Grand Challenges in Signal Processing for Communications

Changyang She[1]*, Peng Cheng[2], Ang Li[3] and Yonghui Li[1]*

[1]School of Electrical and Information Engineering, the University of Sydney, Sydney, NSW, Australia, [2]Department of Computer Science and Information Technology, School of Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC, Australia, [3]School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, China

INTRODUCTION

Mobile communications have become the integral part of our everyday life and until now pursuing ever higher data rates has been central to the development of all its previous generations. The next generation mobile networks will depart from this pursuit and focus on various design objectives in terms of data rate, intelligence, energy and spectral efficiency, number of connections, scalability, latency and reliability. The upcoming cellular network, 5G, promises to expand its personal communications capabilities to enable the Industrial Internet applications. The key to this 5G transformation lies in its three new connectivity types: enhanced mobile broadband (eMBB), massive IoT (Internet of Things) and ultra-reliable low-latency communications (URLLC). However, 5G will not meet all requirements of the future emerging applications (You et al., 2020). Researchers now start working on the sixth generation (6G) communication networks. Compared to 5G, 6G will achieve much higher data rate and spectral/energy efficiency, extreme low latency, high reliability and scalability. Machine learning technologies will play a critical role in 6G networks to achieve automated network management, control and optimization. The traditional signal processing techniques and communication design framework are inadequate to meet the highly dynamic and stringent demands of future applications and complex network scenarios. Moving from the state-of-the-art small pilot projects to a global large-scale deployment requires the design of new signal processing technologies and optimization tools in beyond 5G or 6G wireless systems. In this editorial article, several grand challenges in Signal Processing for Communications are outlined in the following sections.

MULTI-ANTENNA SYSTEMS

It has been widely acknowledged that increasing the size of multi-antenna systems, i.e., the use of a large-scale antenna array, will be indispensable for future wireless communication systems (Larsson et al., 2014; Lu et al., 2014). The main idea behind large-scale antenna arrays is to increase the size of the antenna array to hundreds or even thousands to provide tremendous array gains over small-scale MIMO systems, and the channel hardening effect further enables the simplification of channel estimation process. However, the direct implementation of fully-digital large-scale antenna arrays will result in a significant hardware cost, and the consequent power consumption is also prohibitive for practical deployment. Therefore, the grand challenge for large-scale antenna arrays is the introduction of both hardware-efficient and power-efficient transmit and receive architectures to address the practical implementation issues, and meanwhile handling the performance degradation from hardware impairments such as distortion, phase-noize, and quantization noise, etc., due to the use of low-cost hardware components. More recently, the concept of Holographic MIMO and the use of reconfigurable intelligent surfaces have also emerged as novel multi-antenna architectures, which also

bring new challenges in terms of both hardware design and signal processing techniques (Basar et al., 2019).

MMWAVE AND THZ COMMUNICATIONS

Radio mmwave and THz communications are envisioned as a key technology to enable future ubiquitous wireless communications supporting efficient and flexible massive connectivity with ultra-high user data rates in the order of Tb/s (Akyildiz et al., 2014). However, utilizing the mmwave and THz frequency bands brings unique challenges that require novel RF frontend and antenna design, channel and interference modeling, and resource management. Channel modeling considering molecular noise and blockage probability should be developed to re-evaluate the line-of-sight and non-line-of-sight reflected and scattered components (Priebe and Kurner, 2013). The use of the narrow beam structure requires fundamentally different interference coordination principles (Alkhateeb et al., 2014). Fast handover procedures have to incorporate the time required for discovery, localization, and tracking functionalities. This calls for the theoretical breakthrough and practical implementation of mmwave/THz communications.

MULTI-USER SIGNAL PROCESSING

Interference management is one of the most important topics in multi-user transmission. In traditional view, inter-stream or inter-user interference has been seen as a performance limiting factor in multi-user transmission, and therefore wireless communication systems usually attempt to avoid, mitigate or minimize interference, which is done through precoding at the transmitter side for multi-user multi-antenna transmission (Peel et al., 2005). More recently, a new line of research shows that by manipulating the interfering signals on a symbol level such that they add up constructively at each receiver side, known interference can indeed be made beneficial and further improve the performance of wireless communication systems, achieved via symbol-level precoding based on the concept of constructive interference (Li et al., 2020). The main idea of symbol-level precoding is to exploit both the data symbol information and the channel state information available at the transmitter side to design the precoding strategy, which has provided a new vision toward interference management. Nevertheless, there still exist design challenges for symbol-level precoding toward its practical implementation, which includes computational complexity, applicability to adaptive modulation and coding, SINR estimation, combination with channel coding, etc.

SPARSE SIGNAL PROCESSING

The principle of signal processing aims to decompose complex signals via elementary functions that are easier to handle. In the field of signal processing and wireless communications, there exist discrete signals that are usually sparse in some domain (time, frequency, space, etc.), i.e., most of the entries are zero with only a few non-zero coefficients. Compressed sensing, as a paradigm to recover sparse signals from a small set of linear measurements, has received extensive research attention in recent years (Choi et al., 2017). More relevant to wireless communication, there are a variety of wireless scenarios in which compressed sensing techniques can be applied to. For example, sparse estimation can be used for wireless channel estimation and interference cancellation, support identification can be used for spectrum sensing in cognitive radio, direction (AoA, AoD) estimation, and localization (Bajwa et al., 2010). Nevertheless, there exist some remaining design challenges for compressed sensing in wireless communications, for example the flexible design of compressed sensing and sparse recovery algorithms that are adaptive to diverse wireless environments, the design principle for deterministic sensing matrices, combination with machine learning and deep neural networks, design and implementation complexity, etc.

SPECTRUM AND ENERGY EFFICIENT COMMUNICATIONS

The wireless revolution is fueling an insatiable demand for access to the RF spectrum (Chen and Oh, 2016). Rapidly emerging wireless systems, such as beyond 5G cellular networks and machine-to-machine communications with multi-vendor networks, are competing for precious spectrum resource. On the other hand, enabling energy-efficient green protocols and network architectures can significantly reduce signal transmission power for massive interference coordinations and extend the battery life of wireless devices (Liu and Ansari, 2019). Therefore, it is highly desirable to develop new components, techniques, and architectures to achieve spectrum and energy-efficient wireless communications and networking. These may include advanced spectrum sharing techniques, low power cost and energy efficient resource allocation schemes, novel transceiver design for smart and fast data processing, and wireless charging control and energy management for IoT devices.

CROSS-LAYER OPTIMIZATION

Cross-layer optimization is critical for improving the End-to-End (E2E) performance and understanding the performance limits of communication systems (She et al., 2021). Based on the Open Systems Interconnection model, communication networks can be divided into seven layers, from physical layer to application layer (Jiang et al., 2019). Although the interactions across different layers have significant impacts on the E2E performance, they are not considered in most of the existing studies that only focus on one of the seven layers. One of the bottlenecks in cross-layer optimization is that a cross-layer model is complicated and analytically tractable in general (Amjad et al., 2019). As a result, the objective function or the constraints do not have closed-form expressions. Even if we can obtain closed-form expressions of some problems, they are usually non-convex or NP-hard. Therefore, solving cross-layer optimization problems are extremely challenging, especially for the systems that need to

make decisions in real time. Based on the universal approximation theorem of artificial deep neural networks (DNN) (Hornik et al., 1989), a promising approach to solve cross-layer optimization problems is to use DNNs to approximate the optimal policies. Such a concept has been validated in the existing literature and has the potential to be implemented in real-world systems (Dong et al., 2021; Gu et al., 2021).

CONVEX AND NON-CONVEX OPTIMIZATION

Depending on the properties of feasible regions and objective functions, optimization problems can be either convex or non-convex (Boyd and Vandeberghe, 2004). In communications, there are two branches of studies in convex and non-convex optimization. The first one is to investigate the performance limits of communication systems by deriving the closed-form optimal solutions of the problems, where the first order necessary conditions, also known as the Karush–Kuhn–Tucker (KKT) conditions, have been widely applied (Goldsmith, 2005). It is worth noting that by solving the KKT conditions, one can only obtain a local optimal solution for non-convex problems. Only for convex problems, a local optimal solution is also global optimal. The other branch of works is to design practical algorithms for solving convex or non-convex problems. For convex problems, there are some well-developed low-complexity algorithms that can converge to the global optimal solution, such as the interior-point method (Boyd and Vandeberghe, 2004). For non-convex problems, we can hardly obtain the global optimal solution with low-complexity algorithms in general. Most of existing works developed specific algorithms for their problems case by case. Furthermore, for a variety number of problems in communications, the systems need to make decisions every few milliseconds in real time, e.g., power control, beamforming, scheduling, etc. Even for theoretical low-complexity algorithms like the interior-point method, the computational complexity is still too high to be implemented in real time. To address this issue, unsupervised deep learning has been considered as a promising approach (Eisen et al., 2019), where the DNN used to approximate the optimal policy is trained by the primal-dual method. In this way, the theoretical models in communications are integrated into the unsupervised deep learning for solving optimization problems. After off-line training, we can obtain the output of the DNN (i.e., near optimal solution of the optimization problem) by using the forward propagation algorithm, which can be implemented in real time.

MACHINE LEARNING FOR WIRELESS COMMUNICATIONS

Machine learning and data driven methods have been conceived as a key enabler for future wireless networks evolution (O'Shea and Hoydis, 2017). However, the pathway to the learning-empowered wireless networks is still not clearly identified. A fundamental question is how machine learning can truly benefit the well-established and well-verified communication systems over the last decades. In particular, it is not clear how to adapt the machine learning algorithms widely developed in computer vision and natural language processing to the lower layers such as PHY and MAC in communication protocols, where the design objectives and inherent constraints have dramatically changed (Qin et al., 2019; Wen et al., 2018). Moreover, it necessitates to develop novel learning-based performance limits beyond the classical Shannon theory to establish diversified quality-of-service guarantees in future applications.

CONCLUSION AND OUTLOOK

With the rapid development of emerging new technologies and applications in communications, future communications networks will be a highly complex and dynamic environment. Some important challenges in signal processing technologies that must be catered for future communications to achieve ultra-high data rate, intelligence, high energy and spectral efficiency scalability. Advancement in signal processing technologies and optimization tools will pave the way and fuel innovation for future communication systems. In this editorial article, we have outlined some important areas in emerging signal processing techniques for communications, including large-scale massive MIMO, Holographic MIMO, reconfigurable intelligent surfaces, multi-user signal processing, machine learning, MmWave and THz communications, spectrum and energy efficient communications, cross-layer optimization and convex and non-convex optimization.

AUTHOR CONTRIBUTIONS

CS contributed to Cross-layer Optimization and Convex and non-convex Optimization sections. PC contributed to MmWave and THz Communications, Spectrum and Energy Efficient Communications and Machine Learning for Wireless Communications sections. AL contributed to Multi-Antenna Systems, Multi-User Signal Processing and Sparse Signal Processing Sections. Li coordinated the whole article and wrote the introduction, conclusions and outlook.

REFERENCES

Akyildiz, I. F., Jornet, J. M., and Han, C. (2014). Terahertz band: next Frontier for wireless communications. *Phys. Commun.* 12, 16–32. doi:10.1016/j.phycom.2014.01.006

Alkhateeb, A., Mo, J., Gonzalez-Prelcic, N., and Heath, R. W., Jr. (2014). MIMO precoding and combining solutions for millimeter-wave systems. *IEEE Commun. Mag.* 52 (12), 122–131. doi:10.1109/MCOM.2014.6979963

Amjad, M., Musavian, L., and Rehmani, M. H. (2019). Effective capacity in wireless networks: a comprehensive survey. *IEEE Commun. Surv. Tutorials* 21 (4), 3007–3038. doi:10.1109/COMST.2019.2929001

Bajwa, W. U., Haupt, J., Sayeed, A. M., and Nowak, R. (2010). Compressed channel sensing: a new approach to estimating sparse multipath channels. Proc. *IEEE* 98 (6), 1058–1076. doi:10.1109/JPROC.2010.2042415

Basar, E., Di Renzo, M., De Rosny, J., Debbah, M., Alouini, M., and Zhang, R. (2019). Wireless communications through reconfigurable intelligent surfaces. *IEEE Access* 7, 116753–116773. doi:10.1109/ACCESS.2019.2935192

Boyd, S. P., and Vandenberghe, L. (2004). Convex optimization. Cambridge, United Kingdom: Cambridge University Press.

Chen, Y., and Oh, H. S. (2016). Survey of measurement-based spectrum occupancy modelling for cognitive radios. *IEEE Commun. Surv. Tutorials* 18 (1), 848–859. 10.1109/COMST.2014.2364316

Choi, J. W., Shim, B., Ding, Y., Rao, B., and Kim, D. I.. (2017). Compressed sensing for wireless communications: useful tips and tricks. *IEEE Commun. Surv. Tutorials* 19 (3), 1527–1550. doi:10.1109/COMST.2017.2664421

Dong, R., She, C., Hardjawana, W., Li, Y., and Vucetic, B. (2021). Deep learning for radio resource allocation with diverse quality-of-service requirements in 5G. *IEEE Trans. Wireless Commun. Early Access.* doi:10.1109/TWC.2020.3041319

Eisen, M., Zhang, C., Chamon, L. F. O., Lee, D. D., and Ribeiro, A. (2019). Learning optimal resource allocations in wireless systems. *IEEE Trans. Signal. Process.* 67 (10), 2775–2790. doi:10.1109/TSP.2019.2908906

Goldsmith, A. (2005). Wireless communications. Cambridge, United Kingdom: Cambridge University Press.

Gu, Z., She, C., Hardjawana, W., Lumb, S., McKechnie, D., Essery, T., et al. (2021). Knowledge-assisted deep reinforcement learning in 5g scheduler design: from theoretical framework to implementation. J. Sel. Areas Commun.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks* 2 (5), 359–366. doi:10.1016/0893-6080(89)90020-8

Jiang, X., Shokri-Ghadikolaei, H., Fodor, G., Modiano, E., Pang, Z., Zorzi, M., et al. (2019). Low-latency networking: where latency lurks and how to tame it. *Proc. IEEE* 107 (2), 280–306. doi:10.1109/JPROC.2018.2863960

Larsson, E. G., Edfors, O., Tufvesson, F., and Marzetta, T. L. (2014). Massive MIMO for next generation wireless systems. *IEEE Commun. Mag.* 52 (2), 186–195. doi:10.1109/MCOM.2014.6736761

Li, A., Spano, D., Krivochiza, J., Domouchtsidis, S., Tsinos, C. G., Masouros, C., et al. (2020). A tutorial on interference exploitation via symbol-level precoding: overview, state-of-the-art and future directions. *IEEE Commun. Surv. TutorialsSecondquarter* 22 (2), 796–839. doi:10.1109/COMST.2020.2980570

Liu, X., and Ansari, N. (2019). `Toward green IoT: energy solutions and key challenges. *IEEE Commun. Mag.* 57 (3), 104–110. 10.1109/MCOM.2019.1800175

Lu, L., Li, G. Y., Swindlehurst, A. L., Ashikhmin, A., and Zhang, R. (2014). An overview of massive MIMO: benefits and challenges. *IEEE J. Selected Top. Signal Process.* 8 (5), 742–758. doi:10.1109/JSTSP.2014.2317671

O'Shea, T., and Hoydis, J. (2017). Introduction to deep learning for the physical layer. *IEEE Trans. Cogn. Commun. Netw.* 3 (4), 563. doi:10.1109/TCCN.2017.2758370

Peel, C. B., Hochwald, B. M., and Swindlehurst, A. L. (2005). A vector-perturbation technique for near-capacity multiantenna multiuser communication-Part I: channel inversion and regularization. *IEEE Trans. Commun.* 53 (1), 195–202. doi:10.1109/TCOMM.2004.840638

Priebe, S., and Kurner, T. (2013). Stochastic modeling of THz indoor radio channels. *IEEE Trans. Wireless Commun.* 12 (9), 4445–4455. doi:10.1109/TWC.2013.072313.121581

Qin, Z., Ye, H., Li, G. Y., and Juang, B. F. (2019). Deep learning in physical layer communications, *IEEE Wireless Commun.*. 26, 93–99. 10.1109/MWC.2019.1800601

She, C., Sun, C., Gu, Z., Li, Y., Yang, C., Poor, H. V., et al. (2021). A Tutorial on Ultrareliable and low-latency communications in 6G: integrating domain knowledge into deep learning. *Proc. IEEE* 109 (3), 204–246. doi:10.1109/JPROC.2021.3053601

Wen, C. K., Shih, W. T., and Jin, S. (2018). Deep learning for massive MIMO CSI feedback. *IEEE Wireless Commun. Lett.* 7 (5), 748–751. 10.1109/LWC.2018.2818160

You, X., Wang, C. X., Huang, J., Gao, X., Zhang, Z., Wang, M., et al. (2020). Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. *Sci. China Inf. Sci.* 64, 110301. doi:10.1007/s11432-020-2955-6