On the Limits of Predictability in Real-World Radio Spectrum State Dynamics: From Entropy Theory to 5G Spectrum Sharing

Guoru Ding, Jinlong Wang, Qihui Wu, Yu-Dong Yao, Rongpeng Li, Honggang Zhang, and Yulong Zou

ABSTRACT

A range of applications in cognitive radio networks, from adaptive spectrum sensing to predictive spectrum mobility and dynamic spectrum access, depend on our ability to foresee the state evolution of radio spectrum, raising a fundamental question: To what degree is radio spectrum state (RSS) predictable? In this article we explore the fundamental limits of predictability in RSS dynamics by studying the RSS evolution patterns in spectrum bands of several popular services, including TV bands, ISM bands, cellular bands, and so on. From an information theory perspective, we introduce a methodology of using statistical entropy measures and Fano inequality to quantify the degree of predictability underlying real-world spectrum measurements. Despite the apparent randomness, we find a remarkable predictability, as large as 90 percent, in real-world RSS dynamics over a number of spectrum bands for all popular services. Furthermore, we discuss the potential applications of prediction-based spectrum sharing in 5G wireless communications.

INTRODUCTION

During the past decades we have witnessed a dramatic growth in wireless access along with the popularity of smart phones, mobile TVs, and many other wireless services. The ever-increasing demand for high data rates in the face of limited radio spectrum resources has motivated the introduction of cognitive radio (CR), which opens a potential communication paradigm to improve spectrum utilization by allowing secondary users to opportunistically access spectrum holes or white spaces unused by primary users. To enable CR, one fundamental challenge is how to reliably identify when and where spectrum holes exist.

Spectrum sensing and spectrum prediction are known as two effective enabling techniques to identify spectrum holes. Briefly, spectrum sensing determines radio spectrum state (RSS) using various signal detection methods, which have been investigated extensively in the literature (e.g. a survey in [1]). Complementarily, spectrum prediction infers unknown/unmeasured RSS from historical known/measured spectrum data by exploiting the inherent correlation and/or regularity among them, which has gained increasing attention recently (e.g. a survey in [2]). Spectrum prediction has many merits, for example, reducing sensing time and energy consumption involved in adaptive spectrum sensing [3] and increasing system throughput via prediction-based dynamic spectrum access [4], and so on.

To reap these benefits, a number of spectrum prediction techniques have been proposed, such as time series-based prediction, autoregressive model-based prediction, hidden Markov model-based prediction, neural networks-based prediction, and Bayesian inference-based prediction, etc. (e.g. the survey in [2] and the references therein). However, so far it is not clear what the upper-bound performance of various prediction techniques could be for various frequency bands. Moreover, RSS evolution patterns are generally determined by people's usage of radio spectrum. Although we rarely consider human activity in the radio domain to be totally random, current models of RSS evolution are fundamentally stochastic (see the most widely used continuous/discrete-time Markov chain models in [5]). Yet the probabilistic nature of the existing modeling framework raises fundamental questions: What is the role of randomness in RSS evolution? To what degree are RSS dynamics predictable?

This article attempts to study the interplay between the regular (and thus predictable) and the random (and thus unforeseeable) underlying real-world RSS dynamics theoretically and provide certain guidance over how to apply the pre-

Guoru Ding, Jinlong Wang, and Qihui Wu are with PLA University of Science and Technology.

Yu-Dong Yao is with Stevens Institute of Technology.

Rongpeng Li and Honggang Zhang are with Zhejiang University.

Yulong Zou is with Nanjing University of Posts and Telecommunications.

dicted RSS to the design of future wireless communication systems technically. Specifically, from an information theory perspective, we introduce a methodology of using statistical entropy measures and Fano inequality to quantify the degree of predictability underlying real-world spectrum measurements and provide some intuitive thoughts and conclusions. After validating the fundamental limits of predictability in RSS dynamics, this article moves forward by addressing the potential applications of the predicted RSS in 5G wireless communications.

SPECTRUM DATA DESCRIPTION AND PREPROCESSING

In order to ensure the reproducibility of the spectrum prediction analysis in this article for other researchers, we use a well known open source real-world spectrum dataset from the RWTH Aachen University spectrum measurement campaign [5]. In this article we are primarily interested in several popular services, including TV bands, ISM bands, cellular bands, and so on. The resolution bandwidth of each individual spectrum band is 200 kHz. The inter-sample time is about three minutes, which corresponds to 3360 samples in one week for each 200 kHz spectrum band.[1]

As an illustrative example, Fig. 1 shows the evolution trajectories of a one week real-world RSS, that is, measured power spectral density (PSD) values, in TV bands. Several interesting phenomena can be observed. First, the RSS dynamics for various frequency bands are significantly different; several bands are heavily loaded but others not. Moreover, randomness and regularities coexist in the RSS evolution. Very strong signals can be identified in several TV bands, and it appears that the temporal variations of signals in these bands are not as significant as those in other bands.

To further show the spectrum utilization of each 200 kHz spectrum band, Fig. 2 plots the duty cycle over the frequency under two well known detection thresholds. One threshold, 107 dBm/200 kHz, has initially been proposed in the IEEE 802.22 working group for detection of wireless microphones in 200 kHz channels in the TV bands; the other more sensitive threshold, 114 dBm/200 kHz, has been specified in the FCC's final rules [5]. As shown in Fig. 2, binary spectrum occupancy (BSO) is highly dependent on the selection of the specific detection threshold.

Until now, to characterize RSS dynamics, most of the existing studies have focused on BSO traces by analyzing the ON and OFF state evolution over time. Instead, in this article we will investigate the continuously measured PSD traces and analyze the predictability of PSD evolution over time, mainly for the following concerns: the PSD is the original raw data, while the BSO, obtained from the PSD by comparing it with a detection threshold, inevitably introduces detection or sensing errors (e.g. false alarms and missed detections) [6].
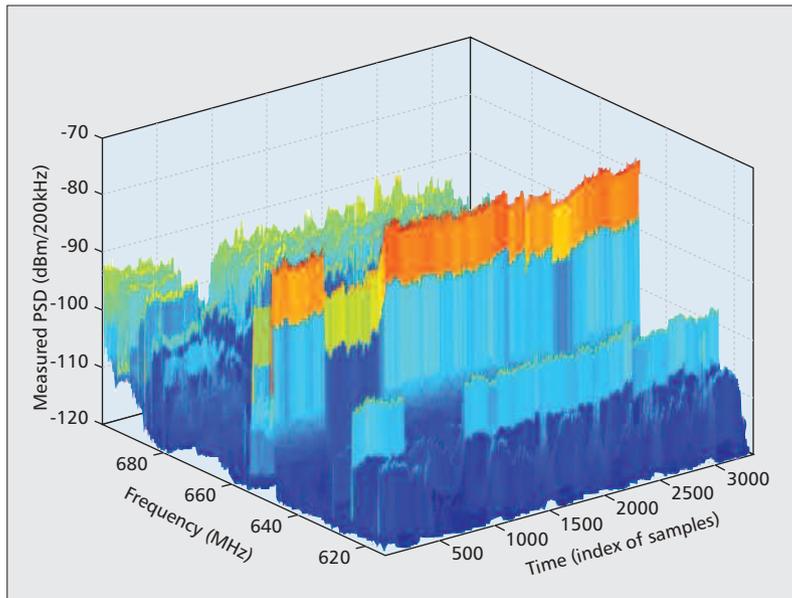
Figure 1. The 3-D view of the evolution trajectories of one week real-world RSS in the TV bands (614∼698 MHz). For each 200 kHz spectrum band, about 3360 samples one week and thus 480 samples one day are plotted.

SPECTRUM PREDICTION ANALYSIS: TO WHAT DEGREE IS RADIO SPECTRUM STATE PREDICTABLE?

In this section we first perform prediction analysis on each individual 200 kHz spectrum band separately, and then on the entire spectrum band allocated to each service statistically.

ENTROPY ANALYSIS

For a given spectrum band, let X_i be a random variable representing its state at time slot i. The state of this band from time slot 1 to time slot n is a random variable series $X_1, X_2, ..., X_n$. Entropy is probably the most fundamental quantity characterizing the degree of predictability of a random variable series. In general, lower entropy implies higher predictability, and vice versa. Recently, entropy-based analysis has been introduced in various prediction scenarios such as atmosphere [7], network traffic [8], and human mobility [9]. The basic idea is that entropy offers a precise definition of the informational content of predictions and it is renowned for its generality due to minimal assumption on the model of the studied scenario.

Specifically, to facilitate the following entropy analysis of RSS dynamics, we first quantize the PSD values for each individual spectrum band into Q RSS levels. Then, let $S = \{X_1, X_2, ..., X_n\}$ denote the series or sequence of RSS levels occurred at n consecutive time slots and we have the following three entropy measures to characterize RSS dynamics:

• Random entropy $E^{\mathrm{rand}} = \log_2 Q$, capturing the degree of predictability of the given spectrum band's evolution if each RSS level occurs with equal probability in each time slot.

• Temporal-uncorrelated entropy $E^{\mathrm{unc}} = -\Sigma_{i=1}^n p_i \log_2 p_i$, where p_i is the probability that the i-th RSS level occurred in the sequence S. E^{unc},

also known as Shannon entropy or classical information theoretical entropy, is by far the most often used entropy metric, which characterizes the heterogeneity of the RSS evolution patterns without taking into account the history of the process.

• Actual entropy $E^{actual} = -\Sigma_{S_i \subset S} P(S_i) \log_2 P(S_i)$, where $P(S_i)$ is the probability of a particular time-ordered subsequence S_i occurring in the trajectory of S. Thus, E^{actual} depends not only on the occurrence frequency of each RSS level, but
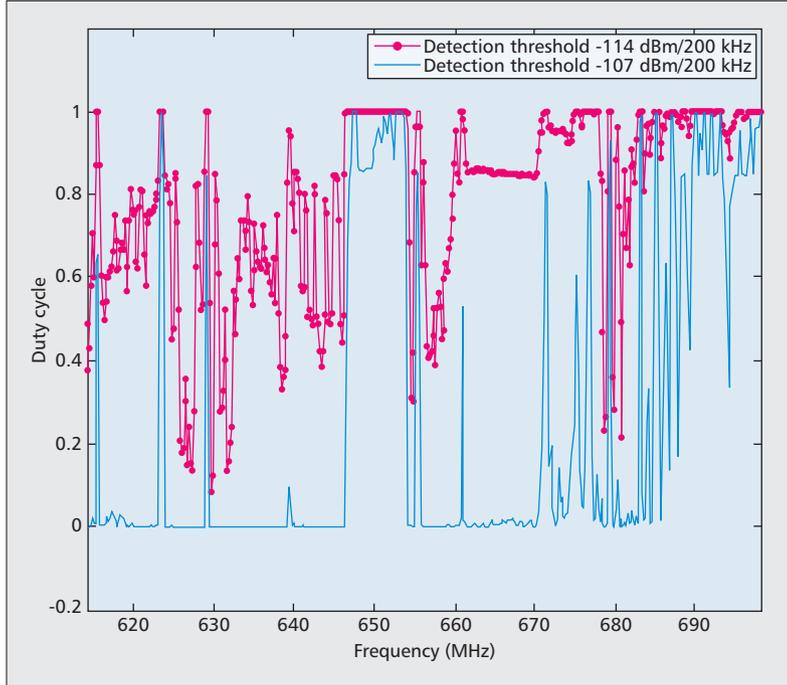
Figure 2. Impact of the detection threshold on the duty cycle in TV bands (614~698 MHz).
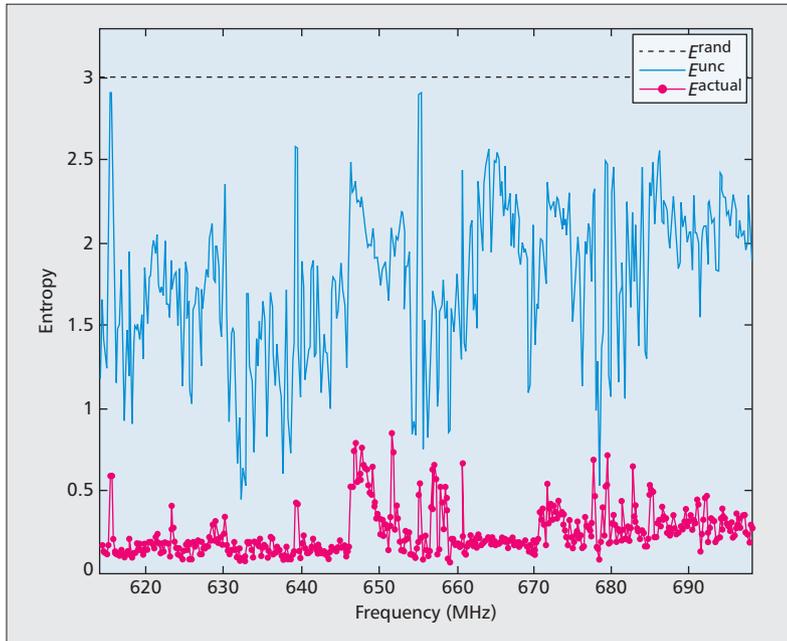
Figure 3. The entropy of the RSS dynamics in TV bands. The number of RSS levels is set as $Q = 8$ for each individual 200 kHz spectrum band and thus the random entropy $E^{rand} = \log_2(Q = 8) = 3$ bits.

also the temporal order in which the RSS levels occurred, and it captures the full frequency-time structure present in a given spectrum band's revolution pattern. In practice, to calculate the actual entropy from the historical spectrum measurements, we use an estimator based on Lempel-Ziv data compression [10], which is known to rapidly converge to the actual entropy of a time series. For a time series with length n, the entropy is estimated by $E_{est}^{actual} = ((1/n)\Sigma_{i=1}^{n} \Lambda_i)^{-1} \ln n$, where Λ_i is the length of the shortest subsequence starting at the i-th time slot which does not previously appear from time slot 1 to time slot i.

Intuitively, we have $0 \leq E^{actual} \leq E^{unc} \leq E^{rand}$, which is illustrated in Fig. 3 via analyzing the real-world spectrum measurements in TV bands. Extremely, if a spectrum band has actual entropy $E^{actual} = 0$, its RSS evolution is completely regular and thus fully predictable. However, if a spectrum band's actual entropy $E^{actual} = E^{rand} = \log_2 Q$, its trajectory is expected to follow a quite random pattern and thus we cannot predict it with an accuracy exceeding $1/Q$. As shown in Fig. 3, all spectrum bands have finite actual entropies between 0 and E^{rand}, indicating that not only a certain amount of *randomness* governs their future whereabouts, but also that there is some *regularity* in their dynamics that can be exploited for predictive purposes.

Based on the obtained actual entropy, in the following we aim to quantify the limits of the predictability of a spectrum band's next state based on its trajectory history. That is, we want to answer the question: How predictable is a spectrum band's next state given the entropy of its historical trajectory?

PREDICTABILITY ANALYSIS

An important measure of predictability is the probability Π that an appropriate predictive algorithm can correctly predict a spectrum band's future state. This quantity is subject to Fano's inequality [11]. That is, if an individual spectrum band with an actual entropy E^{actual} evolutes between Q RSS levels, its predictability $\Pi \leq \Pi^{max}$, where Π^{max} is determined by

$$E^{actual} = -[\Pi^{max}\log_2\Pi^{max} + (1 - \Pi^{max})\log_2(1 - \Pi^{max})] + (1 - \Pi^{max})\log_2(Q - 1).$$

Based on this relationship, for each spectrum band, we can obtain the upper-bound predictability, Π^{max}, through numerical calculations given Q and E^{actual}.

As an illustrative example, Fig. 4 shows the upper-bound predictability P^{max} over each 200 kHz TV band separately when the number of RSS levels is set as $Q = 8$. For comparison, the predictability of independent identical distributed (i.i.d.) Gaussian noise data with one-week samples is also plotted. We have the following observations:

• The predictability of real-world RSS data varies significantly for different spectrum bands. For example, there are a number of TV bands with the predictability higher than 0.95, which means that at most five percent of the time these spectrum bands change their states in a manner that appears to be random, and in the

remaining 95 percent of the time we can expect to predict their whereabouts. On the other hand, we also see that there are a few TV bands with predictability lower than 0.9, which means that regardless of how good our predictive algorithms are, we cannot predict with better than 90 percent accuracy the future states of these spectrum bands.

• For all TV bands, the predictability of real-world RSS data are much higher than that of the i.i.d Gaussian noise data. This demonstrates that the temporal correlation or regularity in real-world RSS data benefits the predictability.

Furthermore, from a statistical perspective, Fig. 5 shows the cumulative distribution functions (CDFs) for the predictability (with $Q = 8$) of various services, including TV bands (614~698 MHz), ISM bands (2400.1~2483.3 MHz), cellular bands (GSM1800 uplink 1710.2~1784.8 MHz and GSM1800 downlink 1820.2~1875.4 MHz), and 2.3 GHz bands (2300~2400 MHz).[2] We have the following observations:

• Among all services, TV bands have the steepest CDF, with minimum predictability of 0.8836. Comparatively, most bands in 2.3 GHz have relatively low predictability, with the minimum close to the predictability of i.i.d. Gaussian noise data, 0.7623. ISM bands have a CDF between TV bands and cellular bands, which implies that a larger (lower) proportion of ISM bands have higher predictability than those of cellular (TV) bands.

• A predictability superiority of the GSM1800 downlink is observed over the GSM1800 uplink for spectrum bands with predictability levels in the bottom 70 percent. However, for spectrum bands with predictability levels in the top 30 percent, a predictability superiority of the GSM1800 uplink is observed over the GSM1800 downlink. That is, although a majority of the GSM1800 downlink bands have superior predictability, there are some GSM1800 uplink bands that have very high predictability levels. This somewhat conflicting observation might result from the fact that quite regular patterns of human spectrum usage exist in few GSM1800 uplink bands.

APPLICATIONS: 5G SPECTRUM SHARING

Radio spectrum usage is an essential issue in 5G wireless communications [12]. The explosion of data rates offered by mobile Internet and the Internet of Things (IoTs) is overwhelming the allocated 2G/3G/4G radio spectrum. In the past, new cellular spectrum has typically been made available through spectrum refarming. However, clearing radio spectrum from an allocated but under-utilized usage to repurpose the spectrum band to another usage often requires many years to accomplish, which makes it difficult to keep pace with user demand of gigabit per second (Gb/s) data rates for 5G [13]. On the other hand, technological innovations such as millimeter wave communications and visible light communications can offer very high data rates; however, these disruptive technologies are mainly for small cells and low mobility usage. To provide wide area cell types, spectrum resources below 3
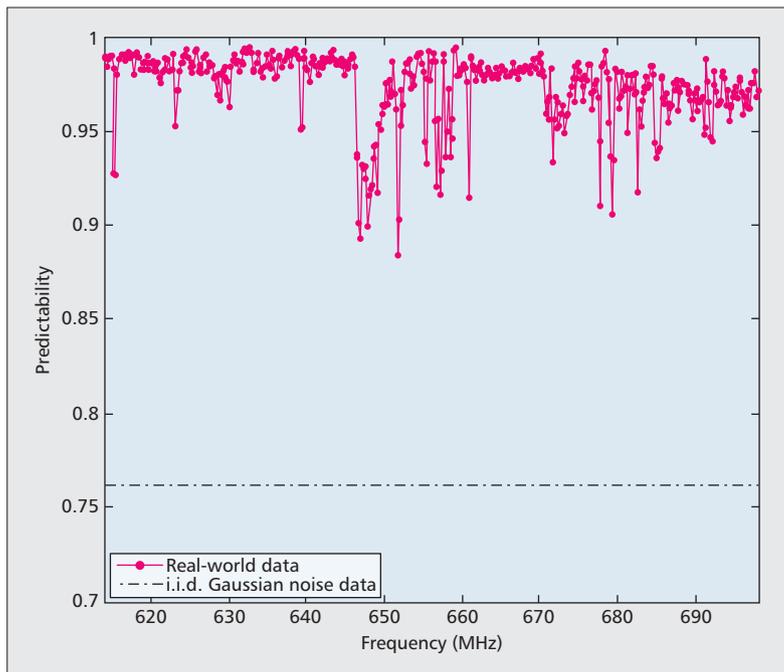

Figure 4. The predictability in RSS dynamics for TV bands.
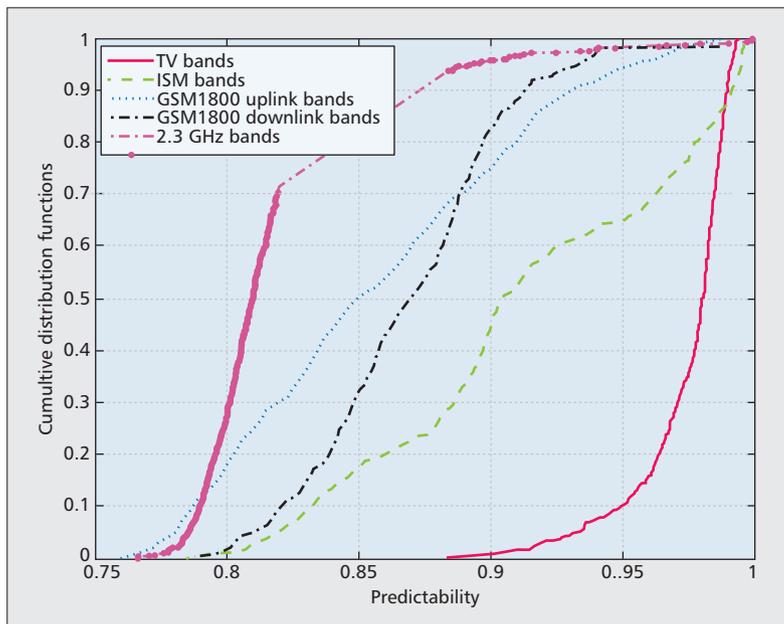

Figure 5. The cumulative distribution functions (CDFs) for the predictability of various services.

GHz will be needed [14].

To address these challenges, spectrum sharing is contemplated as the primary candidate, which has been well recognized as an affordable, near-term solution to meet 5G radio spectrum requirements and increase radio access network capacities for 5G content delivery. Specifically, *5G spectrum sharing is well beyond the previous studies on cognitive radio-based spectrum sharing*, since one main feature of the latter is the opportunistic primary-secondary access in unlicensed bands (such as TV white space). In contrast, as shown in Fig. 6, 5G spectrum sharing may occur

[2] The predictability results on 2.3 GHz bands are include in Fig. 5, since Europe is currently looking at deploying licensed/authorized shared access within these bands.
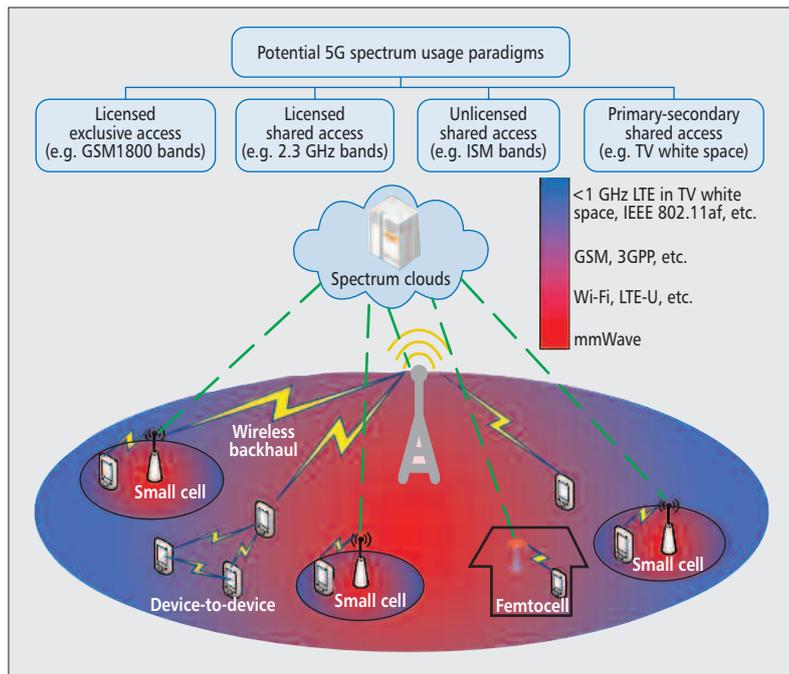
Figure 6. A vision for potential 5G spectrum usage paradigms.

frequency selection and predictive interference mitigation.

CONCLUSION AND DISCUSSIONS

Predicting radio spectrum state evolution has been gaining increasing attention because of the explosive growth in demand for dynamic spectrum access. In this article statistical entropy measures and Fano inequality are exploited to quantify the degree of predictability underlying real-world spectrum measurements. The results in this article, serving as the upper-bound prediction performance, can provide a performance bound of various predictive algorithms and a general guide to the design of future wireless communication systems. Notably, it remains a challenge for the state-of-the-art prediction techniques to obtain prediction precision approaching the upper-bound predictability. Further improvement in the forecast accuracy of spectrum prediction techniques in a real-time mode are thus required.

in both licensed bands (e.g. GSM1800 bands, 2.3 GHz bands) and unlicensed bands (e.g. ISM bands, TV bands). Moreover, one distinguishing feature of potential 5G spectrum usage is the diversity, that is, besides the licensed exclusive access in traditional cellular networks, licensed/authorized shared access, unlicensed shared access (known as LTE in unlicensed bands), and primary-secondary access will coexist [15].

Spectrum prediction will play a significant role in 5G spectrum sharing. Several potential applications are described below.

Cost-Efficient Wideband Carrier Aggregation: To meet 5G capacity requirements, it is known that no single band or air interface standard by itself fully suffices, and it is inevitable for 5G devices to aggregate the benefits of multiple (non-continuous) spectrum bands of a very wide range, possibly from several hundreds MHz bands to 30-300 GHz millimeter wave bands. Consequently, proactive schemes are expected to exploit the evolution dynamics of various spectrum bands of such a wide range, and enable wideband carrier selection and aggregation in a timely and cost-efficient manner.

Dynamic Frequency Selection and Predictive Interference Mitigation: One dominant theme for wireless evolution into 5G is network densification, which is realized mainly by increasing the density of infrastructure nodes (such as base stations and relays) in a given geographic area. It is anticipated that hyper-dense small cells are largely privately owned, and of unplanned deployment. The small cells thus need to be capable of being configured, optimized, and healed by themselves to select the communication frequency bands and not to cause any noticeable interference to the existing neighborhood networks. The knowledge from spectrum prediction can be used by the small cells to assist such autonomous processes through dynamic

REFERENCES

[1] E. Axell *et al.*, "Spectrum Sensing for Cognitive Radio: State-of-the-Art and Recent Advances," *IEEE Signal Process. Mag.*, vol. 29, no.3, May 2012, pp. 101–16.
[2] X. Xing *et al.*, "Spectrum Prediction in Cognitive Radio Networks," *IEEE Wireless Commun.*, vol. 20, no. 2, April 2013, pp. 90–96.
[3] G. Ning and P. Nintanavongsa, "Time Prediction Based Spectrum Usage Detection in Centralized Cognitive Radio Networks," *IEEE WCNC*, April 2012, pp. 300–05.
[4] S. Yin *et al.*, "Prediction-Based Throughput Optimization for Dynamic Spectrum Access," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, Mar. 2011, pp. 1284–89.
[5] M. Wellens, "Empirical Modelling of Spectrum Use and Evaluation of Adaptive Spectrum Sensing in Dynamic Spectrum Access Networks," Ph.D. dissertation, RWTH Aachen University, May 2010.
[6] G. Ding *et al.*, "Kernel-Based Learning for Statistical Signal Processing in Cognitive Radio Networks: Theoretical Foundations, Example Applications, and Future Directions," *IEEE Signal Process. Mag.*, vol. 30, no. 4, July 2013, pp. 126–36.
[7] R. Kleeman, A. J. Majda, and I. Timofeyev, "Quantifying Predictability in a Model with Statistical Features of the Atmosphere," *Proc. National Academy of Sciences of the United States of America*, vol. 99, no. 24, Nov. 2002, pp. 15291–96.
[8] R. Li *et al.*, "The Prediction Analysis of Cellular Radio Access Network Traffic: From Entropy Theory to Networking Practice," *IEEE Commun. Mag.*, vol. 52, no. 6, June 2014, pp. 234–40.
[9] C. Song *et al.*, "Limits of Predictability in Human Mobility," *Science*, vol. 327, no. 5968, Feb. 2010, pp. 1018–21.
[10] I. Kontoyiannis *et al.*, "Nonparametric Entropy Estimation for Stationary Processes and Random Fields, with Applications to English Text," *IEEE Trans. Inf. Theory*, vol. 44, vol. 3, May 1998, pp. 1319–27.
[11] R. Fano and D. Hawkins, "Transmission of Information:

A Statistical Theory of Communications," *Am. J. Phys.*, vol. 29, no. 793, 1961.

[12] J. G. Andrews *et al.*, "What will 5G Be?" *IEEE JSAC*, vol. 32, no. 6, June 2014, pp. 1065–82.

[13] J. Mitola *et al.*, "Accelerating 5G QoE via Public-Private Spectrum Sharing," *IEEE Commun. Mag.*, vol. 52, no. 5, May 2014, pp. 77–85.

[14] Z. Khan, *et al.*, "Carrier Aggregation/Channel Bonding in Next Generation Cellular Networks: Methods and Challenges," *IEEE Network*, vol. 28, no. 6, Nov.-Dec. 2014, pp. 34–40.

[15] Special Interest Group on Cognitive Radio for 5G, "White Paper Novel Spectrum Usage Paradigms for 5G," *IEEE Technical Committee on Cognitive Networks (TCCN)*, Nov. 2014; http://cms.comsoc.org/eprise/main/SiteGen/TCCN/Content/Home/WhitePapers.html}

BIOGRAPHIES

GUORU DING (dingguoru@gmail.com) received his B.S. degree in electrical engineering from Xidian University, Xi'an, China, in 2008, and his Ph.D. degree in communications and information systems from the College of Communications Engineering, PLA University of Science and Technology, Nanjing, China, in 2014. His research interests include wireless security, cognitive radio networks, machine learning, and big data analytics over wireless networks. He was a recipient of the Best Paper Awards from IEEE VTC 2014-Fall and IEEE WCSP 2009. He actively participates in the international standardization association IEEE DySPAN Standards Committee and acts as the Secretary of IEEE 1900.6 and one of the voting members both in IEEE 1900.7 and IEEE 1900.6.

JINLONG WANG (wjl543@sina.com) received his B.S. degree in wireless communications, and the M.S. degree and Ph.D. degree in communications and electronic systems from the Institute of Communications Engineering, Nanjing, China, in 1983, 1986, and 1992, respectively. He is currently a chair professor at PLA University of Science and Technology, Nanjing, China. He is also the co-chair of the IEEE Nanjing Section. He has published widely in the areas of signal processing for wireless communications and networking. His current research interests include software defined radio, cognitive radio, and green wireless communication systems.

QIHUI WU (wqhtxdk@aliyun.com) received his Ph.D. degree in communications and information systems from the Institute of Communications Engineering, Nanjing, China, in 2000. From 2003 to 2005, he was a postdoctoral research associate at Southeast University, Nanjing, China. From 2005 to 2007 he was an associate professor with PLA University of Science and Technology, Nanjing, China, where he is currently a full professor. From March 2011 to September 2011 he was an advanced visiting scholar at Stevens Institute of Technology, Hoboken, NJ, USA. His research interests include software defined radio, cognitive radio, and smart radio.

YU-DONG YAO (yyao@stevens.edu) has been with Stevens Institute of Technology, Hoboken, NJ, USA since 2000, and is currently a professor and department director of electrical and computer engineering. From 1989 to 1990 he was at Carleton University, Ottawa, Canada, as a research associate working on mobile radio communications. From 1990 to 1994 he was with Spar Aerospace Ltd., Montreal, Canada, where he was involved in research on satellite communications. From 1994 to 2000 he was with Qualcomm Inc., San Diego, California, where he participated in research and development in wireless CDMA systems.

RONGPENG LI (lirongpeng@zju.edu.cn) is pursuing his doctorate degree in the Department of Information Science and Electronic Engineering, Zhejiang University, China. From September 2013 to December 2013 he was a visiting doctoral student in Supélec, France. From 2006 to 2010 he studied in the Honor Class, School of Telecommunications Engineering, Xidian University, China, and received his B.E. as an "Outstanding Graduate" in June, 2010. His research interests focus on green cellular networks, applications of reinforcement learning, and analysis of cellular network data.

HONGGANG ZHANG (honggangzhang@zju.edu.cn) is a professor at Zhejiang University, China, and was the International Chair Professor of Excellence for Université Européenne de Bretagne (UEB) & Supélec, France (2012-2014). He is also an Honorary Visiting Professor at the University of York, UK. He served as the Chair of the Technical Committee on Cognitive Networks of the IEEE Communications Society from 2011 to 2012. He is currently involved in research on green communications and was the lead guest editor of the *IEEE Communications Magazine* special issues on "Green Communications."

YULONG ZOU (yulong.zou@njupt.edu.cn) is a full professor and doctoral supervisor at Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China. He received a B.Eng. degree in information engineering from NUPT, Nanjing, China, in July 2006, a first Ph.D. degree in electrical engineering from Stevens Institute of Technology, Hoboken, New Jersey, USA, in May 2012, and a second Ph.D. degree in signal and information processing from NUPT, Nanjing, China, in July 2012. His research interests span a wide range of topics in wireless communications and signal processing, including cooperative communications, cognitive radio, wireless security, and energy-efficient communications. He is currently serving as an editor for *IEEE Communications Surveys & Tutorials*, *IEEE Communications Letters*, *EURASIP Journal on Advances in Signal Processing*, and *KSII Transactions on Internet and Information Systems*, among others.