearly each), and *a factor of about 13* (12.91, to be precise) on the right side of the identity (as EI (FAN) is considered 10 times larger than EI (WAN), and EI (RAN) even 30 times larger but with a lower weight of 16%).

The root of the issue lies in the unsolved mismatch between mainstream academic estimates of the energy intensity of the WAN and estimates of its overall energy, as also presented in rows 1–3 of Table 3. While even the low end of energy estimates devises over 100 TWh yearly (Malmodin and Lundén 2018), several sources (Aslan et al. 2018; Wu, Ryan, and Smith 2019) – some of which are regarded as the state-of-the-art of current understanding – indicate a low energy intensity of the WAN, which, when multiplied by the traffic only yield 21 TWh yearly; a factor of 5 lower than the lowest overall energy estimates.

To partially address this issue, we choose for row 10 in Table 3 the highest EI_{2020} (WAN) number from Table 1, i.e., the 0.02 kWh/GB we computed based on (Telekom 2020), with the assumptions described in Section 4.1.3. This leads to an $EI_{2020}(WAN) * T_{2020}(WAN)$ value of 61 TWh/year, still barely half as large as the 110 TWh from (Malmodin and Lundén 2018), but already much closer.

Keeping the other EI values identical to row 9, we need to compensate on the total energy side. As there is quite a split between the values for $EI_{2020}(FAN)$ between (Malmodin and Lundén 2018) on one side and (Andrae and Edler 2015) and (Van Heddeghem et al. 2014) on the other, we choose a value in-between the 88 TWh from (Malmodin and Lundén 2018) and the 175 TWh from (Van Heddeghem et al. 2014) – **E_{2020} (FAN) = 131.5 TWh**. Overall, the values thus are

- *EI_{2020} (WAN) = 0.02 kWh/GB* (→ EI_{2020} (WAN) *T_{2020} (WAN) = 61 TWh),
- *EI_{2020} (FAN) = 0.07 kWh/GB* (→ EI_{2020} (FAN) *T_{2020} (FAN) = 179 TWh),
- *EI_{2020} (RAN) = 0.2 kWh/GB* (→ EI_{2020} (RAN) *T_{2020} (RAN) = 98 TWh),
- *E_{2020} (WAN) = 110 TWh,*
- *E_{2020} (FAN) = 132 TWh,*
- *E_{2020} (RAN) = 99 TWh.*

With these values, the identity in Equation (7) is again largely satisfied, while the discrepancy between the two sides of the identity from Equation (8) is no longer 13 to 2, but 4.54 to 2.09. As the two sides of Equation (8) should be equal, this difference is still substantial, but a clear progress compared to the results from the end of Section 5.1.

Finally, the final level of rigour would be to satisfy Equation (6) for all three individual types of networks: WAN, FAN, and RAN. This is expressed with the rather trivial equations below:

$$E_y(WAN) = EI_y(WAN) * T_y(WAN) \tag{9}$$

$$E_y(FAN) = EI_y(FAN) * T_y(FAN) \tag{10}$$

$$E_y(RAN) = EI_y(RAN) * T_y(RAN) \tag{11}$$

If Equations (9)–(11) are simultaneously fulfilled, both Equation (7) and also the supplemental identity from Equation (8) will be fulfilled as well. Three levels of rigour for harmonising estimates of the overall energy with those of the energy intensity can thus be identified:

1. Fulfilling the identity from Equation (7),
2. fulfilling the identity from Equation (7) and additionally that from Equation (8), and
3. fulfilling Equations (9)–(11) simultaneously.

Level 1 is already a large progress over no validation whatsoever. It implies a macro-level harmonisation, although there can be remaining inconsistencies (with opposing directions) among the individual network types – as we had with the values at the end of Section 5.1. Achieving level 2 means an additional harmonisation within the WAN, and one within the ANs taken together, while achieving level 3 implies a harmonisation within each type of network.

Given current estimates, we could only achieve level 1 (values in row 9 of Table 3) and make some progress towards level 2 (values in row 10 of Table 3). As we argue below, further research is thus required. While a higher level of harmonisation is generally beneficial, however, it must also be noted that higher levels depend on additional uncertainties: While level 1 depends mainly on reliable overall Internet traffic numbers, level 2 also crucially depends on the assumption that all WAN traffic is further sent over ANs (which, due to Equation (3), is also implicit in Equation (7) at level 1, but not as critical there). Level 3 further depends on a correct value for s_{RAN}, the share of radio access among all access. While this number was also relevant to earlier levels, for satisfying the identities on level 3 it becomes crucial. Worthy pursuing, the additional constraints introduced in Equation (8) and Equations (9)–(11), respectively, can thus also be a case of no longer seeing the forest for the trees.

5.3 Methodological sources for the current discrepancies

Section 5.1 established a consistent misfit between estimates of the energy intensity of the Internet and estimates of its overall energy consumption. According to the identity put forward in Equation (6), further refined in Equation (7), multiplying the former by the overall traffic yields consistently lower results than the latter, often by quite a substantial factor.

Two important sources for historic discrepancies were discussed in Section 2. However, they can no longer explain today's mismatch: First, DCs are no longer considered within system boundaries. And even if they were, they would no longer dominate the end result. Second, while the reference year of the assessment continues to be a crucial parameter, it is easily accounted for, and including the parameter Y in our Equations (3-7) reflects this need. All the mismatches described in Section 5.1 occur for the same year of reference, excluding it as a possible source.

Analysing the results, methods and assumptions of the academic and industry studies presented in Section 4, two main sources for the current discrepancies become evident: i) modelling the access networks, in particular of CPE for the FAN and of the RAN altogether, and ii) top-down versus bottom-up modelling, which tend to set the boundaries too wide and too narrow, respectively, and thus yield inconsistent results. They are both discussed below.

5.3.1 Inconsistent system boundaries and modelling access networks

As can be observed in the overview given by Table 1, the energy intensity of the AN is higher than that of the WAN. This is particularly valid for the RAN, for which current state-of-the-art estimates range around 0.12 – 0.2 kWh/GB (Pihkola et al. 2018; Wu, Ryan, and Smith 2019; Bieser et al. 2020; Elisa 2020; Telekom 2020) as compared to the 0.006 – 0.007 kWh/GB for the WAN (Wu, Ryan, and Smith 2019; Aslan et al. 2018); a factor of about 20 more.

With the wide adoption of the much more energy efficient (NGMN Alliance 2015; Fernández-Fernández, Cervelló-Pastor, and Ochoa-Aday 2017; Gourhant and Tuffin 2020, 5; Nokia 2020) 5G radio, RANs might ultimately become more energy efficient than FANs (Wu, Ryan, and Smith 2019). They will both, however, remain most likely less energy efficient than the WAN. Additionally, with sharply increasing data transfers worldwide, the trend of mirroring content at the provider edge is likely to continue. In particular popular content such as the newest episodes of trending series or recent software updates are likely to be delivered not across the core network, but from CDNs, edge servers of the ISP, or servers of content providers collocated at edge sites of ISPs. This will contribute to a further increase of the energy efficiency (and thus the decrease of the energy intensity) of the WAN, which for an increasing portion of the traffic will consist only of few edge and/or metro routers.

Accounting for the AN is thus of paramount importance for consistent system boundaries and for eloquent results that reflect the entire reality. This means in particular accounting for the CPE part of the FAN and the entirety of the RAN infrastructure. Not considering the AN can lead to estimates such as (Aslan et al. 2018), which are then easily and often misinterpreted as 'the energy intensity of the Internet' and can lead to massive understatements such as the 21.3 TWh/year we computed according to Equation (6).

5.3.2 Inherent biases of top-down versus bottom-up modelling

A typical flaw of early bottom-up studies was to model an idealised network based on state-of-the-art technology and ignore, among others, legacy or redundancy equipment. This error that underwent to, e.g., (Baliga et al. 2009; 2011), led to early underestimated outliers, such as the energy intensity of 0.0064 kWh/GB for the entire Internet in 2010 put forward by (Baliga et al. 2011). Such assumptions have been rightfully criticised by (Aslan et al. 2018): "studies should consider the full range of equipment in use within the network under study. This includes considering the legacy equipment within networks. Estimates based on specific or state-of-the-art equipment, such as Baliga and colleagues (2009), omit the less efficient legacy equipment (i.e., equipment with higher electricity use per GB of data transferred) in use within country-wide Internet networks, resulting in a substantial underestimate of electricity intensity at the lower end of the observed range (0.004 kWh/GB for 2008)" (Aslan et al. 2018).

Despite this rightful criticism, however, (Aslan et al. 2018) itself puts forward an estimate of 0.007 kWh/GB which, even when correctly interpreted as valid for the WAN exclusively, still leads to estimates of the overall consumption that are a factor of 5 – 26 below top-down assessments of the energy consumption of the entire WAN (Malmodin and Lundén 2018; Andrae and Edler 2015; Van Heddeghem et al. 2014), as discussed in the beginning of Section 5.1.

Along similar lines, the bottom-up assessment from (Schien et al. 2013) yields an energy intensity of just 0.293 kWh/GB for the RAN in 2012, despite also accounting for legacy technologies. The top-down assessment in (Pihkola et al. 2018) devises an energy intensity of 4.49 kWh/GB for the RAN in 2012 – a factor of 15 higher.

It thus appears that even when they aim to be as accurate as possible, accounting for legacy and redundancy equipment, aware that numerous technologies are coexisting and old technologies tend to be much less energy efficient than newer generations, bottom-up studies nevertheless lead to consistently smaller estimates than top-down ones. As argued by several authors (Schien et al. 2013; Coroama and Hilty 2014; Aslan et al. 2018), part of the explanation is that bottom-up modelling often ignores devices such as legacy technology, redundancy equipment, or simply necessary equipment that can easily be overseen in an assessment, such as uninterrupted power supply (UPS). In fact, due to the sheer complexity of the networks, their topologies and the plethora of equipment involved, bottom-up modelling might be inherently bound to ignore parts of the network, set its system boundaries to narrow and thus contain an inherent bias towards underestimates.

Top-down modelling, however, might have the exact opposite bias at its core: setting the system boundaries unintentionally too wide, considering equipment and processes that should not be included, and thus contain an inherent bias towards overestimates. To stress this point, we use here the (hypothetical) example of a top-down analysis of the energy consumption of WAN and LAN in Switzerland based on a recent corporate report by Swisscom, a Swiss telecom (Swisscom 2019).

In quite a transparent manner, the report informs that Swisscom's overall energy consumption in 2018 has been 548 GWh. Given that Swisscom is the former monopolist and still dominating telecom company in Switzerland, owing a market share of 60% of the Swiss mobile and 67% of the Swiss broadband markets, one might quickly conclude that this number is composed to a large extent of the energy consumption of the WAN, FAN, and RAN networks operated by Swisscom. A closer look at the report, however, reveals that 40.9 GWh were in fact fuel for building heating and 35.5 GWh were the fuel consumption of the company's fleet. While these numbers certainly belong in a company's sustainability reporting, they do not belong to the ICT sector, as they are covered already in other sectors (buildings and transport, respectively).

Moreover, of the remaining 485 GWh electricity consumption of Swisscom, a substantial part is the consumption of its content delivery servers – be it the cloud computing services offered by the company, its TV offerings (including cloud storage for the end customers), or the servers of other content providers such as Netflix possibly collocated on Swisscom's premises. While certainly IT devices, all of these all already accounted for in DC statistics and should not be double counted for networks as well. The report does not reveal which share of Swisscom's electricity consumption pertains to content servers and not to networks; it states, however, that "this server virtualisation requires less IT infrastructure than before and has enabled additional savings of 21.5 GWh in 2018" (Swisscom 2019). If only the savings in this domain were around 5% of the total electricity consumption of the company, its share must certainly be substantial. Finally, Swisscom reused a part of its DC heat: It "supplied 8.4 GWh of thermal energy to the neighbouring areas as district heating" (Swisscom 2019). There are no methods yet established to account for such reuse of DC energy; however, following general LCA principles, it should partially be allocated to its secondary use.

All the distinctions in this example represent potential sources of error for a top-down assessment. Either due to lack of fine-grained data (by contrast to our example), or due to a rushed assessment, each of the processes and devices discussed could be wrongly included within the system boundaries, thus attributing too large an energy consumption to the WAN, FAN, and RAN of Swisscom.

Table 4 summarises these insights on top-down and bottom-up assessments. Their likely inherent biases for overestimates and underestimates, respectively, although not conclusively proven, represent a compelling argument for the existing mismatch highlighted throughout this report.

Table 4 High-level summary of typical flaws of top-down and bottom-up assessments.

	Limitations of top-down assessments	Limitations of bottom-up assessments
Devices and processes	Top-down assessments can easily count devices and processes that should not be counted,	Bottom-up assessments can easily forget devices and processes that should be counted,
System boundaries	thus (unwillingly) defining too wide system boundaries,	thus (unwillingly) defining too narrow system boundaries,
Counting	leading to double counting,	leading to 'no counting',
Bias	and thus contain an inherent bias towards overestimates.	and thus contain an inherent bias towards underestimates.

6 Conclusions

6.1 Summary of the analysis

In this report, we performed a comparative analysis of well-known estimates of the energy intensity and the overall energy consumption of the Internet. Starting from the observation that for two decades already there has been a substantial inconsistency between existing estimates, we discussed the early sources of these discrepancies (i.e., inclusion of DCs into assessments and different years of the analysis), but also why these reasons are only marginally relevant today.

Nevertheless, there is a continuing – and consistent – inconsistency between assessments of the energy intensity of the Internet and estimates of its overall energy consumption. While the two should be connected through the overall Internet traffic, as shown by the identities expressed in Equations (6) and (7), they are in fact quite far apart, as shown by Table 3. The table summarises the insights generated from the in-depth analysis of several well-known academic studies of the last decade, but also of a couple of current industry assessments, in particular of the RAN. In Section 3, we laid out the system boundaries for the WAN, FAN, and RAN, general assessment principles and the equations deployed in the analysis. In the comprehensive analysis of Section 4, we have shown how these principles were instantiated for each individual study, and have transparently addressed any further assumption deployed.

The analysis identified two key sources for the current discrepancies: the inconsistent (and often untransparent) treatment of the access networks (in particular of the RAN and the CPE part of the FAN), and the inherent but opposing biases of top-down and bottom-up assessments. Referring to the first source, bottom-up studies in particular compute the energy intensity of data transfers without clearly stating whether this value represents just the EI of the WAN, of the WAN plus FAN, or the average energy intensity of the entire Internet, which accounts for the WAN as well as for (weighted) access via both FAN and RAN. As for the second source, we have shown evidence, although not definitive proof, that top-down assessments seem to have a bias towards overestimates, while bottom-up assessments might often exhibit the opposite bias, towards underestimates. This is an artefact of top-down vs. bottom-up modelling and not of the aim of the study, i.e., energy or energy intensity (while it is true that the EI is often computed in a bottom-up manner, and the E is often computed top-down, this is not always the case, (Pihkola et al. 2018) being an example for a top-down study that yields energy intensity as result). One way to work towards satisfying the identity from Equation (7) is, in fact, to substitute intensities obtained from a bottom-up process with those computed in a top-down manner; we performed such a substitution between rows 5 and 6 of Table 3.

Taken together, these two factors go a long way in explaining the existing discrepancies today. They also imply that the true value of both the EI and the E of the Internet lies somewhere in between the two, albeit not necessarily in the middle. As a consequence, bottom-up analyses could use well-known top-down studies to validate their results, and vice-versa.

The second factor probably also explains the persisting mismatch of the WAN values. Despite our best efforts, we could find any numbers in the literature or any reasonable assumptions based on this literature to the specific WAN identity from Equation (9). Choosing an intensity value on the very high end of existing estimates – based on (Telekom 2020) and own assumptions – could bring the right-hand side of Equation (9) to 61 TWh and thus somewhat close to the lowest current estimate for E (WAN), which is 110 TWh. Due to this persisting inconsistency, the “levels 2 and 3 or harmonisation” (as defined in Section 5.2) could not be achieved.

6.2 Metrics for the different energy intensities

As discussed in the Introduction, next to the analysis of existing inconsistencies, another important aim of this study was to establish metrics for the energy intensities of different parts of the network and for the Internet as a whole, and principles for their computation.

A substantial part of the literature, especially while the RANs were still relatively negligible in terms of data delivered, did not take them into account and computed one joint value for the energy intensity of the WAN and FAN taken together, $EI_y(WAN + FAN)$. Examples for such studies include, for example, (Baliga et al. 2009; Coroama et al. 2013); a notable early exception is (Schien et al. 2013). Other studies, such as (Aslan et al. 2018), ignored the AN altogether. While the system boundaries are clearly communicated in (Aslan et al. 2018), its title “Electricity Intensity of Internet Data Transmission” can still be misleading, as “the Internet” is usually understood as more than just the WAN (Baliga et al. 2009; Schien et al. 2013; Coroama et al. 2013; 2015; Schien et al. 2015; Malmodin and Lundén 2018; Ryan, Smith, and Wu 2019).

Instead, we argue that the three parts of the network be clearly distinguished. The scope of an assessment must obviously not cover them all, but it should be clearly stated which parts of the network are under study. And any of $EI_y(WAN)$, $EI_y(FAN)$, and $EI_y(RAN)$ that is addressed in the study, should necessarily be devised separately. With this demand for clarity and granularity, we mirror our call from several years ago, when we similarly argued that end devices and data centres should not be included in the computation of the energy intensity of the Internet: “It therefore seems to be more useful and also more transparent to first estimate (a), (b), and (c) separately and adding the partial results up when necessary, as opposed to calculating just the total from the outset. In general, knowing which fraction of the overall energy demand is caused by which part of the supply chain of a final service is key to efficiently allocate investments in energy efficiency. If such results are to be used by decision-makers, they should clearly differentiate between the three components” (Coroama and Hilty 2014).

We further argue that $EI_y(FAN)$ should include not only the street-level part of the AN, but also the CPE (i.e., essentially WiFi routers nowadays). Although the CPE is located within the customers’ premises, it is equipment used to deliver the Internet and thus semantically belonging to the network, not to end-user devices. An additionally argument is that if CPE is not modelled as part of FAN, comparative assessments of $EI_y(FAN)$ and $EI_y(RAN)$ will be skewed to the detriment of RANs, as the lion's share of FAN energy consumption will be ignored. This can have important policy implications, if for example RANs are to become more energy efficient than FANs, as envisioned by (Wu, Ryan, and Smith 2019).

This being said, the energy intensity of WAN and AN together is obviously a meaningful value to compute for various contexts. After devising them separately, studies can thus compute the energy intensity of the Internet accessed over fixed and mobile networks, respectively, by adding the energy intensity of the WAN to that of the corresponding access network, with the rather trivial equations:

$$E_y(fixed\ Internet) = E_y(WAN) + E_y(FAN) \quad (12)$$

$$E_y(mobile\ Internet) = E_y(WAN) + E_y(RAN) \quad (13)$$

A further motivation behind this work was to study whether the average energy intensity of the Internet is a meaningful concept and how it should be computed. After the analysis, this seems both beneficial and feasible. As users are typically unaware of the technical infrastructure behind their Internet connection but might be interested and motivated to reduce their energy and carbon footprints, an estimate of the average is probably valuable. It must, however, account for the general topology of the Internet, as presented in Figure 3, and in particular account for

i) the additive nature of the EIs of the WAN and the AN,
ii) the alternative access via FAN or RAN, and
iii) the (rapidly changing) shares of the two types of access networks.

With Equation (5), we have provided a computation formula that fulfils these requirements:

$$EI_y(Internet) = EI_y(WAN) + (1 - s_{RAN}) * EI_y(FAN) + s_{RAN} * EI_y(RAN) \quad (5)$$

where s_{RAN} represents the share of data accessed via RAN among all Internet-delivered data for that same year.

This need for a compound average energy intensity for the Internet is also one of the reasons to stick for ANs to an energy intensity expressed as 'energy per data' [kWh/GB] rather than 'energy per time' (i.e., power) [W], as suggested by (Coroama et al. 2015) and (Schien et al. 2015). The other reason against such change is that, for RAN at least, the advanced sleep modes built into the 5G technology (Gourhant and Tuffin 2020) will increase their load elasticity, i.e. enable them to adapt their power consumption much more to the actual load, making the 'energy per data' modelling a closer reflection of reality than an 'energy per time' one.

6.3 Current values and short-term developments

To be able to harmonise estimates of the energy intensity and of the energy consumption of the Internet in such a way that they satisfy the identity expressed in Equation (7), and trying to work towards satisfying the identity from Equation (8) as well, in Section 5.2 we had to use some of the more pessimistic assumptions of energy intensity estimates together with some of the most optimistic estimates for the overall consumption of WAN, FAN, and RAN.

These estimates, whose exact values are presented in row 10 of Table 3 and in Section 5.2, can be rounded as follows:

- *EI_{2020} (WAN) = 0.02 kWh/GB,*
- *EI_{2020} (FAN) = 0.07 kWh/GB,*
- *EI_{2020} (RAN) = 0.2 kWh/GB,*
- *E_{2020} (WAN) = 110 TWh,*
- *E_{2020} (FAN) = 130 TWh,*

- *E_{2020} (RAN) = 100 TWh*

The high EI-values were chosen higher than all academic estimates as follows: EI (WAN) and EI (RAN) follow our computation based on (Telekom 2020), EI (FAN) an own assumption as described in Section 5.1; the low E-values are: E (WAN) from (Malmodin and Lundén 2018) very conservatively extrapolated 2015-2020, E (FAN) an average between (Malmodin and Lundén 2018) and (Van Heddeghem et al. 2014), E (RAN) from (Andrae and Edler 2015).

Although they are towards the two distinct ends of their respective spectra in the literature, they were chosen in such a way as to satisfy the identity reflected in Equation (7), together with trying to fulfil to some extent also the level-2 and level-3 identities from Equations (8) and (9)–(11), respectively. Only by selecting such numbers can estimates of the overall energy and the energy intensity of the Internet validate each other. Additionally, top-down and bottom-up modelling (typically used for overall energy and energy intensity estimates, respectively) tend to induce opposite biases, as discussed in Section 5.3.2. It then appears plausible to pick numbers from the end of the spectrum that is opposite to the respective biases, i.e., high values for intensities, low values for overall energy.

With all their caveats and uncertainties, these numbers are thus far more plausible than the numbers at the opposing ends of their respective spectra, i.e., the low energy intensity and the high overall energy numbers. Before new studies appear that will take into consideration the biases of bottom-up and top-down assessments, and mutually validate the energy and the energy intensity results, we suggest using them as fair approximations for the year 2020.

These numbers probably also represent a reasonable basis for extrapolation for a couple of years into the future. The yearly efficiency gains (corresponding to the yearly decreases of the energy intensity) need to be taken into account according to the rather trivial equation:

$$EI_{y+1}(X) = eirf_y(X) * EI_y(X) \tag{14}$$

for all $X \in \{WAN, FAN, RAN\}$ and where $eirf_y(X)$ represents the yearly energy intensity reduction factor for the year Y. There are approximations of the intensity reduction factors in the literature; their caveat, however, is that – as shown above – they typically resulted in too optimistic results in the past. Until updated factors validated against top-down studies emerge, we suggest to use slightly more pessimistic values than those of the literature, as follows:

- $eirf_{2020}(WAN) = 0.7\ (\text{Aslan et al. } 2018)$ + 0.1 = **0.8**
- $eirf_{2020}(FAN) = 0.75\ (\text{Wu } 2021)$ + 0.1 = **0.85**
- $eirf_{2020}(RAN) = 0.7\ (\text{Pihkola et al. } 2018) - 0.8\ (\text{Elisa } 2020)$. Choose the more pessimistic – and directly measured – **0.8**.

Of course, Equation (14) can be used for several years if assuming a constant reduction factor:

$$EI_{y+n}(X) = (eirf_y(X))^n * EI_y(X) \tag{15}$$

This is a reasonable assumption for a couple of years; and equation (15) should not be used for a longer time horizon anyway, given the uncertainties of the current data (see also next section) and the need for mutual validation between energy and energy intensity values in the new studies to come, which will thus hopefully decrease current uncertainties.

7 Outlook: Largest uncertainties and future research

Identifying the important sources for inconsistencies as well as plausible, harmonised data were necessary first steps; using these insights to minimise the discrepancies is the obvious next one. From our results and insights, we can provide valuable guidelines for sounder analyses in future research.

The most obvious implication of our work is that researchers and practitioners should be aware of the existing inconsistencies and control the results of their study at least against Equation (7), or its simpler form from Equation (6). The traffic data from the Cisco studies is a valuable, highly reliable, and thoroughly tested link between EIs and Es. If possible, they should aim higher for a level 2 harmonisation (i.e., Equation (8) jointly with Equation (7)) or even level 3 (Equations (9)–(11) simultaneously). When using just one type of methodology (bottom-up or top-down), the authors should use established results obtained via the other method to validate their results.

Tables 1 and 2, individually and in comparison, provide some interesting high-level insights that have not been discussed so far: One is the relative coherence of the estimates for the RAN. Despite the various parallel generations of radio standards and the greatest traffic growth of all types of networks that occurs within RANs, existing assessments are relatively homogeneous (between 0.1-0.5 kWh/GB for the intensity, 100-220 TWh/year for their overall energy, and a solid connection between the two with a general tendency towards the lower values). In particular the EI (RAN) estimates seem quite solid, as the top-down studies (Pihkola et al. 2018) and (Elisa 2020) reach very similar results to the bottom-up study from (Bieser et al. 2020), while (Telekom 2020) is also close.

Unexpectedly, the largest data inconsistencies are for the WAN: While the intensity estimates are close together, the values for the total energy of the WAN differ by a factor of more than 5. And even the lowest of these estimates is 5 times higher than the bottom-up results, leading to a divergence factor of 5-26 between the two types of studies. This discrepancy also leads to an impossible harmonisation of EI (WAN) and E (WAN) within the currently available studies, which also makes a “level 2” or “level 3” harmonisation, as defined in Section 5.2, impossible. This domain thus requires immediate attention.

This also leads to another phenomenon that needs clarification: bottom-up studies clearly indicate that the EI of the WAN is substantially lower than the EI of both types of ANs (albeit there are only few estimates of the FAN, which in itself is another important topic for future research). Several of the top-down studies, meanwhile, seem to point towards a total energy consumption of the WAN substantially higher than those of the ANs combined. Given Equation (8), these two statements cannot coexist; future research must either raise the EI (WAN) value and/or lower the E (WAN) one.

To address these issues, it would be methodologically interesting to study whether a hybrid method can be developed that unifies to some extent both bottom-up and top-down modelling, while mitigating their individual biases. Related methodological questions that also need to be addressed refer to the challenges in data availability, their age and rapid dynamics, the (cross) validation of models, and the treatment of uncertainty.

Acknowledgements

The author wishes to thank Michael Moser and Roland Brüniger for the opportunity to perform this study, Kurt Roth and Anson Wu for the permission to include their Figures 1 and 6, respectively, Roland Hischier and again Anson Wu for the insightful personal communications which clarified key details of their studies, and in particular Roland Brüniger and Friedemann Mattern for the thoughtful comments and far-reaching feedback to earlier versions of this manuscript.

References

Andrae, Anders, and Tomas Edler. 2015. “On Global Electricity Usage of Communication Technology: Trends to 2030.” *Challenges* 6 (1): 117–57. https://doi.org/10.3390/challe6010117.

Aslan, Joshua, Kieren Mayers, Jonathan G. Koomey, and Chris France. 2018. “Electricity Intensity of Internet Data Transmission: Untangling the Estimates: Electricity Intensity of Data Transmission.” *Journal of Industrial Ecology* 22 (4): 785–98. https://doi.org/10.1111/jiec.12630.

BAFU. 2018. “How Climate-Friendly Is the Swiss Energy Supply?” October 2, 2018. https://www.bafu.admin.ch/bafu/en/home/topics/climate/questions-answers.html#2097182071.

Baliga, J., R. Ayre, K. Hinton, W.V. Sorin, and R.S. Tucker. 2009. “Energy Consumption in Optical IP Networks.” *Journal of Lightwave Technology* 27 (13): 2391–2403. https://doi.org/10.1109/JLT.2008.2010142.

Baliga, J, R W A Ayre, K Hinton, and R S Tucker. 2011. “Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport.” *Proceedings of the IEEE* 99 (1): 149–67. https://doi.org/10.1109/JPROC.2010.2060451.

Belkhir, Lotfi, and Ahmed Elmeligi. 2018. “Assessing ICT Global Emissions Footprint: Trends to 2040 & Recommendations.” *Journal of Cleaner Production* 177 (March): 448–63. https://doi.org/10.1016/j.jclepro.2017.12.239.

BFE. 2020. “Schweizerische Elektrizitätsstatistik 2020.” Bundesamt für Energie BFE. https://www.bfe.admin.ch/bfe/de/home/versorgung/statistik-und-geodaten/energiestatistiken/elektrizitaetsstatistik.exturl.html/aHR0cHM6Ly9wdWJkYi5iZmUuYWRtaW4uY2gvZGUvcHVibGljYX/Rpb24vZG93bmxvYWQvMTA1MjA=.html.

Bieser, Jan, Beatrice Salieri, Roland Hischier, and Lorenz Hilty. 2020. “Next Generation Mobile Networks: Problem or Opportunity for Climate Protection?” University of Zurich. https://plus.empa.ch/images/5G%20climate%20protection_University%20of%20Zurich_Empa.pdf.

Cisco. 2013a. “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017.” https://de.slideshare.net/sbrnyak/white-paper-c11520862-29256455.

———. 2013b. “Cisco Visual Networking Index: Forecast and Methodology, 2012–2017.” https://web.archive.org/web/20140327074533/http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf.

———. 2015. “Cisco Visual Networking Index: Forecast and Methodology, 2014–2019.” https://web.archive.org/web/20150811104530/http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf.

———. 2016. “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020.” https://www.cisco.com/c/dam/m/en_in/innovation/enterprise/assets/mobile-white-paper-c11-520862.pdf.

———. 2018. “Cisco Visual Networking Index: Global IP Traffic Forecast, 2017–2022.” Cisco.

———. 2019. “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022.” Cisco.

———. 2020. “Cisco Annual Internet Report (2018-2023).” Cisco. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

Coffman, Kerry G., and Andrew M. Odlyzko. 2002. “Growth of the Internet.” In *Optical Fiber Telecommunications IV-B*, 17–56. Elsevier. https://doi.org/10.1016/B978-012395173-1/50002-5.

Coroama, Vlad C., and Lorenz M. Hilty. 2014. “Assessing Internet Energy Intensity: A Review of Methods and Results.” *Environmental Impact Assessment Review* 45 (February): 63–68. https://doi.org/10.1016/j.eiar.2013.12.004.

Coroama, Vlad C., Lorenz M. Hilty, Ernst Heiri, and Frank M. Horn. 2013. “The Direct Energy Demand of Internet Data Flows.” *Journal of Industrial Ecology* 17 (5): 680–88. https://doi.org/10.1111/jiec.12048.

Coroamă, Vlad C, and Friedemann Mattern. 2019. “Digital Rebound – Why Digitalization Will Not Redeem Us Our Environmental Sins.” In *Proc. of the 6th Int. Conf. on ICT for Sustainability (ICT4S)*. Lappeenranta, Finland. http://ceur-ws.org/Vol-2382/ICT4S2019_paper_31.pdf.

Coroama, Vlad C., Daniel Schien, Chris Preist, and Lorenz M. Hilty. 2015. “The Energy Intensity of the Internet: Home and Access Networks.” In *ICT Innovations for Sustainability*, edited by Lorenz M. Hilty and Bernard Aebischer, 310:137–55. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-09228-7_8.

Elisa. 2020. “Responsibility Report 2020.” https://static.elisa.com/v2/image/2tqybbhjs47b/4zwayOcLhlPoe9eqRKl84W/Elisa_Responsibility_report_2020.pdf.

Fernández-Fernández, Adriana, Cristina Cervelló-Pastor, and Leonardo Ochoa-Aday. 2017. “Energy Efficiency and Network Performance: A Reality Check in SDN-Based 5G Systems.” *Energies* 10 (12): 2132. https://doi.org/10.3390/en10122132.

Gourhant, Yvon, and Stéphane Tuffin. 2020. “5G: Energy Efficiency ‘by Design.’” February 10, 2020. https://hellofuture.orange.com/en/5g-energy-efficiency-by-design/.

Gupta, Maruti, and Suresh Singh. 2003. “Greening of the Internet.” In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications - SIGCOMM ’03*, 19. Karlsruhe, Germany: ACM Press. https://doi.org/10.1145/863955.863959.

Hintemann, Ralph, and Simon Hinterholzer. 2020. “Videostreaming: Energiebedarf und CO2-Emissionen.” borderstep. https://www.borderstep.de/wp-content/uploads/2020/06/Videostreaming-2020.pdf.

Hischier, Roland. 2021. “AW: Einige Verständnisfragen Zu Eurer ‘next Gen. Mob. Netw.’ Studie Für Swisscom (Direct Impact),” February 26, 2021.

ITU-T. 2020. “L.1331 : Assessment of Mobile Network Energy Efficiency.” SERIES L: ENVIRONMENT AND ICTS, CLIMATE CHANGE, E-WASTE, ENERGY EFFICIENCY; CONSTRUCTION, INSTALLATION AND PROTECTION OF CABLES AND OTHER ELEMENTS OF OUTSIDE PLANT. Telecommunication Standardization Sector of the International Telecommunication Union. https://www.itu.int/rec/T-REC-L.1331-202009-I/en.

Koomey, Jonathan, Huimin Chong, Woonsien Loh, Bruce Nordman, and Michele Blazek. 2004. “Network Electricity Use Associated with Wireless Personal Digital Assistants.” *Journal of Infrastructure Systems* 10 (3): 131–37.

Lanzisera, Steven, Bruce Nordman, and Richard E. Brown. 2012. “Data Network Equipment Energy Use and Savings Potential in Buildings.” *Energy Efficiency* 5 (2): 149–62. https://doi.org/10.1007/s12053-011-9136-4.

Malmodin, Jens, and Dag Lundén. 2018. “The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015.” *Sustainability* 10 (9): 3027. https://doi.org/10.3390/su10093027.

Malmodin, Jens, Dag Lundén, Mikael Nilsson, and Greger Andersson. 2012. “LCA of Data Transmission and IP Core Networks.” In *Proc. of Electronics Goes Green (EGG)*, 6.

NGMN Alliance. 2015. “5G White Paper: Next Generation Mobile Networks.” Next Generation Mobile Networks. https://www.ngmn.org/work-programme/5g-white-paper.html.

Nokia. 2020. “Nokia Confirms 5G as 90 Percent More Energy Efficient.” December 2, 2020. https://www.nokia.com/about-us/news/releases/2020/12/02/nokia-confirms-5g-as-90-percent-more-energy-efficient/.

Pihkola, Hanna, Mikko Hongisto, Olli Apilo, and Mika Lasanen. 2018. “Evaluating the Energy Consumption of Mobile Data Transfer—From Technology Development to Consumer Behaviour and Life Cycle Thinking.” *Sustainability* 10 (7): 2494. https://doi.org/10.3390/su10072494.

Roth, Kurt W., Fred Goldstein, and Jonathan Kleinman. 2002. “Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings - Volume I: Energy Consumption

Baseline." Washington, DC: US Department of Energy, Office of Energy Efficiency and Renewable Energy. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.8518&rep=rep1&type=pdf.

Ryan, Paul, Terence Smith, and Anson Wu. 2019. "Total Energy Model for Connected Devices." Internation Energy Agency. https://www.iea-4e.org/wp-content/uploads/2020/11/A2b_-_EDNA_TEM_Report_V1.0.pdf.

Schien, Daniel, Vlad C. Coroama, Lorenz M. Hilty, and Chris Preist. 2015. "The Energy Intensity of the Internet: Edge and Core Networks." In *ICT Innovations for Sustainability*, edited by Lorenz M. Hilty and Bernard Aebischer, 310:157–70. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-09228-7_9.

Schien, Daniel, Chris Preist, M. Yearworth, and Paul Shabajee. 2012. "Impact of Location on the Energy Footprint of Digital Media." In *2012 IEEE International Symposium on Sustainable Systems and Technology (ISSST)*, 1–6. Boston, MA: IEEE. https://doi.org/10.1109/ISSST.2012.6228017.

Schien, Daniel, Paul Shabajee, Mike Yearworth, and Chris Preist. 2013. "Modeling and Assessing Variability in Energy Consumption During the Use Stage of Online Multimedia Services." *Journal of Industrial Ecology* 17 (6): 800–813. https://doi.org/10.1111/jiec.12065.

Swisscom. 2019. "Annual Report 2018." https://reports.swisscom.ch/download/2018/en/swisscom_geschaeftsbericht_gesamt_2018_en.pdf.

Telekom. 2020. "Deutsche Telekom: The 2020 Financial Year." https://report.telekom.com/annual-report-2020/_assets/downloads/entire-dtag-ar20.pdf.

Van Heddeghem, Ward, Sofie Lambert, Bart Lannoo, Didier Colle, Mario Pickavet, and Piet Demeester. 2014. "Trends in Worldwide ICT Electricity Consumption from 2007 to 2012." *Computer Communications* 50 (September): 64–76. https://doi.org/10.1016/j.comcom.2014.02.008.

Williams, Daniel R., and Yinshan Tang. 2012. "Methodology To Model the Energy and Greenhouse Gas Emissions of Electronic Software Distributions." *Environmental Science & Technology* 46 (2): 1087–95. https://doi.org/10.1021/es202125j.

Wu, Anson. 2021. "Re: EDNA-Study 'Intelligent Efficiency for DCs and WANs': Request for Underlying Data," April 18, 2021.

Wu, Anson, Paul Ryan, and Terence Smith. 2019. "Intelligent Efficiency for Data Centres and Wide Area Networks." International Energy Agency. https://www.iea-4e.org/wp-content/uploads/2021/01/A1b_-_DC_WAN_V1.0.pdf.