

Article

Approximate Networking for Universal Internet Access

Junaid Qadir ¹, Arjuna Sathiseelan ², Umar Bin Farooq ^{1,*}, Muhammad Usama ¹,
Muhammad Ali Imran ³ and Muhammad Shafique ⁴

¹ Department of Electrical Engineering, Information Technology University (ITU)-Punjab, Lahore 54000, Pakistan; junaid.qadir@itu.edu.pk (J.Q.); muhammad.usama@itu.edu.pk (M.U.)

² Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK; arjuna.sathiseelan@cl.cam.ac.uk

³ School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK; Muhammad.Imran@glasgow.ac.uk

⁴ Institute of Computer Engineering, Vienna University of Technology (TU Wien), Wien 1040, Austria; muhammad.shafique@tuwien.ac.at

* Correspondence: mscs16019@itu.edu.pk

Received: 19 October 2017; Accepted: 4 December 2017; Published: 11 December 2017

Abstract: Despite the best efforts of networking researchers and practitioners, an ideal Internet experience is inaccessible to an overwhelming majority of people the world over, mainly due to the lack of cost-efficient ways of provisioning high-performance, global Internet. In this paper, we argue that instead of an exclusive focus on a utopian goal of universally accessible “ideal networking” (in which we have a high throughput and quality of service as well as low latency and congestion), we should consider providing “approximate networking” through the adoption of context-appropriate trade-offs. In this regard, we propose to leverage the advances in the emerging trend of “approximate computing” that rely on relaxing the bounds of precise/exact computing to provide new opportunities for improving the area, power, and performance efficiency of systems by orders of magnitude by embracing output errors in resilient applications. Furthermore, we propose to extend the dimensions of approximate computing towards various knobs available at network layers. Approximate networking can be used to provision “Global Access to the Internet for All” (GAIA) in a pragmatically tiered fashion, in which different users around the world are provided a different context-appropriate (but still contextually functional) Internet experience.

Keywords: universal Internet access; approximate networking; Global Access to the Internet for All (GAIA)

1. Introduction

The new global development agenda “Transforming our world: the 2030 Agenda for Sustainable Development”, composed of 17 Sustainable Development Goals (SDGs), has recently been adopted by the United Nations (UN) General Assembly in 2016. An important pillar of this movement is the need to ensure *social inclusion*, wherewith the society strives to achieve shared prosperity which reaches everyone in the society, including women, people from minorities and the bottom strata of human society. Due to the importance of Internet access—which has now become a key indicator of the potential of economic progress, with impact imprinted on all spheres of human life (personal, societal, political, economical, and educational) in both developing and developed countries—the provisioning of universal Internet access becomes an important stepping stone towards sustainable development the world over.

The fact that Internet access can play a large role in facilitating development motivates the vision of *Global Access to the Internet for All* (GAIA), currently being formally pursued in the Internet Research Task Force (IRTF). While Internet has the capability of fostering development and growth,

this potential is being thwarted by the inability of billions of people to access the Internet. According to recent statistics, almost six billion people do not have high-speed internet, which makes them unable to fully participate in the digital economy [1]. Bringing the Internet to the remaining billions of people left without will democratize knowledge, open up new opportunities, and undoubtedly open up avenues for sustained development.

The overwhelming focus of the Internet research community has been on improving the ideal networking experience by providing increasingly higher throughputs along with lower latencies. However, this focus has led to an Internet design that is very costly, which has precluded the global deployment of the Internet. We see this in wired technologies such as the modern fiber-based broadband high-speed network, which come close to providing “ideal network” performance, have largely been restricted to urban centers and advanced countries, with economical reasons (primarily the high cost of laying fiber) precluding their universal deployment. Similarly, cellular technology—despite its great success—has not been able to ensure GAIA (since it is mainly an urban phenomena that cannot be used to cost effectively serve rural and remote areas [2,3]). Since Internet is over-engineered for many practical applications and needs (i.e., not all applications and users of the Internet require high-fidelity Internet services), we argue that a viable GAIA-enabling approach is the use of “*approximate networking*”, where context-appropriate trade-offs are adopted to deal with different challenges and impairments characterizing a certain region. We can loosely define approximate networks as networks that are close to ideal in terms of quality, nature, and quantity. We proposed the concept of approximate networking previously in [4], where the presentation of the concept focused on the use of simple approximate good-enough services to tame the complexity of the networking infrastructure in a future world afflicted with hard limits due to the exhaustion of natural resources, such as fossil fuels. In this paper, we argue that apart from its clear use in reducing network complexity with the complementary benefits of more sustainable, cheaper services, approximate networking can also be used to satisfy the widely-differing and diverse user requirements by taking *context-appropriate trade-offs* and thereby help in realizing the vision of Global Access to the Internet for All (GAIA).

Our main idea is, for universal Internet provisioning of mobile and Internet services, that it is time to move away from pursuing over-engineered “perfect products” and focus instead on developing appropriate “good enough” solutions. Our “approximate networking” idea can be thought of as the network analog of the emerging computer architectural trend called “*approximate computing*” [5,6], which we discuss next.

1.1. What Is Approximate Computing?

Broadly speaking, approximate computing leverages the capability of many computing systems and applications to tolerate some loss of quality and optimality by trading off “precision” for “efficiency” (where efficiency can be in the terms of increased performance or reduced costs in terms of energy consumed or system cost/area). Approximate computing systems are able to optimize the efficiency of systems by relaxing the commonly applied notion of exact (numerical or Boolean) equivalence between the specification and implementation at multiple layers of the hardware and software stacks (see Figure 1 for a depiction). The use of approximate computing is motivated by the following factors: (1) modern “big data” applications are based on noisy real-world data; (2) many computing applications (e.g., recommendation and web search) do not have a single golden answer; (3) the perceptual limitations of users mean that some approximations may not even be noticed; and (4) many applications exist—e.g., images, video, and sound—where minor errors and approximations can be tolerated by different users. Some recent case studies for applying approximate computing to video processing [7], signal processing [8] and communication systems [9] have shown early feasibility. The research in the field of approximate computing has been led by seminal contributions from the industry players such as Intel [10], IBM [11,12], and Microsoft [13], as well as several research groups from academia [5,6,14–16].

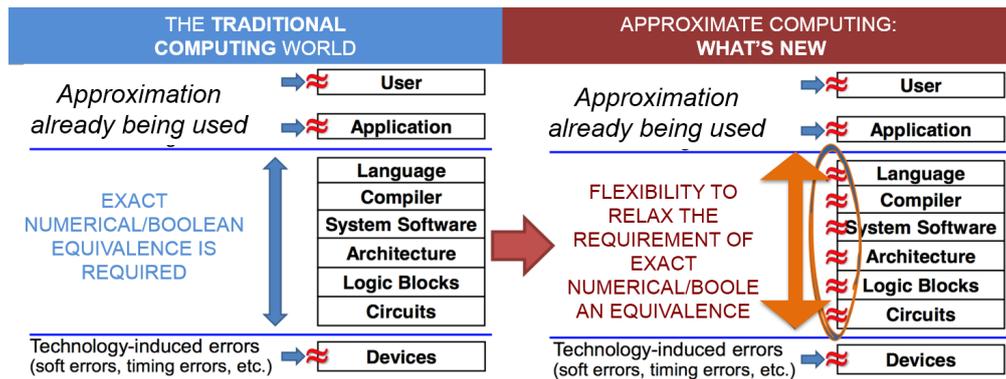


Figure 1. What’s new about approximate computing? (Adapted from [6,17]).

1.2. What Is Approximate Networking?

With approximate networking, we intend to seamlessly integrate the concepts from approximate computing along with traditional mechanisms for approximations in networking, in terms of approximations adopted by networking protocols and algorithms. The concept of “approximate networking” is necessary since experience has shown us that universal commissioning of “ideal networks”, which have extremely high capacity, bandwidth, and reliability, in addition to extremely low or negligible delays, errors, and congestion, is non-practical. It is important to emphasize that approximate networking is not a single standalone technique nor is it the first time that approximation has been proposed in networking. Indeed, a number of existing networking techniques already utilize approximation and are best effort. Our idea of “approximate networking” generalizes these classical ideas and importantly supplements these ideas with recent developments in the field of *cross-layered approximate computing* (particularly at the hardware level) to facilitate the design of future energy-efficient and optimized network infrastructure as well as algorithms and protocols. We aim to enable end to end approximation principles/frameworks engaging hardware/software approximation as well as network layer approximations for systematic approximate networking. As this point in time, there has only been rudimentary work done in efficiently combining low-level, approximate computing modules to construct larger high-level modules and architectural components. Approximate computing has been deployed for a large number of applications including image processing, signal processing, machine learning, scientific computing, financial analysis, database search, and distributed computing, *however, its extension to the field of networking is practically non-existent at this point in time, with only some recent preliminary works as exceptions [18,19]*. We anticipate that these hardware-focused approximate computing advances will percolate into the field of networking and in the future, there will be an increased interest in synergistic approximation management at different layers of the hardware and software stacks in networking.

1.3. Why Adopt Approximate Networking?

1.3.1. Affordable Universal Internet (GAIA)

The right of affordable access to broadband Internet is enshrined in the 2015 sustainable development goals of the United Nations. The International Telecommunication Union (ITU) broadband “Goal 20-20” initiative aims at an optimistic target of universal broadband Internet speeds of 20 Mbps for \$20 a month, accessible to everyone in the world by 2020 (Source: Alliance for Affordable Internet (A4AI) Report, 2014). Such an approach, which aims at providing an “ideal networking” experience universally, has historically always failed (due to various socioeconomical and technical issues). An important reason is that most modern technologies (such as 3G/ 4G Long-Term Evolution (LTE) and the planned 5G) are urban focused as rural systems (being sparsely populated by definition) do not thus hold much business potential for mobile carriers [2,3]. The Internet is also

large unaffordable when we consider that on average the mobile broadband price and the fixed-line broadband prices are 12 and 40% of the average person’s monthly income, where women and rural populations are hit the hardest [20]. Approximate networking is a particularly appealing option to reach out to the offline human population by providing an affordable, contextually “good enough” service.

1.3.2. Diversity of User & Application Profiles

The Internet’s “digital divide” is not a binary divide. There is a spectrum of connectivity options and digital capabilities accessible to people around the world (see Figure 2). In some places, ultra high-speed broadband connections are available, while there are hosts of places where there is no connectivity at all, however, most places lie somewhere in between. User and application profiles and requirement vary greatly: at one extreme, we have applications that require extremely high throughput (e.g., video on demand) and low latency (e.g., tactile Internet); on the other extreme, we have applications that have minimal throughput and latency requirements (e.g., smart meters, which report back low-volume data relatively infrequently). Users can also have vastly different service requirements and financial strength. In the face of such great diversity, the approximate networking framework can avoid the difficulties of single-size-fits-all networking solutions, furthermore, these diversities can be exploited and applications and users provided services and resources commensurate to their requirements.

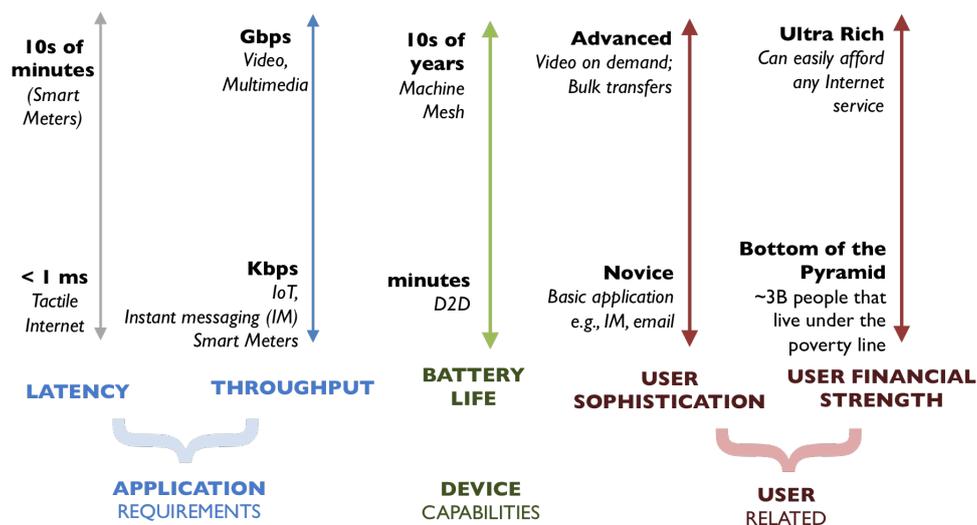


Figure 2. Ensuring Global Access to the Internet for All (GAIA) requires provisioning ‘good enough’ quality of service (QoS) that accommodates the diversity of applications requirements, device capabilities, user profile and requirements.

1.3.3. The Pareto Principle (80–20 Law): The Power of “Good Enough”

To help manage the approximate networking trade-offs, it is instructive to remember the *Pareto principle*, alternatively called the 80–20 rule [21], which states roughly that 20% of the factors result in 80% of the overall effect. This principle has big implications for approximate networking since this allows us to provide adequate fidelity to ideal networking by only focusing on the most important 20% of the effects. The key challenge in approximate networking then becomes the task of separating the all-important, essential, non-trivial factors from the trivial factors, which may be omitted or approximated. In this regard, we can leverage previous human-computer-interaction (HCI) research that has shown that human quality of service (QoS) perception can be flawed (e.g., relatively fast service may be judged to be unacceptable if the service is not predictable, visually appealing and

reliable [22]) in choosing the precise approximate networking trade-off to adopt such that the users perceive the least inconvenience.

1.3.4. Need of Energy Efficiency

It has been reported that information and communication technology (ICT) is one of the biggest consumer of the world's electrical energy, using up to 5% of the overall energy (2012 statistics) [23]. The urgency of delivering on the front of energy efficiency is reinforced by the impending decline of non-renewable energy resources along with the concomitant increase in ICT demand. The approximate networking trend can augment the hardware-focused "approximate computing" trend in managing the brewing energy crisis through the ingenious use of approximation. In particular, approximate networking can help generalize the performance and efficiency improvements offered by approximate computing, which have largely been limited to local speedups on a single device, to broader network settings. Optimizing communication/networking cost is important since these costs can be significant (e.g., on mobile phones, the Wi-Fi and cellular radios require, on average, an order of magnitude more power than the CPU or memory [18]).

1.4. Contributions of This Paper

The main contribution of this paper is to investigate the extension of the concept of approximate computing to the field of networking. We propose approximate networking as an overarching, cross-layered framework that encompasses classical approximation techniques, as well as recently-developed techniques in the field of approximate computing, to implement context-appropriate networking trade-offs that are necessary for the aims of "Global Access to the Internet for All" (GAIA). In order to facilitate these trade-offs, apart from the classical approximation techniques adopted in networking (in the areas of software/hardware, algorithms, protocols, and architecture), approximate networking will also leverage the advances in the fast-emerging field of approximate computing as an extra degree-of-freedom for finer-grained tradeoff optimization. We also propose approximate networking as an overarching framework for systematically thinking about networking trade-offs that must be adopted for ensuring GAIA. Furthermore, we also present an application of approximate networking in 5G with a case for low income and rural regions.

This paper is organized as follows. Section 2 describes approximate networking technologies. In Section 3, we present context-appropriate approximate trade-offs for networking. We describe a case study for approximate 5G networks in rural and low income areas in Section 4. We present discussion issues for approximate networking in Section 5 and conclude in Section 6.

2. Approximate Networking Technologies

There exist some error-tolerant networking applications that are constrained by the needs for energy efficiency and real-time packet delivery. By using approximate computing, these applications can be deconstrained by the relaxation of the integrity requirements for the approximate data, thereby allowing these applications to communicate more efficiently (i.e., these applications can transmit faster, over a longer range, and using less power; see [19] and references therein).

2.1. Approximate Networking: Old Wine in a New Bottle?

We do not claim that approximate networking is a new invention. Taking a broad view, we see that many established existing technologies are in fact examples of approximate networking. Indeed, delay-tolerant networking (DTN), information-centric networking (ICN), the concept of lowest-common denomination networking (LCD-NET) [24], the use of caching and opportunistic communication are all approximate networking solutions in this classical sense. The User Datagram Protocol (UDP) protocol approximates the transport service provided by Transmission Control Protocol (TCP), but it trades reliability for performance gains. We can also have approximate networks that provide only a tenuous approximation of the quality, nature, or quantity of the Internet: e.g., services

that rely on data mules (e.g., DakNet [25]) are only infrequently connected to the Internet; there are also services (such as Outernet [26] and Internet in a Box [27]) that approximate the Internet experience without actually connecting to the Internet.

The novel aspect of approximate networking is that it can leverage advances in hardware-based approximate computing developments. In particular, researchers can utilize cross-layer approximation across the computing stack (refer to Figure 1), where the stack contains, in addition to the networking, programming languages, compilers, operating systems/databases, and hardware architectures. For instance, approximate programming languages (such as EnerJ, EnerC, etc.) can be used to specify the critical and the non-critical aspects of computation (e.g., EnerJ is a general purpose programming language built as an extension to Java that exposes hardware faults in a safe, principled manner; results have shown large amounts of energy savings with slight sacrifices to QoS). The same general idea applies to networking devices (such as switches and routers) where not all operations are equivalent—some aspects have to be precise, while others can be approximate. Approximate computing can be used to differentially specify the critical parts of the program and the less critical parts depending upon their inherent resilience properties (see resilience characterization in [14]); in this manner, approximate networking can have greater support from the hardware in implementing context-appropriate approximations.

A representative Taxonomy of approximate networking concepts is shown in Figure 3. A summary of some sample works related to the various facets of approximate networking taxonomy highlighted in the taxonomy presented in is also presented in Table 1 for illustrative purposes.

2.2. Approximate Networking Hardware

With physical limits beginning to stall the exponential growth of electronics, due to Moore's law and the Dennard's scaling, it seems to be the case that future hardware devices, and by extension future network switches and routers, will use approximate computing in one way or another. In terms of hardware support, networking in most end user devices (such as smartphones, tablets, laptops, and embedded devices such as smart TV) is implemented using application-specific integrated circuits (ASICs). These end-nodes are general purpose computing devices that have network-interface-cards (NICs) built in for connecting to networks. In contrast, we also have specialized networking devices (such as routers, switches, hubs, and firewalls) that function as the building blocks of networks; such devices have, in addition to NICs, also switching fabrics and use various techniques such as memory-based computing through ternary-content-addressable-memories (TCAMs). Networking devices also utilize a number of components: memory hierarchy of static RAM (SRAM), dynamic RAM (DRAM), non-volatile RAM (NVRAM), hardware parallelism (using ternary content addressable memory (TCAM)), multiplexers and demultiplexers, and counters, which are all amenable to approximate computing (as explained in [28]).

Two principal components of network hardware are (1) the implementation of packet processing logic through combinational logic, and (2) memory based components. In *implementing combinational logic*, there are two approximating options. Firstly, the elementary building blocks of approximate adder and multiplier, which are typically implemented in arithmetic logic units and are used in counters, already have an approximate computing implementation that return impressive gains [28].

Such adders and counters can be used in implementing the set of counters maintained by the simple network management protocol (SNMP) that is implemented by all routers today. Secondly, the complex logic in a larger circuit can be replaced with an approximate, optimized, pseudo-equivalent circuit. When *implementing memory in networking devices*, approximation can be implemented at the level of SRAM, DRAM, and TCAMs. Approximate caches have been built that avoid cache misses by using approximation based on temporal and spatial correlation present in data stored in memory (this can be used to avoid expensive power-hungry access of DRAM). Furthermore, the power consumed in accessing DRAM can be reduced through refresh rate control using which the error/power tradeoff can be managed.

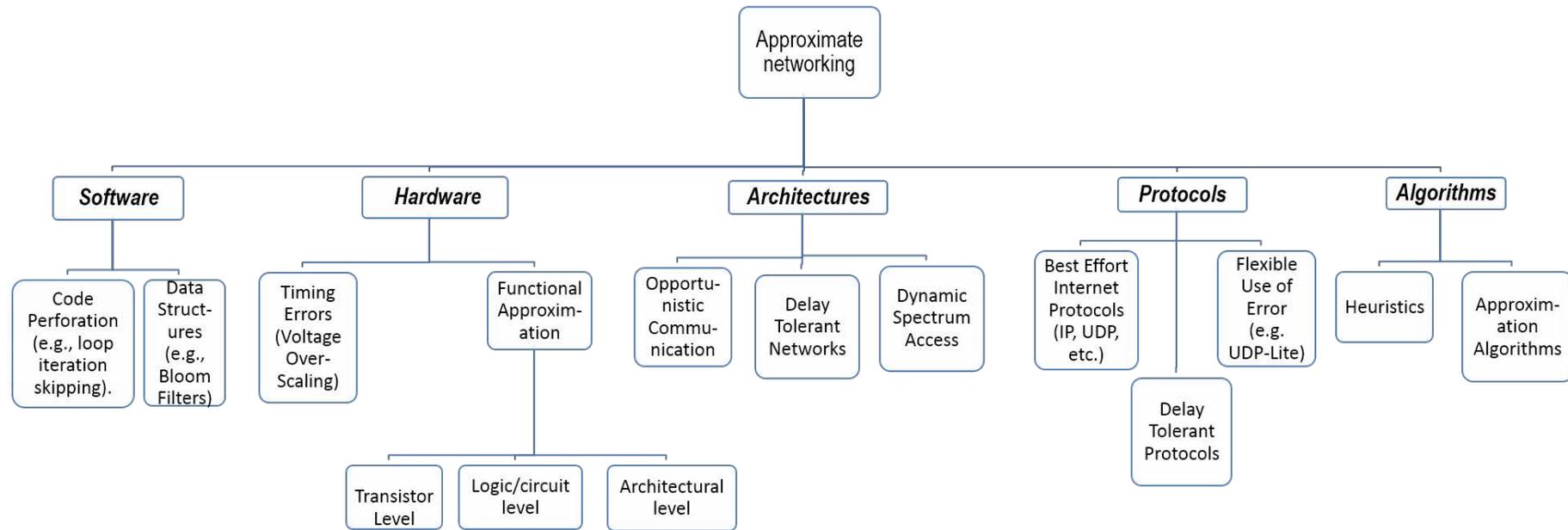


Figure 3. An (approximate) taxonomy of approximate networking concepts.

Table 1. A summary of sample work related the categories of approximate networking taxonomy.

Reference	Task	Brief Summary	How Approximation Is Used to Increase Performance
<i>Software</i>			
Sampson et al. [29]	Approximation-based Compiler Framework	Introduces a compiler framework for practical approximate computing.	The approximation compiler framework substantially improves the end-to-end performance with little quality degradation.
Sampson et al. [30]	Language of Approximate Computing	Proposes a programming language model (EnerJ) for approximate computing	An approximate data type for low power consumption devices is proposed.
Esmailzadeh et al. [31]	Programmable Accelerator	Proposes a new class of neural programmable unit (NPU) accelerator that uses approximate computing to get better performance and energy efficiency.	A general purpose approximate computing NPU saves 3× more energy and speeds up the process by 2.3×.
Jokela et al. [32]	Multicast Forwarding	LIPSIN incorporate Bloom filter properties for large scale topic based Publish-Subscribe systems.	Bloom filters reduce the forwarding table size, and increase multicast forwarding efficiency, at the cost of small false positives.

Table 1. Cont.

Reference	Task	Brief Summary	How Approximation Is Used to Increase Performance
<i>Hardware</i>			
Talla et al. [33]	Network Hardware Approximation	Power over Wi-Fi delivers power to low-power sensors and network devices.	A new approximate-computing-enabled energy harvesting design that provides far-field power delivery to Wi-Fi enabled is provided.
Jouppi et al. [34]	Custom Hardware Chip for Machine Learning (ML)	Google's Tensor Processing Unit (TPU) provides tolerance for reduced computational precision in ML programs.	Google is using TPUs in datacenters since 2016, thereby achieving better-optimized ML performance per watt.
Esmailzadeh et al. [35]	Neural Processing Unit (NPU)	NPU's software and hardware design is presented.	With learning, code transfer, and approximate computing enabled instruction set architecture, 2× performance and energy-saving improvement is achieved.
Mazahir et al. [36]	Consolidated Error Correction (CEC)	CEC: Correction is applied to errors accumulated from several additions. Low latency generic accuracy configurable hardware combined with error recovery circuit for applications requiring high accuracy.	CEC is used in Approximate Hardware Accelerators for area saving and speed enhancement. Adder provides a better accuracy, area and speed tradeoff as compared to previous counterparts.
Shafique et al. [37]	Low Latency Adder	Low latency generic accuracy configurable hardware combined with error recovery circuit for applications requiring high accuracy.	Adder provides a better accuracy, area and speed tradeoff as compared to previous counterparts.
Mishra et al. [10]	Approximate Computing Toolkit	Intel's approximate computing (iACT) toolkit comprises a run-time compiler and a simulated hardware testbed.	Intel's iACT is a approximate computing toolkit designed for promoting industry and academia research.
<i>Architectures</i>			
Baker et al. [38]	Opportunistic Communication for Delay Tolerant Networks	A routing platform for delay-tolerant social networks.	Packets from source to destination reaches in cooperative communication fashion.
Sermpezis et al. [39]	Opportunistic Communication	Describes how content-centric applications perform in opportunistic scenarios.	QoS of content-centric networks is improved by approximating delays, content popularity and availability.
Rehman et al. [16]	Architectural Exploration of Approximate Multipliers	Using variants of approximate/accurate adders/multipliers and approximate LSBs for exploring apace of approximate multipliers.	Open Source Library for further Research and Development of approximate Computing at higher abstraction level of HW/SW stack.
Esmailzadeh et al. [13]	Architectural Support for Approximate Programming	A new ISA extension which provides approximate operations and storage, due ti which energy is saved at the cost of small degradation in accuracy.	When proposed scheme is tested with several applications up to 43% energy is saved.
<i>Protocols</i>			
Larzon et al. [40]	Flexible Best Effort Protocol	Proposes a UDP variant called UDP-Lite that uses partial checksums.	UDP-Lite allows for error tolerance and this approximation can significantly improve the network throughput.
Shelby et al. [41]	Best Effort Protocol	Proposes a best-effort application layer protocol for constrained devices.	Constraint application protocol uses UDP and UDP-Lite as the underlying approximation transport layer protocol to facilitate error tolerance.
Ransford et al. [18]	Cross Layer Approximation Protocol	Selective Approximate Protocol (SAP) enables network applications to receive potentially damaged network data.	Approximation introduced in SAP increased the throughput and reduce the retransmission rate of wireless communication networks.
<i>Algorithms</i>			
Krishnan et al. [42]	Incremental Approximation Algorithm	An incremental approximate computing algorithm (IncApprox) is presented for network and Twitter data analytics.	IncApprox combines incremental and approximate computing paradigms to achieve 2.1× the throughput achieved by either.
Gupta et al. [43]	Approximation Algorithms	Approximation algorithms for network design are presented.	Different emerging solutions for minimum spanning tree problem using different approximation assumption are discussed.
Gandhi et al. [44]	Approximation Algorithms	A one-to-all approximate wireless broadcasting algorithm is presented.	An approximate solution is proposed for an NP-Complete optimization problem with routing, scheduling and QoS applications.

Approximate computing can be done through circuit-level techniques such as *voltage over-scaling* using which the voltage is deliberately reduced for improving power efficiency of circuits. Such a technique can be used in approximate networking to build hybrid memories with power efficient unreliable memory arrays that can drastically improve power efficiency through aggressive voltage over-scaling. Approximate networking can also use *functional approximation* at various levels (particularly, at the architectural, circuit-level, and at the transistor level). This can include the use of neural networks for learning a simplified approximation of code [28].

We note here that approximate computing technology is not limited to general-purpose central processing unit (CPU) architecture only. Companies are already resorting to approximate computing to obtain energy and cost optimized application-specific integrated circuits (ASICs), e.g., Google has made a custom ASIC named the Tensor Processing Unit to run machine learning-based tasks at scale in their data centers, while requiring fewer transistor per operation. Similarly, IBM exploited the error resilience of Deep Neural Networks to loss of numerical precision for better area and power efficiency systems [12]. We believe it is only a matter of time before these ideas would find way to network ASICs and that more research is needed on how these technologies will interplay with the network-layer stack.

2.3. Approximate Networking Software: Algorithms

Implementing approximate networking can be useful in many scenarios where aspects of a system, such as ease of use, functionality, and robustness, may be more important than performance alone. The idea of approximation is an oft-used tool in networking algorithms [45]. Approximate networking algorithms [46] (also called heuristics) are often required in networking to tackle discrete optimization problems (many of which are NP-hard, and thus there are no efficient algorithms to find optimal solutions). Such algorithms have been widely used in scheduling, routing, and QoS problems in networking.

In the book *Network Algorithmics* [45], Varghese has distilled 15 important high-performance principles that have wide applicability in networks, many of which are based on efficient approximations and trade-offs. In particular, Principle 3 talks of trading certainty for time (P3a) and trading accuracy for time (P3b). With P3a, randomized strategies are used when deterministic algorithms may be too slow (e.g., the use of randomization in deciding when to transmit after a collision in millions of Ethernets worldwide). With P3b, the idea is to relax accuracy requirements for speed (e.g., by using lossy compression, approximation thresholds, approximate sorting, and approximate set membership query algorithms such as the Bloom filter and Cuckoo filter).

In particular, Bloom filters, which is a high-speed, approximate set membership query algorithm whose tests that can return false positives (but never false negatives), have been extensively applied in networking in a wide variety of settings [47]. Bloom filters are important since a large number of applications require fast matching of arbitrary identifiers to values. Since it is common to have millions, or even billions of entries, we require efficient and scalable methods for storing, updating, and querying tables. Although Bloom filters can return a few false positives, their real utility is in alleviating the scaling problems that a number of diverse network applications face. Broder and Mitzenmacher have articulated the *Bloom filter principle* [47]: “whenever a list or set is used, and space is at a premium, consider using a Bloom filter if the effect of false positives can be mitigated”. Bloom filters have been used for diverse ends in networking and distributed systems such as caching, peer-to-peer systems, routing and forward, and monitoring and measurement [47].

Code perforation is another software-based approximate computing technique that can be used to automatically identify error-resilient portions within a code that can be skipped while keeping the error within a predefined range [6].

2.4. Approximate Networking Protocol Stack

Various approximations have already been adopted by transport-layer protocols to handle multimedia traffic over the best-effort Internet. Datagram Congestion Control Protocol (DCCP) can provide congestion control services without the reliability overhead of TCP. UDP is approximate in the sense that it trades off reliability for efficiency and rapidness. UDP-Lite [40] is a connectionless protocol that approximates the performance of UDP but it uses *partial checksums*, thus, while UDP always discards packets that fail checksums, UDP-Lite relaxes the integrity checks to deliver timely but potentially partially erroneous packets to multimedia applications. UDP-Lite is designed for accelerating error-tolerant, real-time multimedia streaming applications that can tolerate slight errors. While UDP-Lite can provide a significant throughput improvement by relaying partially corrupted packets to the application layer [18], care must be adopted to ensure that (1) there does not exist a conflict between the link-layer checksum (e.g., 802.11 FCS) and the UDP-Lite checksum, which can result in the link layer refusing to carry damaged data, and (2) the additional throughput is not in carrying “useless data” for the application—this may require the use of erasure-error protection at the application layer.

Approximate packet processing circuits that return imprecise answers will have the benefit that they will be much smaller than those today. They would consume less power, and many more of them could fit on a single chip, greatly increasing the number of packets it could process at once. However, a critical question is how these imprecise packet processing affect the performance of higher-layer networking stack. Since the approximate networking technology is inchoate, we cannot answer with certainty how the low-level hardware-based approximate computing developments will affect the performance of the network upper-layer stack. However, we can definitely foresee the need to approximation semantics at the upper layers of the networking stack so that hardware-level approximate computing features can be efficiently utilized. Otherwise, the potential benefits of approximation will be frittered away (as the network may attempt a reliable and precise transfer of approximate data). Some approximate networking semantics naturally exists as part of the TCP/IP stack (e.g., through the best effort service provided by the UDP and IP protocols). There have been many disparate efforts at the transport layer which can be categorized as “approximate networking” efforts. Standard reliable transport protocols like TCP work well for file transfer applications but are too unwieldy for multimedia applications that prioritize speed of transfer and have some error tolerance. To ensure that we have appropriate communication support for approximate data, some essential mechanisms that should be supported include [19]:

1. *Optional multi-layer integrity check support*: currently, the different network layers perform redundant checksums (e.g., TCP over Wi-Fi uses three separate checksums, namely, the TCP layer, the IP layer, and the link layer). In an approximate networking context, it is useful to permit some errors in approximate payloads.
2. *Partial integrity checking for critical data* (e.g., addresses and ports must be precise): it is typical in networking to discard erroneous packets that have been received with checksum errors. Both TCP and UDP discard erroneous packets (TCP also asks for a retransmission to ensure reliability). However, in the spirit of approximation, partial errors in non-critical data can be tolerated. UDP-Lite [40] is an example transport protocol that performs partial integrity checking through the use of a configurable checksum (which specifies how many bits are protected by checksum).
3. *Application-provided approximation specification*, and switching between these specifications, for a given socket at the level of different layers. As an example work, Selective Approximate Protocol (SAP) [18] allows applications to coordinate with multiple networks layers to accept potentially damaged data. The authors of SAP, which is built over UDP-Lite, have reported a 30% speedup for an error-tolerant file transfer application over Wi-Fi.

3. Context-Appropriate Approximate Networking Trade-Offs

3.1. Trade-Offs in Networking

In a complex system such as the Internet and the networking TCP/IP ecosystem, in which multiple subsystems work in silos isolated from each other, there is a danger that improvement/simplification of one subsystem may increase the complexity of some of the other subsystems and deteriorate their performance. We propose that approximate networking ecosystem as a holistic systems-oriented framework that can offer context-appropriate management of the various trade-offs involved in networking in terms of performance, cost, and complexity. Some of these trade-offs are discussed below:

- *Fidelity versus affordability/convenience*: a lot of research has shown that customers are willing to sacrifice considerable fidelity for a more convenient and accessible service [48]. The notion of fidelity matches with the QoS/ Quality of Experience (QoE) concept. Convenience subsumes concepts such as the cost, accessibility/availability, and simplicity of the service. The fact that users are willing to tradeoff fidelity for convenience and affordability is an extremely important insight for our topic.
- *Latency versus throughput*: it is well known in literature that throughput-optimal solutions can compromise performance in terms of delay [49]. The *Sneakernet* concept, long known in networking folklore (“Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway.”—Andrew Tanenbaum, 1981) is the embodiment of the latency-throughput trade-off. In a similar vein, DTN routing protocols also tradeoff latency for throughput and connectivity (i.e., DTN Bundles can achieve the same throughput as IP protocols but with longer latency).
- *Throughput versus coverage/reliability*: in wireless networks, there is a tradeoff between the throughput and the coverage (and the reliability) of a transmission, i.e., for higher-rate transmissions, the coverage area is typically smaller and the bit error rate higher. The idea of approximate networking can be used to provide context-appropriate QoS to 5G users [3], by provisioning higher rates to users and applications where feasible and desired, while still allowing everyone access to basic connectivity (allowing users who are currently offline to come online).
- *Coverage versus consumed power*: in wireless networks, the coverage of a transmission is directly proportional to the transmission power. Since nodes do not need to communicate at all times, researchers have proposed putting to sleep parts of the infrastructure, such as the base transceiver station (BTS) of cellular systems, to save on energy costs.
- *Other trade-offs*: many innovative solutions are able to improve performance by inventing a new tradeoff. For example, Vulimiri et al. discovered that an interesting way to reduce latency is to tradeoff some additional capacity or redundancy (i.e., the authors showed that latency can be reduced by initiating redundant operations across diverse resources and using the first complete response) [50]. Future approximate networking solutions can derive much utility by focusing on discovering new ways of developing context-appropriate new trade-offs.

3.2. Leveraging Approximation

When dealing with context-appropriate trade-offs that come with approximate networking, an important concept is that of *Pareto optimality*, which refers to a state of resource allocation in which it is not possible to make any one individual better off without making at least one individual worse off. The set of values that are each Pareto optimal together constitute the *Pareto frontier*. The main benefit of using approximate networking is that it can leverage the degree-of-freedom of leveraging errors to shift the Pareto frontier (see the tradeoff being depicted in Figure 4) such that performance and cost may be improved simultaneously at the cost of some inaccuracy (which is designed to be under the threshold necessary for acceptable QoS).

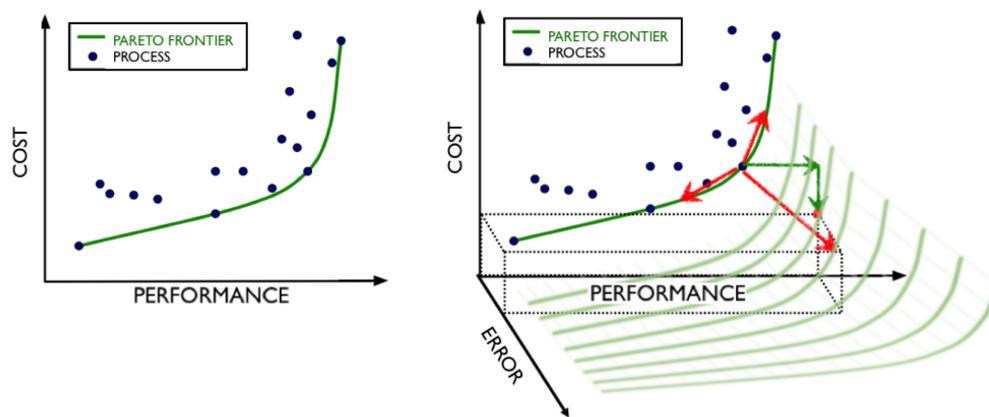


Figure 4. Leveraging the extra degree of freedom of exploiting errors can improve performance while reducing cost.

3.3. How Can We Visualize the Trade-Offs?

An interesting approach to understanding trade-offs is to use visualization techniques. In approximate networking, the task for optimizing for one explicit parameter is easier than optimizing for multiple optimization variables (such as throughput, delay, and energy). The main problem arises when the various objectives are conflicting and have to compete for dominance, e.g., it is impossible to jointly minimize both bit error rate (BER) and transmit power simultaneously [51]. One approach to solving such a problem is to look for a solution on the so called ‘Pareto Frontier’ that defines the set of input parameters that define non-dominated solutions in any dimension. The use of a *tradeoff curve* [52] can be used to visualize bi-objective problems. The problem of visualizing high-dimensional trade-offs is more challenging. One approach that has been proposed is to utilize Pareto front, which defines the set of values that are each Pareto optimal. When considering multiple objectives, it is often useful to consider the *Pareto frontier* or the *Pareto front* that comprises of the set of choices that are Pareto efficient instead of considering the full range of every parameter. A *radar chart* is another planar visualization technique that can be useful for visualizing multivariate network trade-offs. For example, Chang et al. [53] have used radar chart for visualizing QoE in multiple QoS dimensions.

3.4. Open Questions

While we have described the main trade-offs involved in approximate networking, the all-important question still remains to be addressed: *how can we effectively manage these approximate networking trade-offs?* This bigger question is very much an open issue requiring more research. Some important more specific follow-up research questions regarding trade-offs are:

1. Which approximation to apply where in the hardware/software stack, and to which degree, such that the end-to-end QoS requirements are fulfilled?
2. How to estimate end-to-end error degradation due to approximations?
3. How do we quantify when our approximation is working and when it is not?
4. How to measure success in managing the service quality/accessibility tradeoff?
5. How do we measure the cost of approximation in terms of performance degradation?
6. How to dynamically control the approximation trade-offs according to the network condition?
7. Can the degradation models for approximation errors and channel errors be consolidated?

In answering these questions, we can leverage the copious literature produced in the field of multi-objective optimization for some tradeoff-related questions, however, answering other questions related to approximate networking would require new and original future research investigations. It is also important to point out that computing the right tradeoff requires the incorporation of factors such

as the (1) the technology-focused concept of *quality of service* (QoS), and (2) the user-focused concept of *quality of experience* (QoE) [22]. While most QoS works have focused on objective measurable metrics such as delay, jitter, throughput, and packet loss, both the objective and the subjective quality measures are needed to provide a holistic multi-dimensional assessment. The subjective quality measures can incorporate factors such as the subjective user preferences; the subjective and objective user perception of the QoS [22]; and the objective application/ service's QoS utility.

An open challenge in defining and managing context-appropriate trade-offs is to determine the right granularity of the conception of context-appropriate trade-offs, which supports network-level efficiency as well as the user-level and application-level needs. Since the network will likely be used by many users and applications, there is also the problem that instead of a single approximate networking trade-off, there will be multiple interacting approximate networking trade-offs that have to be harmonized. In such an environment, a holistic or systems-thinking based approach will be useful to model the interaction between the various approximate networking point solutions.

Another challenge is to support applications with approximate networking that are intolerant of even small transmission errors. For instance, implementing encryption in an approximate setting is going to be challenging since encryption transforms even error-resilient applications (such as images, audio, video) to be error intolerant [18]. Other error-intolerant applications may also exist and more research is needed to investigate how such applications can be supported with approximate networking.

4. Case Study: Approximate 5G Networks in Rural/Low Income Areas

In this section, we investigate how approximate networking may be applied in the concrete setting focused on democratizing 5G wireless services universally, particularly in rural and low-income poor regions that are mostly found in developing countries. According to estimates, at least two billion people living mostly in *rural and low-income* regions experience a complete lack of wireless network coverage. Figure 5 shows the cost of mobile broadband for developing and least developed regions is a big percentage of gross national income (GNI) per capita. As a result, a considerable population of these regions cannot afford a mere 100 MB mobile data per month (Figure 6).

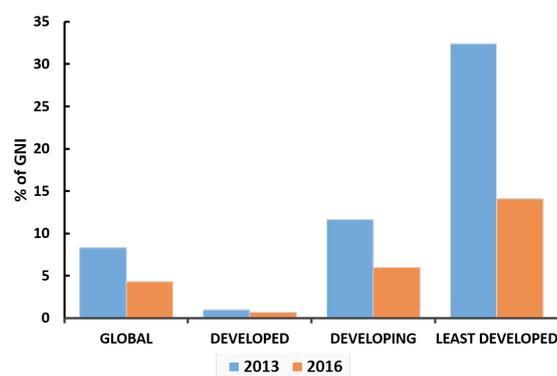


Figure 5. Mobile broadband prices as a percentage of Gross National Income (GNI) per capita for different regions [54].

Without the enabling technology of the Internet and communication technology, these economically-backward areas suffer from a vicious cycle that pushes them even more backwards. Approximate networking services that can provide universal “good-enough” services can help bring the benefits of the Internet and communication to such disadvantaged communities. In contrast to the plethora of 5G research projects aiming for high performance, the coverage of rural and low-income areas in future 5G networks despite its great importance has received relatively little attention although that has begun to change with few recent works [3,55,56]. For rural areas, the main

challenge is cost-effective coverage rather than the urban focus on high data rates and low latency. The high-performance urban 5G requirements dictate the need of a complex and expensive network comprising inter alia high-capacity fronthaul and backhaul networks, dense heterogenous networks, and large datacenters. Supporting such an architecture requires significant revenues, which will be hard to obtain in rural environments that have very few inhabitants compared to urban environments. The weak business cases for telecommunication operators in serving rural areas can be observed from a previous study [57], which showed that the approximate revenue potential for operators in the least-densely populated areas is merely 262 USD per square mile of service, compared to 248,000 USD per square mile of service for highly-dense urban populations.

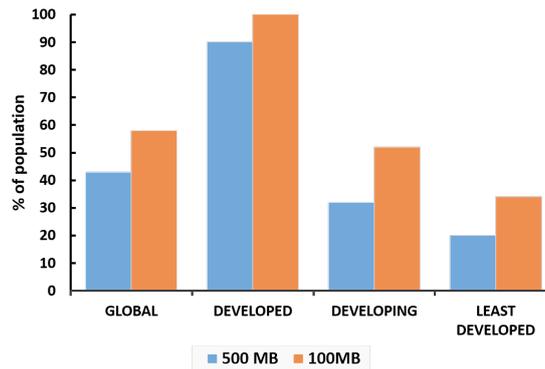


Figure 6. Percentage of population who can afford 500 MB and 100 MB pre-paid mobile data per month [58].

Researchers are actively sketching out the details of the technologies that will drive the future 5G wireless networks. The unprecedented performance requirements of 5G will drive an increase in the overall energy consumption of cellular networks and in its carbon footprints. The evolution of cellular networks over its various generations, and the massive increase in performance as well as energy consumed by cellular Radio Access Networks (RANs) over time can be seen clearly in Figure 7. Satisfying these exacting requirements requires (1) greater *spectrum efficiency*, which may be defined as the *number of bits transferred per second per hertz or b/s/Hz*, as well as (2) better *energy efficiency*, which refers to the *energy consumed to communicate an information bit measured in Joules/bit*. The spectral efficiency (SE) and energy efficiency (EE) of wireless networks has been well studied in literature [59,60]. Unfortunately, there is a trade-off between higher spectral efficiency and higher energy efficiency, which may be expressed by the following equation for the case of an Additive White Gaussian Noise (AWGN) Channel:

$$\eta_{EE} = \frac{\eta_{SE}}{N_o(2^{\eta_{SE}} - 1)} \tag{1}$$

where N_o is the noise power spectral density. The SE vs EE tradeoff, however, is complex for practical systems and becomes bell shaped when circuit power P_c is considered. Figure 8 shows the tradeoff comparison for different values of circuit power. Some of the emerging 5G trends such as small cells can be good for energy efficiency as it has been shown that reducing the cell size can increase the number of bits delivered per unit energy for given user density and total power in the service area; similarly, the HetNet arrangement of overlaying a macrocell with micro-/pico-cells at the edge can also help save energy [61]. Figure 9 illustrates a better SE vs EE tradeoff for LTE pico cells as compared to LTE micro cells and Global System for Mobile Communication (GSM).

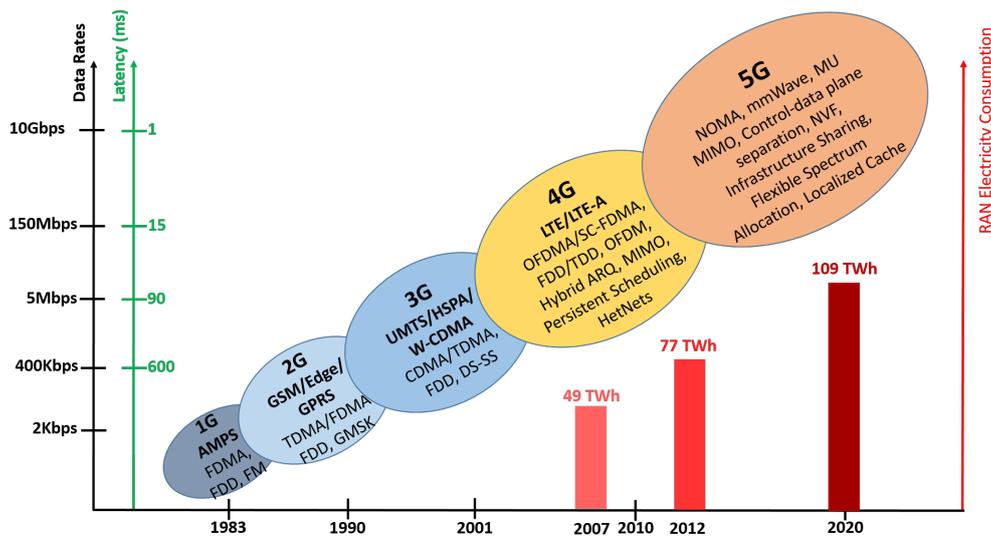


Figure 7. A timeline of historical evolution of cellular networks and the expected electricity consumption of Radio Access Network(RAN) [62].

Researchers have demonstrated that 5G has a rich design space with many potential trade-offs [60], such as *deployment efficiency vs. energy efficiency*; *spectrum efficiency vs. energy efficiency*; *bandwidth vs. power*; and *latency vs. power*. A comparison of parameters for the various 5G use cases are shown in Figure 10. Mainstream 5G research has taken up the challenge of providing high-quality 5G service with gusto, but such an approach will result in an over-engineered system that will be too expensive to install and maintain for low-income and/or rural areas. By suitably scaling the level of service, 5G can encompass a broader coverage base. As an example, the requirement for supporting high-definition (HD) video and tactile Internet may be dropped in low-income scenarios but may be supported for urban and high-income rural environments. Since the rural users would likely prefer “good-enough” service to no service, researchers can explore using the traditional 5G trade-offs in conjunction with the newly developed flexibility of approximate computing and approximate networking to produce cost effective solutions for low-income and/or rural areas.

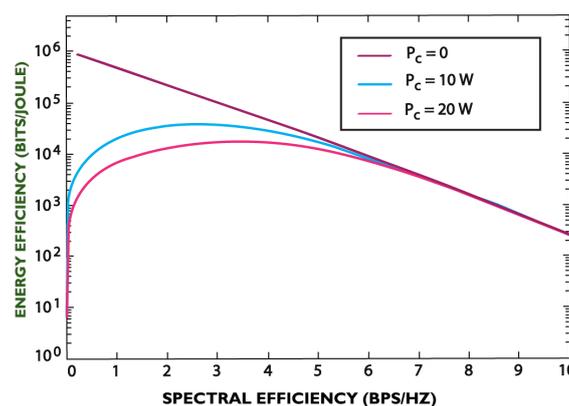


Figure 8. Spectral energy vs. energy efficiency trade-offs for different circuit powers (Adapted from [63]).

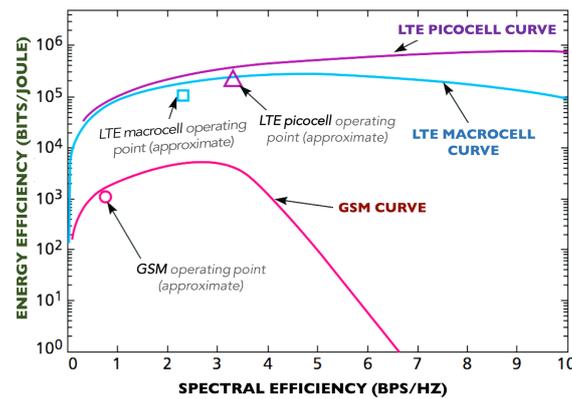


Figure 9. Spectral energy vs. energy efficiency trade-offs for various wireless technologies (Adapted from [63]).

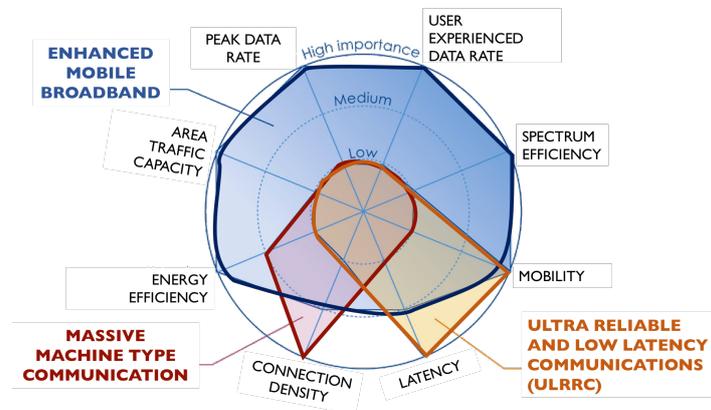


Figure 10. The different requirements of the various 5G use cases. Adapted from resources made available by the European Telecommunications Standards Institute (ETSI), (<http://www.etsi.org/>).

We can also use techniques that bring us more flexibility and convenience in reducing the cost of networking with techniques such as using (1) renewable energy sources such as solar power; (2) base station sleep modes; (3) the building blocks proposed in approximate computing literature (such as the digital baseband processors proposed in [16]) for 5G infrastructure such as base stations; (4) virtualization of network components using techniques such as network functions virtualization (NFV); (5) exploitation of commodity hardware; (6) using community networking and resource pooling principles; (7) achieving flexibility using a separation of control and data planes; and (8) achieving flexibility by integrating Software Defined Networks (SDN) and NFV [64]. Using such technologies, coupled with a well-thought out approximate networking design, it is feasible to provide contextually-appropriate “good-enough” services in a sustainable fashion at a low cost (*in the order of 1 USD for low-income regions and 10 USD for rural regions*).

Chiaraviglio et al. [65] have proposed a mix of low-cost solutions and techniques to bring 5G to rural and low-income regions. For instance, remote radio heads (RRHs) can be mounted on unmanned aerial vehicle and balloons, low cost solar powered large cells (with coverage radius in the order of 50 km), and delay tolerant networks for 5G to provide cellular access in rural areas. This mixture was proposed considering the application context of rural areas, which comprised basic delay tolerant connectivity that could support applications such as e-health, e-learning, and emergency services. An economic analysis of these techniques in [56] showed a monthly subscription fee of 11 euros for rural areas of Italy and Cook Islands and a monthly subscription fee of 1 euro for low-income,

but slightly more populated, rural areas of Zimbabwe. This shows the viability of a low-cost but still functional approximate 5G networking service.

Another way we can adopt approximate networking in 5G in rural and low-income regions is to allow more simultaneous transmissions by multiple users in each orthogonal resource block (be it a time slot, a frequency channel, or a spreading code) using non-orthogonal multiple access (NOMA). NOMA is now being considered as a front runner 5G technology since it is compatible with conventional orthogonal multiple access techniques such as TDMA and OFDMA. NOMA can allow for a context-appropriate tradeoff by catering to varied channel conditions at the users and by allocating more power to the user with poorer channel conditions (or adjusting the power to facilitate the respective context-appropriate QoS requirements). Using the cognitive radio NOMA (CR-NOMA) variant, it can be ensured that some or all of the user's QoS requirements are satisfied. NOMA also allows chunking of users into multiple groups which can then be served in the same orthogonal block using multi-carrier NOMA, and different groups are allocated different orthogonal resource blocks. This can facilitate management of context-appropriate trade-offs for a large numbers of users that can be logically organized in different user classes/groups.

Community networking and resource pooling is another enabling general architectural tool for rural area 5G networking [66]. These efforts are being given a boost by open source technologies and plummeting prices of commodity hardware, e.g., the introduction of low cost software defined radios and open source softwares such as OpenBTS have made community cellular very feasible for rural areas. Using such technologies, community cellular networks can be built in rural area networks in a do-it-yourself (DIY), bottom-up fashion. In contrast to a traditional GSM macrocell for rural areas, which can cost around 500,000 USD; Heimerl et al. [67] have shown a complete deployment of functional 2G community base station in rural Papua, Indonesia, at a much cheaper price of around 10,000 USD. The study showed that 187 subscribers were served with a cheap 10,000 USD base station, thereby providing 830 dollars as monthly revenue with 368 dollars as profit. The GSM band for the project was not purchased due to high spectrum cost and the cell was operating in conjunction with the government officials. This non-availability of legal spectrum can be tackled by exploiting the GSM whitespaces discussed in [68].

It can be anticipated that recent approximate computing hardware advances will soon permeate networking solutions, particularly with 5G appliances becoming virtualized and running as software appliances on cloud data centers with the increased adoption of cloud computing and NFV. The Neural Processing Unit (NPU) is a prominent example of new approximate computing hardware, which can provide significant benefits such as an average speedup and energy savings of $2.3\times$ and $3\times$ respectively with relatively few errors. NPUs have been shown to provide these gains on almost all type of program applications, including signal processing, image processing, machine learning, graphics, and gaming. [31] has tested the gains of NPUs on these applications including Fast fourier Transform which is the one of the most important task of multicarrier communication such as OFDM. Huawei also recently announced Kirin 970, a smart phone processor featuring a dedicated NPU for superior performance and efficiency [69]. With the current developments in approximate computing and the diverse use cases of NPUs, the time is right to apply these concepts to the development of network hardware, such as routers, switches, network interface cards (NICs), as well as to RF components, digital and analog processors, and the various hardware blocks of RANs.

Further studies considering the affects of area and power efficient approximate hardware on cost and efficiency of cellular networks will provide a more detailed overview of the true potential of approximate networking. The time is now ripe to design cost efficient accelerated approximate memory and processors considering the demands of cellular systems. For example, the use of approximate digital blocks to design content addressable memory, widely used for packet classification and forwarding in routers, and the impact of new errors on system level performance will provide deep insights on the feasibility of approximate networking. Similarly, self organizing networks (SON) proposed for future cellular networks [70] use machine learning algorithms for various self

organization phases. Cost, area and power efficient hardware at the cost of 100 percent accuracy are proposed for machine learning and deep learning applications [12]. The use cases for SONs using approximate hardware can reduce the cost of futuristic cellular systems. However, a detailed and careful analysis of the resultant performance degradation and cost benefit is necessary to understand the benefits and penalties of approximate networking. Furthermore, the co-design of approximate hardware and software stack can help to extract maximum benefits of approximate networking.

5. Discussion Issues

5.1. Zero Rating and Net Neutrality

In recent times, the idea that Internet access is a human right is gaining increasing traction (as shown by a recent global Internet survey conducted by the Internet Society [71]). There have been recent efforts that have strived to make basic Internet access accessible to everyone. One approach that has been adopted is the concept of zero-rating through which various mobile companies are offering cheaper (or free) access to selected websites and applications. While the idea at its surface appears noble and harmless, it has become controversial due to its relationship with net neutrality—the idea that all content on the Internet is equal and should be treated such.

One way of provisioning an approximate Internet experience is to curate a *walled garden*. In such an approach, a company or service provider provides access to a limited set of approved sites and services through its platform, typically at a reduced rate or even free. Facebook has adopted the walled garden approach for its *Free Basics* program. This project has however become controversial with India recently ruling that program and others like it infringe the principles of net neutrality. Critics of the program have criticized that Facebook is violating the tenets of net neutrality project and that such a project leads to an unfair market (since in this project, a for-profit company Facebook adopts a position where it can decide which service qualifies as an essential service).

What's Better: Approximate or Zilch?

It has been discussed that how users are often willing to trade-off fidelity for convenience; this would seem to imply that users will be happy with accessible approximate networking solutions when conveniently available. People are divided over the utility of zero rating with one group asserting that zero rating programs can help as a helpful ad-hoc arrangement while others argue that these programs create a tiered Internet ecosystem that is still not able to bridge the digital divide [72]. The net neutrality and zero rating debate is not only a rational debate but also an emotional and moral one. In some cases, people may prefer no networking service to an approximate networking service. Research on the experience of users has however shown that while it is true that “some access is better than none”, users are not always thrilled at being limited to a vastly limited subset of the Internet [73]. To understand why users may turn something for nothing, it is instructive to refer to the classical game-theoretic “*ultimatum game*”. In this game, the first player receives a sum of money and proposes a sharing offer to the second player. The second player can either choose to accept or reject this offer. If the second player accepts, the money is split as proposed, otherwise, neither player gets anything. Under rational economic theory assumptions, the second player should accept any offer since something is better than nothing in utility. However, a number of experiments have shown that typically the responder will not accept “unfair” divisions and will deprive the proposer by rejecting the offer (in the process losing his money as well). This game demonstrates the nuance of the question “what's better: approximate or none?”.

Notwithstanding the problems that current zero-rated approaches are facing, we believe that the goal of democratizing Internet service is worthwhile. Researchers agree that more research is needed to fully assess the impact of specific zero rating initiatives as well as the potential impact of zero rating more generally on Internet adoption in the developing world and in rural and low-income regions [72]. In particular, more research is required to study user preferences regarding limited access

to the complete Internet vs. unlimited access to an incomplete Internet. While Facebook's Free Basics is probably not the right solution, it may be the case that better-designed "good enough" solutions offered at a convenient and affordable price will disrupt the currently established markets.

5.2. HCI Issues: User Perceptions of the Approximation

Previous human-computer-interaction (HCI) research has shown that humans' subjective knowledge can be flawed (relatively fast service may be judged to be unacceptable unless it is also predictable, visually appealing and reliable [74]). More research is needed on which kinds of weaknesses are most perceived by users. This knowledge can then be used to determine the precise approximate networking tradeoff to adopt so that the users perceive least inconvenience. If it turns out that the Pareto principle or the 80/20 rule is applicable, then we can optimize for the approximation that is responsible for 80% of user perceived QoS.

In view of the user's experience, high network latency has an obvious detrimental effect on latency-bound applications such as gaming, voice, and web applications. While bandwidth is admittedly important for QoS (especially for bulk transfer and video applications), it turns out that the round-trip-time (RTT) can dominate performance more than bandwidth for short, bursty HTTP connections [75], especially over long-distance wide-area networks (WANs) [76]. It has been shown in human-computer interaction (HCI) literature that increments of only 100 ms can decrease sales by 1% and human can sense and do react to small differences in the delays of operations [50].

6. Conclusions

The Utopian goal of providing "ideal networking" service universally is an elusive target (due to the moving target nature of "ideal networking" and the lack of affordability of advanced technologies in challenging markets). A lot of experience has highlighted the *fidelity-convenience trade-off*, which is seen when users are willing to tradeoff considerable fidelity for convenience (in terms of accessibility and affordability). In this paper, we have described "approximate networking" as a philosophy that understands that there will no one-size-fits-all, ideal networking solution that will be universally applicable—approximation networking proposes to adopt appropriate context-specific trade-offs to provide "good-enough" service. In addition, we have also provided an overview of approximate networking technologies and have highlighted how a number of existing Internet technologies can be seen as instances of the larger approximation networking vision. Approximate networking, as envisioned in this paper, is a powerful abstraction that has significant promise for Global Access to the Internet for All (GAIA). Realizing approximate networking will require multidisciplinary research in diverse domains such as algorithms, protocols, operating systems, and hardware architectures. We believe that approximate networking can provide a novel fundamental tuning knob that can facilitate context-appropriate trade-offs (by adopting the extra degree-of-freedom of approximate computing at the hardware level) that can help realize the egalitarian vision of GAIA.

Author Contributions: All authors contributed equally to this work. Junaid Qadir and Arjuna Sathiseelan suggested the skeleton of the paper and also provided the supplementary material for paper write up. Junaid Qadir, Umar Bin Farooq and Muhammad Usama wrote the main paper. Muhammad Ali Imran and Muhammad Shafique provided the technical guidance for 5G and Approximate Computing respectively. All authors provided comments on the manuscript at all stages.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. The World Bank. *World Development Report 2016: Digital Dividends*; The World Bank: Washington, DC, USA, 2016; ISBN 978-1-4648-0671-1.
2. Subramanian, L.; Surana, S.; Patra, R.; Ho, M.; Sheth, A.; Brewer, E. Rethinking Wireless for the Developing World. *IRVINE IS BURNING 2006*, 43–48. Available online: <http://conferences.sigcomm.org/hotnets/2006/subramanian06rethinking.pdf> (accessed on 7 December 2017).

3. Onireti, O.; Qadir, J.; Imran, M.A.; Sathiaselalan, A. Will 5G See Its Blind Side? Evolving 5G for Universal Internet Access. In Proceedings of the GAIA '16 2016 Workshop on Global Access to the Internet for All, Florianopolis, Brazil, 22–26 August 2016.
4. Qadir, J.; Sathiaselalan, A.; Wang, L.; Crowcroft, J. Taming Limits with Approximate Networking. In Proceedings of the Second Workshop on Computing within Limits, Irvine, CA, USA, 8–10 June 2016.
5. Han, J.; Orshansky, M. Approximate computing: An emerging paradigm for energy-efficient design. In Proceedings of the 2013 18th IEEE European Test Symposium (ETS), Avignon, France, 27–30 May 2013; pp. 1–6.
6. Shafique, M.; Hafiz, R.; Rehman, S.; El-Harouni, W.; Henkel, J. Cross-layer approximate computing: From logic to architectures. In Proceedings of the 53rd Annual Design Automation Conference (ACM), Austin, TX, USA, 5–9 June 2016; p. 99.
7. El-Harouni, W.; Rehman, S.; Prabakaran, B.S.; Kumar, A.; Hafiz, R.; Shafique, M. Embracing approximate computing for energy-efficient motion estimation in high efficiency video coding. In Proceedings of the 2017 Design, Automation & Test in Europe Conference & Exhibition (DATE), Lausanne, Switzerland, 27–31 March 2017; pp. 1384–1389.
8. Wadhwa, A.; Madhow, U.; Shanbhag, N.R. Slicer Architectures for Analog-to-Information Conversion in Channel Equalizers. *IEEE Trans. Commun.* **2017**, *65*, 1234–1246.
9. Schläfer, P.; Huang, C.H.; Schoeny, C.; Weis, C.; Li, Y.; Wehn, N.; Dolecek, L. Error resilience and energy efficiency: An LDPC decoder design study. In Proceedings of the IEEE 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE), Dresden, Germany, 14–18 March 2016; pp. 588–593.
10. Mishra, A.K.; Barik, R.; Paul, S. iACT: A Software-Hardware Framework for Understanding the Scope of Approximate Computing. In Proceedings of the Workshop on Approximate Computing Across the System Stack (WACAS), Salt Lake City, UT, USA, 2 March 2014.
11. Nair, R. Big data needs approximate computing: Technical perspective. *Commun. ACM* **2015**, *58*, 104.
12. Agrawal, A.; Chen, C.Y.; Choi, J.; Gopalakrishnan, K.; Oh, J.; Shukla, S.; Srinivasan, V.; Venkataramani, S.; Zhang, W. Accelerator Design for Deep Learning Training: Extended Abstract. In Proceedings of the 54th Annual Design Automation Conference 2017 (ACM), Austin, TX, USA, 18–22 June 2017; p. 57.
13. Esmailzadeh, H.; Sampson, A.; Ceze, L.; Burger, D. Architecture support for disciplined approximate programming. *ACM SIGPLAN Notices* **2012**, *47*, 301–312.
14. Chippa, V.; Chakradhar, S.; Roy, K.; Raghunathan, A. Analysis and characterization of inherent application resilience for approximate computing. In Proceedings of the 2013 50th ACM/EDAC/IEEE Design Automation Conference (DAC), Austin, TX, USA, 29 May–7 June 2013.
15. Kugler, L. Is good enough computing good enough? *Commun. ACM* **2015**, *58*, 12–14.
16. Rehman, S.; El-Harouni, W.; Shafique, M.; Kumar, A.; Henkel, J. Architectural-Space Exploration of Approximate Multipliers. In Proceedings of the 2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), Austin, TX, USA, 7–10 November 2016.
17. Venkataramani, S.; Chakradhar, S.T.; Roy, K.; Raghunathan, A. Computing approximately, and efficiently. In Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition, Grenoble, France, 9–13 March 2015; pp. 748–751.
18. Ransford, B.; Ceze, L. SAP: An Architecture for Selectively Approximate Wireless Communication. *arXiv* **2015**, arXiv:1510.03955.
19. Ransford, B.; Sampson, A.; Ceze, L. Approximate Semantics for Wirelessly Networked Applications. Available online: <https://sampa.cs.washington.edu/wacas14/papers/ransford.pdf> (accessed on 7 December 2017).
20. Affordability Report 2014 by Alliance for Affordable Internet. Available online: <http://bit.ly/1BXTS0X> (accessed on 7 December 2017).
21. Koch, R. *The 80/20 Principle: the Secret to Achieving More with Less*; The Crown Publishing Group: New York, NY, USA, 2011.
22. Bouch, A.; Kuchinsky, A.; Bhatti, N. Quality is in the eye of the beholder: Meeting users' requirements for Internet quality of service. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, The Hague, The Netherlands, 1–6 April 2000; pp. 297–304.
23. Raghavan, B.; Ma, J. Networking in the long emergency. In Proceedings of the 2nd ACM SIGCOMM workshop on Green networking, Toronto, ON, Canada, 15–19 August 2011; pp. 37–42.

24. Sathiseelan, A.; Crowcroft, J. LCD-Net: Lowest cost denominator networking. *ACM SIGCOMM Comput. Commun. Rev.* **2013**, *43*, 52–57.
25. Pentland, A.; Fletcher, R.; Hasson, A. Daknet: Rethinking connectivity in developing nations. *Computer* **2004**, *37*, 78–83.
26. Outernet by Alliance for Affordable Internet. Available online: <https://www.outernet.is> (accessed on 7 December 2017).
27. Tyson, G.; Sathiseelan, A.; Ott, J. Could we fit the Internet in a Box? In *Embracing Global Computing in Emerging Economies*; Springer: New York, NY, USA, 2015.
28. Mittal, S. A survey of techniques for approximate computing. *ACM Comput. Surv.* **2016**, *48*, 62.
29. Sampson, A.; Baixo, A.; Ransford, B.; Moreau, T.; Yip, J.; Ceze, L.; Oskin, M. *Accept: A Programmer-Guided Compiler Framework for Practical Approximate Computing*; University of Washington Technical Report UW-CSE-15-01; University of Washington: Washington, DC, USA, 2015.
30. Sampson, A.; Dietl, W.; Fortuna, E.; Gnanapragasam, D.; Ceze, L.; Grossman, D. EnerJ: Approximate data types for safe and general low-power computation. *ACM SIGPLAN Not.* **2011**, *46*, 164–174.
31. Esmaeilzadeh, H.; Sampson, A.; Ceze, L.; Burger, D. Neural acceleration for general-purpose approximate programs. In Proceedings of the 2012 45th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), Vancouver, BC, Canada, 1–5 December 2012; pp. 449–460.
32. Jokela, P.; Zahemszky, A.; Esteve Rothenberg, C.; Arianfar, S.; Nikander, P. LIPSIN: Line speed publish/subscribe inter-networking. *ACM SIGCOMM Comput. Commun. Rev.* **2009**, *39*, 195–206.
33. Talla, V.; Kellogg, B.; Ransford, B.; Naderiparizi, S.; Smith, J.R.; Gollakota, S. Powering the next billion devices with Wi-Fi. *Commun. ACM* **2017**, *60*, 83–91.
34. Jouppi, N. Google Supercharges Machine Learning Tasks with TPU Custom Chip. Available online: <https://cloudplatform.googleblog.com/2016/05/Google-supercharges-machine-learning-tasks-with-custom-chip.html> (accessed on 7 December 2017).
35. Esmaeilzadeh, H.; Sampson, A.; Ceze, L.; Burger, D. Towards neural acceleration for general-purpose approximate computing. In Proceedings of the 2012 45th Annual IEEE/ACM International Symposium on Microarchitecture, Vancouver, BC, Canada, 1–5 December 2012; pp. 449–460.
36. Mazahir, S.; Hasan, O.; Hafiz, R.; Shafique, M.; Henkel, J. An area-efficient consolidated configurable error correction for approximate hardware accelerators. In Proceedings of the 53rd ACM/EDAC/IEEE Design Automation Conference (DAC), Austin, TX, USA, 5–9 June 2016; pp. 1–6.
37. Shafique, M.; Ahmad, W.; Hafiz, R.; Henkel, J. A low latency generic accuracy configurable adder. In Proceedings of the 52nd ACM/EDAC/IEEE Design Automation Conference (DAC), San Francisco, CA, USA, 8–12 June 2015; pp. 1–6.
38. Baker, C.E.; Starke, A.; Xing, S.; McNair, J. Demo Abstract: A Research Platform for Real-World Evaluation of Routing Schemes in Delay Tolerant Social Networks. *arXiv* **2017**, arXiv:1702.05654.
39. Sermpezis, P.; Spyropoulos, T. Not all content is created equal: Effect of popularity and availability for content-centric opportunistic networking. In Proceedings of the 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing, Philadelphia, PA, USA, 11–14 August 2014; pp. 103–112.
40. Larzon, L.A.; Degermark, M.; Pink, S.; Jonsson, L.E.; Fairhurst, G. The Lightweight User Datagram Protocol (UDP-Lite). RFC 3828. Available online: <https://tools.ietf.org/html/rfc3828> (accessed on 7 December 2017).
41. Shelby, Z.; Hartke, K.; Bormann, C. *The Constrained Application Protocol (CoAP)*; RFC 7252; Internet Engineering Task Force: Fremont, CA, USA, 2014. Available online: <https://tools.ietf.org/html/rfc7252> (accessed on 7 December 2017).
42. Krishnan, D.R.; Quoc, D.L.; Bhatotia, P.; Fetzer, C.; Rodrigues, R. Incapprox: A data analytics system for incremental approximate computing. In Proceedings of the 25th International Conference on World Wide Web, Montreal, QC, Canada, 11–15 April 2016; pp. 1133–1144.
43. Gupta, A.; Könemann, J. Approximation algorithms for network design: A survey. *Surv. Oper. Res. Manag. Sci.* **2011**, *16*, 3–20.
44. Gandhi, R.; Kim, Y.A.; Lee, S.; Ryu, J.; Wan, P.J. Approximation algorithms for data broadcast in wireless networks. *IEEE Trans. Mob. Comput.* **2012**, *11*, 1237–1248.
45. Varghese, G. *Network Algorithmics*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2010.
46. Vazirani, V.V. *Approximation Algorithms*; Springer: New York, NY, USA, 2013.
47. Broder, A.; Mitzenmacher, M. Network applications of Bloom filters: A survey. *Internet Math.* **2004**, *1*, 485–509.

48. Maney, K. *Trade-Off: Why Some Things Catch On, and Others Don't*; The Crown Publishing Group: New York, NY, USA, 2010.
49. Bertsekas, D.P.; Gallager, R.G.; Humblet, P. *Data Networks*; Prentice-Hall International: Upper Saddle River, NJ, USA, 1992; Volume 2.
50. Vulimiri, A.; Godfrey, P.B.; Mittal, R.; Sherry, J.; Ratnasamy, S.; Shenker, S. Low latency via redundancy. In Proceedings of the ACM CoNEXT 2013, Santa Barbara, CA, USA, 9–12 December 2013; pp. 283–294.
51. Rondeau, T.W.; Bostian, C.W. *Artificial Intelligence in Wireless Communications*; Artech House: Norwood, MA, USA, 2009.
52. Van Mieghem, P.; Vandenberghe, L. Trade-Off Curves for QoS Routing. In Proceedings of the INFOCOM 2006. 25th IEEE International Conference on Computer Communications, Barcelona, Spain, 23–29 April 2006.
53. Chang, Y.C.; Chang, C.J.; Chen, K.T.; Lei, C.L. Radar chart: Scanning for satisfactory QoE in QoS dimensions. *IEEE Netw.* **2012**, *26*, 25–31.
54. ICT Facts and Figures 2017. International Telecommunication Union. Available online: <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf> (accessed on 7 December 2017).
55. Eriksson, M.; van de Beek, J. Is anyone out there? 5g, rural coverage and the next 1 billion. *IEEE ComSoc Technology News (CTN)*. 2015. Available online: <https://www.comsoc.org/ctn/anyone-out-there-5g-rural-coverage-and-next-1-billion> (accessed on 7 December 2017).
56. Chiaraviglio, L.; Blefari-Melazzi, N.; Liu, W.; Gutiérrez, J.A.; van de Beek, J.; Birke, R.; Chen, L.; Idzikowski, F.; Kilper, D.; Monti, P.; et al. Bringing 5G in Rural and Low-Income Areas: Is it Feasible? *IEEE Commun. Stand. Mag.* **2017**, *1*, 51–57.
57. Smith, D. *The Truth about Spectrum Deployment in Rural America*; Technical Report; Mobile Future: Washington, DC, USA, 2015. Available Online: <http://mobilefuture.org/wp-content/uploads/2015/03/031615-MF-Rural-Paper-FINAL.pdf> (accessed on 7 December 2017)
58. State of Connectivity 2015 A Report on Global Internet Access. Internet.org by Facebook. Available online: <https://fbnewsroomus.files.wordpress.com/2016/02/state-of-connectivity-2015-2016-02-21-final.pdf> (accessed on 7 December 2017).
59. Hasan, Z.; Boostanimehr, H.; Bhargava, V.K. Green cellular networks: A survey, some research issues and challenges. *IEEE Commun. Surv. Tutor.* **2011**, *13*, 524–540.
60. Chen, Y.; Zhang, S.; Xu, S.; Li, Y.G. Fundamental trade-offs on green wireless networks. *IEEE Commun. Mag.* **2011**, *49*, doi:10.1109/MCOM.2011.5783982.
61. Wang, W.; Shen, G. Energy efficiency of heterogeneous cellular network. In Proceedings of the 2010 IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall), Ottawa, ON, Canada, 6–9 September 2010; pp. 1–5.
62. Fehske, A.; Fettweis, G.; Malmudin, J.; Biczok, G. The global footprint of mobile communications: The ecological and economic perspective. *IEEE Commun. Mag.* **2011**, *49*, doi:10.1109/MCOM.2011.5978416.
63. Chih-Lin, I.; Rowell, C.; Han, S.; Xu, Z.; Li, G.; Pan, Z. Toward green and soft: A 5G perspective. *IEEE Commun. Mag.* **2014**, *52*, 66–73.
64. Duan, Q.; Ansari, N.; Toy, M. Software-defined network virtualization: An architectural framework for integrating SDN and NFV for service provisioning in future networks. *IEEE Netw.* **2016**, *30*, 10–16.
65. Chiaraviglio, L.; Blefari-Melazzi, N.; Liu, W.; Gutierrez, J.A.; Van De Beek, J.; Birke, R.; Chen, L.; Idzikowski, F.; Kilper, D.; Monti, J.P.; et al. 5G in rural and low-income areas: Are we ready? In Proceedings of the ITU Kaleidoscope: ICTs for a Sustainable World (ITU WT), Bangkok, Thailand, 14–16 November 2016; pp. 1–8.
66. Qadir, J.; Sathiseelan, A.; Wang, L.; Crowcroft, J. Resource Pooling for Wireless Networks: Solutions for the Developing World. *arXiv* **2016**, arXiv:1602.07808.
67. Heimerl, K.; Hasan, S.; Ali, K.; Brewer, E.; Parikh, T. Local, sustainable, small-scale cellular networks. In Proceedings of the Sixth ICTD Conference (ICTD'13), Cape Town, South Africa, 7–10 December 2013; pp. 2–12.
68. Hasan, S.; Heimerl, K.; Harrison, K.; Ali, K.; Roberts, S.; Sahai, A.; Brewer, E. GSM whitespaces: An opportunity for rural cellular service. In Proceedings of the 2014 IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN), McLean, VA, USA, 1–4 April 2014; pp. 271–282.
69. HUAWEI Reveals the Future of Mobile AI at IFA 2017. Available online: <http://consumer.huawei.com/en/press/news/2017/ifa2017-kirin970/> (accessed on 2 September 2017).

70. Aliu, O.G.; Imran, A.; Imran, M.A.; Evans, B. A survey of self organisation in future cellular networks. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 336–361.
71. Global Internet User Survey 2012 by Interent Society. Available online: <https://www.internetsociety.org/internet/global-internet-user-survey-2012/> (accessed on 7 December 2017).
72. Bates, S.L.; Bavitz, C.T.; Hessekiel, K.H. *Zero Rating & Internet Adoption: The Role of Telcos, ISPs, & Technology Companies in Expanding Global Internet Access*; Berkman Klein Center for Internet & Society Research Publication: Cambridge, MA, USA, 2017. Available online: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33982356> (accessed on 7 December 2017).
73. Pahwa, N. It's a battle for internet freedom. *The Times of India*. 2015. Available online: <https://blogs.timesofindia.indiatimes.com/toi-edit-page/its-a-battle-for-internet-freedom/> (accessed on 7 December 2017).
74. Bouch, A.; Sasse, M.A. It ain't what you charge, it's the way that you do it: A user perspective of network QoS and pricing. In Proceedings of the Sixth IFIP/IEEE International Symposium on Integrated Network Management, 1999, Distributed Management for the Networked Millennium, Boston, MA, USA, 24–28 May 1999; pp. 639–654.
75. Belshe, M. *More Bandwidth Doesn't Matter (Much)*; Google Inc.: Mountain View, CA, USA, 2010.
76. Fall, K.; McCanne, S. You don't know jack about network performance. *Queue* **2005**, *3*, 54–59.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).