

# Sliding Window Spectrum Sensing for Full-Duplex Cognitive Radios with Low Access-Latency

Orion Afisiadis, Andrew C. M. Austin, Alexios Balatsoukas-Stimming and Andreas Burg

**Abstract**—In a cognitive radio system the failure of secondary user (SU) transceivers to promptly vacate the channel can introduce significant *access-latency* for primary or high-priority users (PU). In conventional cognitive radio systems, the backoff latency is exacerbated by frame structures that only allow sensing at periodic intervals. Concurrent transmission and sensing using self-interference suppression has been suggested to improve the performance of cognitive radio systems, allowing decisions to be taken at multiple points within the frame. In this paper, we extend this approach by proposing a sliding-window full-duplex model allowing decisions to be taken on a sample-by-sample basis. We also derive the access-latency for both the existing and the proposed schemes. Our results show that the access-latency of the sliding scheme is decreased by a factor of 2.6 compared to the existing slotted full-duplex scheme and by a factor of approximately 16 compared to a half-duplex cognitive radio system. Moreover, the proposed scheme is significantly more resilient to the destructive effects of residual self-interference compared to previous approaches.

**Index Terms**—Full Duplex Cognitive Radio, Spectrum Sensing, Energy Detection, Self-Interference Suppression.

## I. INTRODUCTION

One of the most significant challenges faced by wireless systems today is the (apparent) scarcity of available spectrum. However, recent studies have shown that while the radio spectrum is densely allocated, it is often not heavily occupied or utilized by the licensed primary users (PU) [1], [2]. Frequency-agile cognitive radio networks have been proposed to take advantage of this situation, by allowing unlicensed secondary users (SU) to opportunistically reuse licensed frequency bands [3]. One of the fundamental requirements of such systems is that SUs should not generate harmful interference to PUs. Consequently, SU transceivers must be capable of sensing the radio channel to determine if a PU is present [4]. Similar to spectrum re-use, we can also consider a scenario that requires low-latency medium access for high-priority users or high-priority transmissions. In such systems, a less latency-sensitive ongoing transmission would need to stop instantly as a PU forcefully accesses the channel to get its urgent message across. Such scenarios are of particular importance for real-time services, such as virtual/augmented reality and autonomous vehicles and therefore the proposed standards for future 5G systems aim to achieve less than 1 ms latency [5].

System architectures that periodically stop SU transmissions to sense the channel have been widely proposed [6], [7] to detect the start of PU transmissions. These approaches introduce *blind-intervals*, where the SU system is transmitting and thus unable to detect the start of a PU transmission until the next sensing slot (at the earliest). Decreasing the

interval between successive sensing-slots will decrease the efficiency and throughput of the SU system [7] but improve its spectrum sensing capabilities. Previous research has examined the trade-off between sensing and throughput, and introduced scheduling algorithms to maximise sensing efficiency [8], and MAC-layer frame structures to maximise SU throughput while adequately protecting the PU [7]. An alternative approach to improve the detection-throughput trade-off, by using self-interference cancellation to enable concurrent transmission and sensing (similar to full-duplex systems [9]), has been proposed by a number of authors [10], [11], [12], [13], [14]. Unfortunately, in practice, perfect self-interference suppression is usually not attained, and for typical operating conditions the residual self-interference component is above the noise floor [15]. Previous research has considered the analysis of the sensing-throughput trade-off for energy detection [11], waveform detection [16], the power-throughput trade-off [14], [17] and various adaptive algorithms for maximizing SU throughput [12], [13], in the presence of different levels of residual self-interference.

In existing work, the PU is assumed to be protected if the probability of detection,  $P_d$ , is sufficiently high (typically above 90%) [6], [13]. However, from a physical-layer perspective, the *access-latency* (defined as the time required by the SU to detect the PU and vacate the channel) gives a better measure of the impact of SU interference and the necessary protection. For example, a high access-latency may harm PU communication, by distorting a too large portion of its training or synchronisation fields. Moreover, the full-duplex<sup>1</sup> cognitive radio schemes considered in [11] and [14] typically take decisions after  $N_s$  samples have been accumulated into a buffer. These schemes can thus have a high access-latency, as the PU may start transmitting at any time.

**Contributions:** In this paper we focus on obtaining analytical expressions for the physical access-latency in various cognitive radio systems. The access-latency results are presented in terms of physical samples and can thus be scaled and applied to arbitrary hardware implementations. Furthermore, to alleviate the issue of high access-latency in full-duplex cognitive radio systems, we introduce a full-duplex *sliding-window* spectrum sensing technique. Unlike existing schemes, our approach takes decisions on a sample-by-sample basis, and

<sup>1</sup> In this paper the conventional cognitive radio system is referred to as a half-duplex scheme, while all the systems using self-interference suppression to enable concurrent transmission and sensing are referred to as full-duplex schemes.

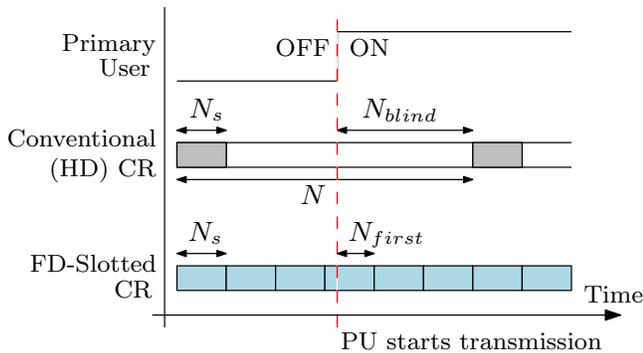


Fig. 1. Frame structures for conventional (HD) and concurrent (FD) cognitive radio systems:  $\square$ : SU transmission;  $\blacksquare$ : sensing only; and  $\blacksquare$ : concurrent transmit and sense.

can detect the presence of PUs more quickly, thereby reducing the access-latency as demonstrated by our simulation results.

*Outline:* This paper is organized as follows. In Section II we describe the system models for the existing half-duplex and slotted full-duplex systems, along with the model of our proposed sliding-window full-duplex system. The throughput-latency trade-off is analysed in Section III, along with derivations of expressions for the access-latency. Theoretical and numerical simulation results are shown in Section IV, and we conclude in Section V.

## II. SYSTEM MODELS

In this section we outline the three cognitive radio system models considered in this paper. For all three approaches, the presence of the PU is detected by comparing a decision metric,  $M$ , (computed from a set of  $N_s$  samples), against a threshold,  $\epsilon$ . If  $M > \epsilon$ , the PU is assumed to be present, otherwise the channel is declared idle. A widely used decision-metric for spectrum sensing is energy detection, with [6]

$$M = \frac{1}{N_s} \sum_{n=1}^{N_s} |r(n)|^2, \quad (1)$$

where  $r(n)$  is the received signal. Other metrics to detect the presence of the PU are not considered in this paper, however, the findings can be readily extended.

Based on the decision metric, the probabilities of detection,  $P_d$ , and false-alarm,  $P_f$ , can be found by applying appropriate hypothesis tests [6], [7]. Denoting  $\mathcal{H}_0$  as the scenario where the PU is inactive, and  $\mathcal{H}_1$  where the PU is active, the probabilities of false alarm and detection are defined as

$$P_f = Pr(M > \epsilon | \mathcal{H}_0) \quad (2)$$

$$P_d = Pr(M > \epsilon | \mathcal{H}_1). \quad (3)$$

The probability of misdetection,  $P_m$  is defined as

$$P_m = Pr(M < \epsilon | \mathcal{H}_1) = 1 - P_d. \quad (4)$$

### A. Half-Duplex Cognitive Radio

Fig. 1 shows the SU transmission frame consisting of  $N$  samples for a conventional half-duplex cognitive radio system [7]. The frame consists of two parts: an  $N_s$  sample sensing window and an  $N - N_s$  sample transmission window. The  $N_s$  samples are used to decide if the PU is present based on the decision metric (e.g., (1)), with transmission continuing if the PU is not detected. Decisions are thus made every  $N$  samples (i.e., once per frame). It is important to note that the SU is unable to detect or react to the presence of the PU while transmitting, leading to a high access-latency. Due to the sensing window, the throughput of the SU system is also reduced by a factor of  $\frac{N - N_s}{N}$ .

### B. Concurrent Sensing and Transmission With Slotted Sensing

By cancelling the self-interference signal, SUs can concurrently transmit and sense the channel. Fig. 1 shows the frame structure of a concurrent sensing and transmission system using self-interference cancellation as proposed in [16] and [14]. The systems analysed in [16] retain the same frame structure as a conventional CR system by including a sensing-only slot at the start of each frame. In this paper, we consider full-duplex systems that only include concurrent transmission and sensing slots, similar to the approach taken in [14], [17]. The SU throughput is not reduced as there is no dedicated sensing-slot. Similarly, the SU may be able to more quickly detect whether the PU starts transmitting during the frame, thereby reducing the access-latency. While decisions are made more frequently (every  $N_s$  samples) than in a conventional cognitive radio system (every  $N$  samples), there still remains a blind interval of  $N_{first}$  samples. Residual self-interference also makes it more difficult to detect the presence of the PU, by effectively increasing the noise-floor. However, as the detection decisions are made more frequently compared to the half-duplex case, lower values of the  $P_d$  may still result in an acceptable access latency.

### C. Concurrent Sensing and Transmission with Sliding Window Sensing

We propose an extension to the concurrent sensing and transmission scheme outlined in the previous section by introducing a sliding-window to take decisions at every sample, i.e., the buffer does not wait to fill with  $N_s$  fresh samples before a decision is made. This approach can be implemented easily in digital hardware via a FIFO buffer. It is important to note that successive decision metrics are *not* independent as only one new sample is added (and one removed). However, the minimum access-latency of this scheme is one sample, unlike the slotted full-duplex method. In our analysis it is assumed that the number of sensing samples,  $N_s$ , remains the same as the slotted full-duplex model.

### III. LATENCY-THROUGHPUT ANALYSIS

#### A. Throughput of the Half-Duplex and Full-Duplex Systems

The capacity of the SU (assuming the PU is not transmitting) is given by

$$C_0 = \log_2(1 + \text{SNR}_{\text{SU}}), \quad (5)$$

where  $\text{SNR}_{\text{SU}}$  is the signal-to-noise ratio of the SU measured at a receiver node. This capacity is achieved when the PU is not transmitting *and* the SU has not raised a false alarm.

Following [7], [14], we express the total throughput for the SU system with full-duplex sensing capability as

$$R_{\text{FD}} = C_0(1 - P_f). \quad (6)$$

For a specific SNR of the SU, the throughput is a parameter of only  $P_f$ . In practice, if a false-alarm occurs, the loss in SU throughput can be significant as an entire data frame may be lost, however the impact of these intermittent SU outages can be mitigated by coding.

The half-duplex system has an additional throughput loss due to the sensing overhead

$$R_{\text{HD}} = \frac{N - N_s}{N} C_0(1 - P_f). \quad (7)$$

Clearly for the half-duplex system there exists a trade-off between detection latency and throughput that is determined by the sensing overhead. For full-duplex systems, the main consideration is the increase of  $P_f$ , due to the residual self-interference. In order for a full-duplex system with residual self-interference to maintain a throughput close to  $C_0$ , it has to operate with lower  $P_d$  values. In the results we show that the sliding window model allows for lower access-latency compared to the slotted full-duplex and half-duplex cases, while maintaining a high SU throughput.

#### B. Average Latency of the Half-Duplex System

Let us denote  $D_k$  as the event that the PU is detected during the  $k$ -th decision after starting transmission, and  $D_k^c$  the complementary event that the PU is not detected during the  $k$ -th decision. Let also  $N_k$  be the number of samples from when the PU starts to the  $k$ -th decision point. The average access-latency for all the schemes can thus be computed from the infinite sum

$$L = N_1 P(D_1) + N_2 P(D_2 | D_1^c) + N_3 P(D_3 | D_1^c \cap D_2^c) + \dots + N_i P(D_i | D_1^c \cap D_2^c \dots \cap D_{i-1}^c) + \dots \quad (8)$$

For the half-duplex case, there are two possible scenarios, depending on which part of the SU activity (sensing or transmitting) the PU starts transmitting. In the first scenario the PU starts transmitting during the blind interval,  $N_{\text{blind}}$  samples before the end of the blind interval. The first decision is made after  $(N_{\text{blind}} + N_s)$  samples, i.e.,  $N_1 = (N_{\text{blind}} + N_s)$ . Let  $P_d(k)$  denote the probability of detection with  $k$  samples of the PU signal and  $P_m(k)$  denote the probability of misdetection for the same case; the probability of detecting the PU at the first decision is thus  $P_d(N_s)$ . If the PU is not

detected, the number of samples for the second decision will be  $N_2 = (N_{\text{blind}} + N_s + N)$ , again with probability of detection  $P_d(N_s)$ , and so on. Thus

$$\begin{aligned} L(N_{\text{blind}}) &= P_d(N_s) \sum_{n=0}^{\infty} (N_{\text{blind}} + N_s + nN) P_m^n(N_s) \\ &= P_d(N_s) \left[ (N_{\text{blind}} + N_s) \frac{1}{P_d(N_s)} + N \frac{P_m(N_s)}{P_d^2(N_s)} \right], \end{aligned} \quad (9)$$

for  $0 < P_d(N_s) < 1$ .

$N_{\text{blind}}$  can take each value between 1 and  $N - N_s$  with probability  $\frac{1}{N}$ , thus the average latency under scenario 1 is

$$L_1 = \frac{1}{N} \sum_{N_{\text{blind}}=1}^{N-N_s} L(N_{\text{blind}}). \quad (10)$$

In the second scenario the PU starts transmitting during the sensing period,  $N_{\text{first}}$  samples before the end of the sensing period. The first decision is taken after  $N_1 = N_{\text{first}}$  samples, with probability of detection  $P_d(N_{\text{first}})$ . If the PU is not detected, the number of samples for the second decision is  $N_2 = (N_{\text{first}} + N)$ , with probability of detection  $P_d(N_s)$ , and so on. Thus

$$\begin{aligned} L(N_{\text{first}}) &= N_{\text{first}} P_d(N_{\text{first}}) + P_d(N_s) P_m(N_{\text{first}}) \times \\ &\quad \sum_{n=0}^{\infty} [N_{\text{first}} + (n+1)N] P_m^n(N_s) \\ &= N_{\text{first}} P_d(N_{\text{first}}) + P_d(N_s) P_m(N_{\text{first}}) \times \\ &\quad \left[ N_{\text{first}} \frac{1}{P_d(N_s)} + N \frac{1}{P_d^2(N_s)} \right] \end{aligned} \quad (11)$$

for  $0 < P_d(N_s) < 1$ .

As  $N_{\text{first}}$  can take any value between 1 and  $N_s$ , with probability  $\frac{1}{N}$ , the average latency under scenario 2 is

$$L_2 = \frac{1}{N} \sum_{N_{\text{first}}=1}^{N_s} L(N_{\text{first}}). \quad (12)$$

Therefore, the total average latency of the half-duplex system is  $L = L_1 + L_2$ .

#### C. Average Latency of the Slotted Full-Duplex System

In the slotted full-duplex scheme there is no blind interval. Let us thus consider the case where the PU turns on  $N_{\text{first}}$  samples before the end of the sensing period. The first decision is made after  $N_1 = N_{\text{first}}$  samples, with probability of detection  $P_d(N_{\text{first}})$ . If the PU is not detected, the second decision has

latency of  $N_2 = (N_{\text{first}} + N_s)$ , with  $P_d(N_s)$ , and so on. Thus the latency is

$$\begin{aligned} L(N_{\text{first}}) &= N_{\text{first}} P_d(N_{\text{first}}) + P_d(N_s) P_m(N_{\text{first}}) \times \\ &\quad \sum_{n=0}^{\infty} [N_{\text{first}} + (n+1)N_s] P_m^n(N_s) \\ &= N_{\text{first}} P_d(N_{\text{first}}) + P_d(N_s) P_m(N_{\text{first}}) \times \\ &\quad \left[ N_{\text{first}} \frac{1}{P_d(N_s)} + N_s \frac{1}{P_d^2(N_s)} \right], \end{aligned} \quad (13)$$

for  $0 < P_d(N_s) < 1$ .

$N_{\text{first}}$  can take any value between 1 and  $N_s$  with probability  $\frac{1}{N_s}$ , which leads to an average latency

$$L = \frac{1}{N_s} \sum_{N_{\text{first}}=1}^{N_s} L(N_{\text{first}}). \quad (14)$$

#### D. Average Latency of the Sliding Full-Duplex System

In the sliding full-duplex model, the average latency can be derived from (8), however the difference in this case is that the decisions are not independent. A decision taken at a sample  $i$  is not independent of decisions taken over the previous  $i + (N_s - 1)$  samples. Only decisions separated by  $N_s$  samples are independent. Accordingly, the slotted model can be regarded as a special case of the sliding model, where the only decisions kept are those separated by  $N_s$  samples. The idea of the sliding approach is that there is no real reason to discard all the decisions in between, as these will potentially reduce the access-latency. It should also be noted that taking decisions every sample is feasible in contemporary systems. However, the effect on the energy consumption, which may be a limitation for battery-driven SUs, remains to be investigated.

Since each decision is not independent of the  $N_s$  previous decisions, the conditional terms of (8) cannot be easily expanded using multiplicative terms as in (11) and (13). The analytical expression for the average access-latency in this case would require the conditional terms to be expressed using a stochastic process model with memory, and is outside the scope of this work. To compare the performance of our proposed sliding-window method with the existing schemes, we have used Monte-Carlo simulations which maintain the dependence between the decisions.

## IV. RESULTS AND DISCUSSION

In Fig. 2 we observe a significant improvement in the average latency of the full-duplex schemes compared to the conventional half-duplex scheme. Moreover an improvement is observed for the sliding full-duplex model compared to the slotted scheme. For these results, no residual self-interference is present, i.e., perfect self-interference suppression is assumed. The minimum average access-latency of the half-duplex scheme is approximately half the length of a SU frame, as the system is unable to sense while transmitting. The slotted full-duplex system has lower access-latency—as decisions are made every  $N_s$  samples—and the minimum

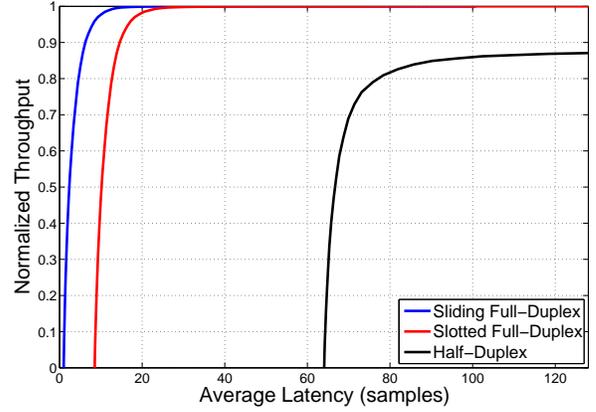


Fig. 2. Average latency - normalized throughput comparison for the 3 systems,  $\text{SNR}_{\text{PU}} = 0$  dB with perfect self-interference suppression,  $N_s = 16$ ,  $N = 128$ .

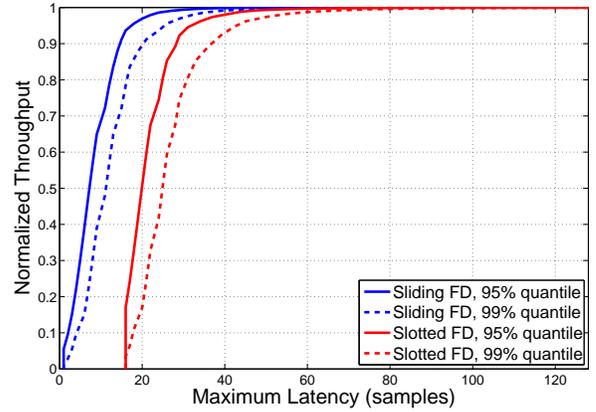


Fig. 3. Maximum acceptable latency - normalized throughput comparison for the 95% and 99% quantiles for the 3 systems, primary user  $\text{SNR}_{\text{PU}} = 0$ ,  $N_s = 16$ ,  $N = 128$ .

average latency is approximately  $\frac{N_s}{2}$ . The sliding-window full-duplex scheme can potentially detect the PU with an average access-latency of a single sample, however, this mode of operation results in very low SU throughput (and is thus not viable in normal circumstances). Similarly, the maximum throughput is achieved for all three schemes—the half-duplex system has a reduced maximum throughput due to the sensing overhead, as given by (7)—as the average latency increases.

Fig. 3 shows the 95% and 99% quantiles of the latency-throughput curves for both full-duplex schemes. The access-latency is a random variable, which depends on the actual realisations of the signal, noise and self-interference. Thus, in order to provide useful metrics of the system performance we examine the maximum access-latency that is reached with a specific confidence.

Fig. 4 compares the average latency for the two full-duplex models with increasing residual self-interference (measured relatively to the noise-floor). The average latency is found by normalizing the throughput to 0.9, i.e.,  $P_f = 0.1$ . The sliding

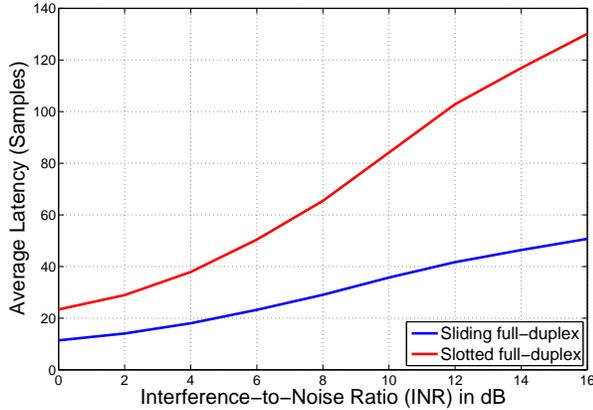


Fig. 4. Average access-latency with increasing residual self-interference, for a normalised throughput of 0.9,  $\text{SNR}_{\text{PU}} = 0$  dB,  $N_s = 16$ ,  $N = 128$ .

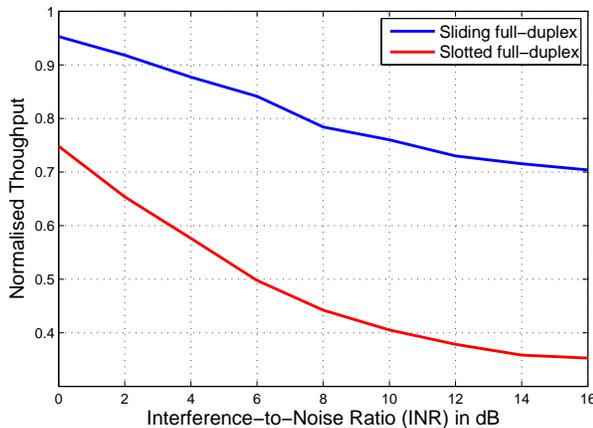


Fig. 5. Normalised throughput with increasing residual self-interference, for an average access-latency of 16 samples,  $\text{SNR}_{\text{PU}} = 0$  dB,  $N_s = 16$ ,  $N = 128$ .

scheme is observed to have lower latency for all values of residual self-interference considered. Moreover, the difference in the slopes indicates that the sliding scheme is more resistant to residual self-interference, by approximately a factor of 2.7 relative to the slotted scheme.

In Fig. 5 we can observe a similar result from a different perspective. In this figure we compare the normalised throughput for the two full-duplex models with increasing residual self-interference. The throughput is found for the same latency of  $L = N_s$  samples for both schemes, i.e., for the same PU protection. The sliding scheme is observed to have higher throughput for all values of residual self-interference considered, again with a difference in the slopes.

## V. CONCLUSIONS

In this paper we have elaborated on the analysis of cognitive radios from the perspective of protecting the primary or high-priority users by reducing the access-latency of the system. By deriving analytical formulas for the access-latency of existing schemes, both half-duplex and full-duplex, we have quantified

the access-latency problem, and showed that it is particularly significant for the half-duplex model and that it is reduced by the slotted full-duplex model. In order to overcome the latency problem even more effectively, we proposed a sliding full-duplex scheme that profits from its ability to take decisions every sample. The results show that there is a significant improvement in the response performance of the system by using full-duplex techniques. The problem of residual self-interference that exists in every full-duplex system was also considered. The proposed sliding method was proven to resist to the residual self-interference 2.7 times more effectively than the slotted scheme.

## REFERENCES

- [1] Spectrum Policy Task Force, "Report of the spectrum efficiency working group," Federal Communications Commission, Tech. Rep., Nov. 2002.
- [2] R. I. C. Chiang, G. B. Rowe, and K. W. Sowerby, "A quantitative analysis of spectral occupancy measurements for cognitive radio," in *Proc. IEEE Vehicular Technology Conf. (Spring)*, Apr. 2007, pp. 3016–3020.
- [3] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.
- [4] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [5] C. Dewar and D. Warren, "Understanding 5G: Perspectives on future technological advancements in mobile," GSMA Intelligence, Tech. Rep., Dec. 2014. [Online]. Available: <https://gsmaintelligence.com/research/?file=141208-5g.pdf>
- [6] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, Mar. 2009.
- [7] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.
- [8] W.-Y. Lee and I. F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3845–3857, Oct. 2008.
- [9] A. Sabharwal, P. Schniter, D. Guo, D. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sept. 2014.
- [10] E. Ahmed, A. Eltawil, and A. Sabharwal, "Simultaneous transmit and sense for cognitive radios using full-duplex: A first study," in *Proc. IEEE Antennas Propag. Int. Symp.*, Jul. 2012.
- [11] W. Afifi and M. Krunz, "Exploiting self-interference suppression for improved spectrum awareness/efficiency in cognitive radio systems," in *IEEE INFOCOM*, Apr. 2013, pp. 1258–1266.
- [12] T. Riihonen and R. Wichman, "Energy detection in full-duplex cognitive radios under residual self-interference," in *Proc. Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, Jun. 2014, pp. 57–60.
- [13] W. Afifi and M. Krunz, "Adaptive transmission-reception-sensing strategy for cognitive radios with full-duplex capabilities," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, Apr. 2014, pp. 149–160.
- [14] Y. Liao, T. Wang, L. Song, and Z. Han, "Listen-and-talk: Full-duplex cognitive radio networks," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, Dec. 2014, pp. 3068–3073.
- [15] A. Balatsoukas-Stimming, A. C. M. Austin, P. Belanovic, and A. P. Burg., "Baseband and RF hardware impairments in full-duplex wireless systems: experimental characterisation and suppression," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 142, May 2015.
- [16] W. Afifi and M. Krunz, "Incorporating self-interference suppression for full-duplex operation in opportunistic spectrum access systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2180–2191, Apr. 2015.
- [17] Y. Liao, L. Song, Z. Han, and Y. Li, "Full duplex cognitive radio: a new design paradigm for enhancing spectrum usage," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 138–145, May 2015.