Alexander von Humboldt Professorship Research Project:

Foundations and Architectures for the Next Generation of Wireless Networks

Progress Report (April 2015 – March 2016)

Giuseppe Caire

Receiving institutions: Technische Universität Berlin
Fraunhofer Institut für Nachrichtentechnik (Heinrich-Hertz-Institut)

1 Research

1.1 Publications

This is a list of publications during the second year of the AvH project:

1. Vishnu V. Ratnam; Andreas F. Molisch; Giuseppe Caire, "Capacity Analysis of Interlaced Clustering in a Distributed Transmission System With/Without CSIT," *IEEE Transactions on Wireless Communications* Year: 2016, Volume: 15, Issue: 4, Pages: 2629 - 2641.

2. M. Ji; G. Caire; A. F. Molisch "Wireless Device-to-Device Caching Networks: Basic Principles and System Performance," IEEE Journal on Selected Areas in Communications Year: 2016, Volume: 34, Issue: 1, Pages: 176 - 189.

3. M. Ji; G. Caire; A. F. Molisch "Fundamental Limits of Caching in Wireless D2D Networks," IEEE Transactions on Information Theory Year: 2016, Volume: 62, Issue: 2, Pages: 849 - 869.

4. D. Bethanabhotla; G. Caire; M. Neely "WiFlix: Adaptive Video Streaming in Massive MU-MIMO Wireless Networks," IEEE Transactions on Wireless Communications Year: 2016, Volume: PP, Issue: 99, Pages: 1 - 1.

5. M. Dai; B. Clerckx; D. Gesbert; G. Caire "A Rate Splitting Strategy for Massive MIMO with Imperfect CSIT," IEEE Transactions on Wireless Communications Year: 2016, Volume: PP, Issue: 99, Pages: 1 - 1.

6. D. Bethanabhotla; O. Y. Bursalioglu; H. C. Papadopoulos; G. Caire "Optimal User-Cell Association for Massive MIMO Wireless Networks," IEEE Transactions on Wireless Communications, Year: 2016, Volume: 15, Issue: 3, Pages: 1835 - 1850.

7. K. Mahler; W. Keusgen; F. Tufvesson; T. Zemen; G. Caire "Tracking of Wideband Multipath Components in a Vehicular Communication Scenario," IEEE Transactions on Vehicular Technology, Year: 2016, Volume: PP, Issue: 99, Pages: 1 - 1.

   item Mingyue Ji; Giuseppe Caire; Andreas F. Molisch, "The Throughput-Outage Tradeoff of Wireless One-Hop Caching Networks," *IEEE Transactions on Information Theory* Year: 2015, Volume: 61, Issue: 12, Pages: 6833 - 6859.

8. V. Ntranos; M. A. Maddah-Ali; G. Caire "Cellular Interference Alignment: Omni-Directional Antennas and Asymmetric Configurations," IEEE Transactions on Information Theory, Year: 2015, Volume: 61, Issue: 12, Pages: 6663 - 6679.

9. Wei Yang; Caire, G.; Durisi, G.; Polyanskiy, Y. "Optimum Power Control at Finite Blocklength," IEEE Transactions on Information Theory, Year: 2015, Volume: 61, Issue: 9 Pages: 4598 - 4615

10. Song-Nam Hong; Caire, G. "Beyond Scaling Laws: On the Rate Performance of Dense Device-to-Device Wireless Networks," IEEE Transactions on Information Theory, Year: 2015, Volume: 61, Issue: 9 Pages: 4735 - 4750

11. Adhikary, A.; Dhillon, H.S.; Caire, G. "Massive-MIMO Meets HetNet: Interference Coordination Through Spatial Blanking," IEEE Journal on Selected Areas in Communications, Year: 2015, Volume: 33, Issue: 6 Pages: 1171 - 1186

12. M. Ji; G. Caire; A. F. Molisch "Wireless Device-to-Device Caching Networks: Basic Principles and System Performance," IEEE Journal on Selected Areas in Communications Year: 2016, Volume: 34, Issue: 1, Pages: 176 - 189.

13. Lim, Y.; Chae, C.; Caire, G. "Performance Analysis of Massive MIMO for Cell-Boundary Users," IEEE Transactions on Wireless Communications, Year: 2015, Volume: PP, Issue: 99 Pages: 1 - 1

14. Miller, K.; Bethanabhotla, D.; Caire, G.; Wolisz, A. "A Control-Theoretic Approach to Adaptive Video Streaming in Dense Wireless Networks," IEEE Transactions on Multimedia, Year: 2015, Volume: 17, Issue: 8 Pages: 1309 - 1322

15. Song-Nam Hong; Caire, G. "Virtual Full-Duplex Relaying With Half-Duplex Relays," IEEE Transactions on Information Theory, Year: 2015, Volume: 61, Issue: 9 Pages: 4700 - 4720

16. Kim, J.; Caire, G.; Molisch, A.F. "Quality-Aware Streaming and Scheduling for Device-to-Device Video Delivery," IEEE/ACM Transactions on Networking, Year: 2015, Volume: PP, Issue: 99 Pages: 1 - 1

17. Dhillon, H.S.; Caire, G. "Wireless Backhaul Networks: Capacity Bound, Scalability Analysis and Design Guidelines," IEEE Transactions on Wireless Communications, Year: 2015, Volume: PP, Issue: 99 Pages: 1 - 1

18. Song-Nam Hong; Caire, G. "Structured Lattice Codes for Some Two-User Gaussian Networks With Cognition, Coordination, and Two Hops," IEEE Transactions on Information Theory, Year: 2015, Volume: 61, Issue: 5 Pages: 2624 - 2646,

19. K. Mahler; W. Keusgen; F. Tufvesson; T. Zemen; G. Caire "Propagation of Multipath Components at an Urban Intersection," Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd Year: 2015, Pages: 1 - 5.

20. A. S. Tehrani; A. F. Molisch; G. Caire "Directional ZigZag: Neighbor Discovery with Directional Antennas," 2015 IEEE Global Communications Conference (GLOBECOM) Year: 2015, Pages: 1 - 6.

21. Kurras, Martin and Thiele, Lars and Caire, Giuseppe, "Interference Mitigation and Multiuser Multiplexing with Beam-Steering Antennas," *19th International ITG Workshop on Smart Antennas WSA 2015*, Ilmenau, Germany, April 2015.

22. Mingyue Ji; Shanmugam, K.; Vettigli, G.; Llorca, J.; Tulino, A.M.; Caire, G. “An efficient multiple-groupcast coded multicasting scheme for finite fractional caching,” Communications (ICC), 2015 IEEE International Conference on Year: 2015, Pages: 3801 - 3806

23. Yongpeng Wu; Schober, R.; Ng, D.W.K.; Chengshan Xiao; Caire, G. “Secure Massive MIMO transmission in the presence of an active eavesdropper,” Communications (ICC), 2015 IEEE International Conference on Year: 2015, Pages: 1434 - 1440

24. Kittichokechai, Kittipong; Caire, Giuseppe “Secret key-based authentication with a privacy constraint,” Information Theory (ISIT), 2015 IEEE International Symposium on Year: 2015, Pages: 1791 - 1795

25. Sang-Woon Jeon; Song-Nam Hong; Mingyue Ji; Caire, G. “On the capacity of multihop device-to-device caching networks,” Information Theory Workshop (ITW), 2015 IEEE Year: 2015, Pages: 1 - 5

26. Yi, Xinping; Caire, Giuseppe “On the optimality of treating interference as noise: A combinatorial optimization perspective,” Information Theory (ISIT), 2015 IEEE International Symposium on Year: 2015, Pages: 1721 - 1725

27. Song-Nam Hong; Maric, I.; Hui, D.; Caire, G. “Multihop virtual full-duplex relay channels,” Information Theory Workshop (ITW), 2015 IEEE Year: 2015, Pages: 1 - 5

28. Ratnam, V.V.; Caire, G.; Molisch, A.F. “Capacity analysis of interlaced clustering in a distributed antenna system,” Communications (ICC), 2015 IEEE International Conference on Year: 2015, Pages: 1727 - 1732

29. Mingyue Ji; Tulino, A.; Llorca, J.; Caire, G. “Caching-aided coded multicasting with multiple random requests,” Information Theory Workshop (ITW), 2015 IEEE Year: 2015, Pages: 1 - 5

30. Sang-Woon Jeon; Song-Nam Hong; Mingyue Ji; Caire, G. “Caching in wireless multihop device-to-device networks,” Communications (ICC), 2015 IEEE International Conference on Year: 2015, Pages: 6732 - 6737

31. Ntranos, Vasilis; Ali, Mohammad-Ali Maddah; Caire, Giuseppe “Cooperation alignment for distributed interference management,” Information Theory (ISIT), 2015 IEEE International Symposium on Year: 2015, Pages: 874 - 878

32. Mingyue Ji; Ming Fai Wong; Tulino, A.M.; Llorca, J.; Caire, G.; Effros, M.; Langberg, M. “On the fundamental limits of caching in combination networks,” Signal Processing Advances in Wireless Communications (SPAWC), 2015 IEEE 16th International Workshop on Year: 2015, Pages: 695 - 699

1.2 Information theoretic study of caching wireless networks

Data traffic generated by wireless and mobile devices is predicted to increase by something between one and two orders of magnitude [1] in the next five years, mainly due to wireless video streaming. Traditional methods for increasing the area spectral efficiency, such as using more spectrum and/or deploying more base stations, are either insufficient to provide the necessary wireless throughput increase, or are too expensive. Thus, exploring alternative strategies that leverage different and cheaper network resources is of great practical and theoretical interest.

The bulk of wireless video traffic is due to asynchronous *video on demand*, where users request video files from some library (e.g., iTunes, Netflix, Hulu or Amazon Prime) at arbitrary times. This type of traffic differs significantly from *live streaming*. The latter is essentially a lossy multicasting problem, for which the broadcast nature of the wireless channel can be naturally exploited (see for example [2, 3, 4, 5, 6]). The theoretical foundation of schemes for live streaming relies on well-known information theoretic settings for one-to-many transmission of a common message with possible refinement information, such as successive refinement [7, 8, 9] or multiple description coding [10, 11, 12].

In contrast, the *asynchronous* nature of video on demand prevents from taking advantage of multicasting, despite the significant overlap of the requests (people wish to watch a few very popular files). Hence, even though users keep requesting the same few popular files, the asynchronism of their requests is large with respect to the duration of the video itself, such that the probability that a single transmission from the base station is useful for more than one user is *essentially zero*. Due to this reason, current practical implementation of video on demand over wireless networks is handled at the application layer, requiring a dedicated data connection (typically TCP/IP) between each client (user) and the server (base station), for each streaming user, as if users were requesting independent information.

One of the most promising approaches to take advantage of the inherent *asynchronous content reuse* is *caching*, widely used in content distribution networks over the (wired) Internet [13]. In [14, 15], the idea of deploying dedicated "helper nodes" with large caches, that can be refreshed via wireless at the cellular network off-peak time, was proposed as a cost-effective alternative to providing large capacity wired backhaul to a network of densely deployed small cells. An even more radical view considers caching directly at the wireless users, exploiting the fact that modern devices have tens and even hundreds of GBytes of largely under-utilized storage space, which represents an enormous, cheap and yet untapped network resource.

Recently, a *coded multicasting* scheme exploiting caching at the user nodes was proposed in [16]. In this scheme, the files in the library are divided in blocks (packets) and users cache carefully designed subsets of such packets. Then, for given set of user demands, the base station sends to all users (multicasting) a common codeword formed by a sequence of packets obtained as linear combinations of the original file packets. As noticed in [16], coded multicasting can handle any form of asynchronism by suitable sub-packetization. Hence, the scheme is able to create multicasting opportunities through coding, exploiting the overlap of demands while eliminating the asynchronism problem. For the case of arbitrary (adversarial) demands, the coded multicasting scheme of [16] is shown to perform within a small gap, independent of the number of users, of the cache size and of the library size, from the cut-set bound of the underlying compound channel.[1] However, the scheme has some significant drawbacks that makes it not easy to be implemented in practice: 1) the construction of the caches is combinatorial and the sub-packetization explodes exponentially with the library size and number of users; 2) changing even a single file in the library requires a significant reconfiguration of the user caches, making the cache update difficult. In [17], similar near-optimal performance of coded caching is shown to be achieved also through a *random caching scheme*, where each user caches a random selection of bits from each file in the library. In this case, though, the combinatorial complexity of the coded caching scheme is transferred from the caching phase to the (coded) delivery phase, where the construction of the multicast codeword requires solving multiple clique cover problems with fixed clique size (known to be NP-complete), for which a greedy algorithm is shown to be efficient.

Our contributions in 2015-2016: We have extended our approach of random independent caching at the user nodes and device-to-device (D2D) communication to the multi-hop case, where users can retrieve the desired content from other nodes which are not immediately reachable in one hop, but in multiple hops through other nodes which operate as relays.

Main results:

[1] The compound nature of this model is due to the fact that the scheme handles adversarial demands.

- The throughput scaling of D2D multihop caching networks has been characterized in our recent works:

  1. Sang-Woon Jeon, Song-Nam Hong, Mingyue Ji, and Giuseppe Caire, “On the Capacity of Multihop Device-to-Device Caching Networks,” *IEEE Information Theory Workshop*, Jerusalem, Israel, April 26 - Mary 1, 2015.
  2. Jeon, Sang-Woon, Song-Nam Hong, Mingyue Ji, and Giuseppe Caire. ”Caching in wireless multihop device-to-device networks.” In Communications (ICC), 2015 IEEE International Conference on, pp. 6732-6737. IEEE, 2015.
  3. M. Ji; G. Caire; A. F. Molisch “Wireless Device-to-Device Caching Networks: Basic Principles and System Performance,” IEEE Journal on Selected Areas in Communications Year: 2016, Volume: 34, Issue: 1, Pages: 176 - 189.

1.3 Network optimization, dynamic scheduling and video quality control for video streaming over wireless networks

With the proliferation of mobile devices and services, industry predicts that the wireless data traffic is going to increase by two to three orders of magnitude within a decade [18]. Although the definition of the next generation of systems and standards is at its initial phase, it is widely agreed that the next generation of wireless networks, generally referred to as “5G”, will involve a combination of multiuser MIMO technology, cell densification, and heterogeneous architectures based on nested tiers of smaller and smaller cells operating at higher and higher frequencies, in order to target traffic hot-spots [19]. These trends have motivated the recent surge of research on massive and dense deployment of base station antennas, both in the form of *Massive MIMO* schemes, with hundreds of antennas at each cell site [20, 21, 22], and in the form of multi-tier networks of densely deployed small-cells [23, 24].

Massive MIMO promises dramatic increases in spectral efficiency by transmitting independent data streams simultaneously to multiple users sharing the same transmission resource (time-frequency slot). The massive MIMO regime [20, 21, 22] distinguishes itself from classical multiuser MIMO [25, 26] by the fact that the number of served users is significantly less than the (very large) number of base station antennas. Operating in *Time-Division Duplexing* (TDD) mode, massive MIMO can provide very large spectral efficiencies, simple per-cell processing, and very attractive power efficiency due to the large array gain [20]. Thanks to the higher and higher carrier frequencies [27], it is possible to implement massive MIMO even in relatively small base stations within a reasonable form factor. Hence, it is envisaged that massive MIMO will not just be applied to large tower-mounted base stations, but also used in conjunction with small cells [28].

The heterogeneous wireless network framework mentioned above may include some of the following features: 1) base stations that may differ significantly by transmit power, number of antennas, and multiplexing gain (e.g., see [29] and references therein); 2) non-homogeneous user spatial distribution, characterized by high-density hot-spots separated by less dense regions [30]; 3) Due to the large beamforming gain of massive MIMO, a user may be in good SINR conditions with respect to several base stations. As a consequence, the rationale that has driven for decades the conventional cellular system design and optimization, based on symmetric lattice-deployed cells (see for example [20, 21, 22]) and/or (roughly) uniform number of users per cell (e.g., see [31] and references therein), must be abandoned in favor of more efficient schemes that include user-cell association into the optimization problem.

Our work focuses on the design of a scheduling policy for VoD streaming in a wireless network formed by many users and helpers, deployed over a localized geographic area and sharing the same channel bandwidth. We focus on the wireless segment of the network, assuming that the video files are already present at the helper nodes. This condition holds when the backhaul connecting the helper nodes to some video server in the core network is fast enough, such that we can neglect the delays introduced by the backhaul. In the case where such fast backhaul is not present, the recently proposed approach of caching at the wireless edge (see [32, 33, 34, 35, 36, 37, 38, 17, 39, 40, 41, 42]) was shown to be able to exploit the inherent asynchronous content reuse of VoD in order to predict and proactively store the popular video files such that, with high probability, the demanded files are effectively already present in the helpers caches. This justifies our assumption of neglecting the effects of the wired backhaul and focusing only on the wireless segment of the system.

In order to obtain a tractable formulation, we follow a "divide and conquer" approach, conceptually organized in the following steps:

i) We formulate a Network Utility Maximization (NUM) problem [43, 44, 45] where the network utility function is a concave and componentwise non-decreasing function of the time-averaged users' *requested* video quality index and the maximization is subject to the stability of all queues in the system. The shape of the network utility function can be chosen in order to enforce some desired notion of fairness across the users [46].

ii) We solve the NUM problem in the framework of Lyapunov Optimization [47], using the *drift plus penalty* (DPP) approach [47]. The obtained solution is provably asymptotically optimal (with respect to the defined NUM problem) on a per-sample path sense (i.e., without assuming stationarity and ergodicity of the underlying network state process [47, 48]). Furthermore, it naturally decomposes into sub-policies that can be implemented in a distributed way, by functions performed at the users and the helpers, requiring only *local* information. The function implemented at the user nodes is referred to as *congestion control,* since it consists of the adaptive selection of the video quality and the serving helper. The function implemented at the helpers is referred to as *transmission scheduling,* since it corresponds to the adaptive selection of the user to be served on the downlink of each helper station.

iii) We observe that, since all queues in the system are stable, all *requested* video chunks shall be eventually *delivered*.

iv) As a consequence, in order to ensure that all the video chunks are delivered within their playback deadline, it is sufficient to ensure that the largest delay among all queues at the helpers serving any given user is not larger than the pre-buffering time allowed for that user at its streaming session startup phase. We refer to the event that a chunk is not delivered within its playback deadline as a *buffer underrun* event. Since such events are perceived as very harmful for the overall quality of the streaming session, the system must operate in the regime where the relative fraction of such chunks (referred to as *buffer underrun rate*) is small. We refer to such desirable regime as the *smooth streaming regime*. In particular, when the maximum delay of each queue in the system admits a deterministic upper bound (e.g., see [49]), setting the pre-buffering time larger than such bound makes the underrun rate equal to zero. However, for a system with arbitrary user mobility, arbitrary per-chunk fluctuations of the video coding rate (as in typical Variable Bit-Rate (VBR) coding [50]), and users joining or leaving the system at arbitrary times, such deterministic delay upper bounds do not exist. Hence, in order to make the system operate in the smooth streaming regime, we propose a method to locally estimate the delays with which the video packets are delivered, such that each user can calculate its pre-buffering and re-buffering time to be larger than the locally estimated maximum queue delay. Through simulations, we demonstrate that the combination of our scheduling policy and adaptive pre-buffering scheme is able to achieve the desired fairness across the users and, at the same time, very small playback buffer underrun rate.

Since the proposed policy achieves NUM optimality on a per-sample path basis and, thanks to the adaptive dimensioning of the users' pre-buffering time, the system operates in the regime of small buffer underrun rate (i.e., in the smooth streaming regime), the resulting system performance is near-optimal in the following sense: for any *bounded* penalty weight assigned to the buffer underrun events, the system network utility (including such penalty) is just a small perturbation away from the optimal NUM value obtained by our DPP policy.

Our contributions in 2015-2016: We have extended and improved our original NUM-based joint transmission scheduling and dynamic video quality selection approach to a new scheme based on a single request queue per user and a "pull" strategy, which guarantees that all video chunks are received in the right order even though they are pulled from different helper stations. Furthermore, we have continued the collaboration with prof. Adam Wolisz and his PhD student, Kostantin Miller, pursuing the "control theoretic approach" as an alternative to the Lyapunov dynamic optimization approach.

Main results: Our contribution are collected in the two published papers:

1. D. Bethanabhotla; G. Caire; M. Neely "WiFlix: Adaptive Video Streaming in Massive MU-MIMO Wireless Networks," IEEE Transactions on Wireless Communications Year: 2016, Volume: PP, Issue: 99, Pages: 1 - 1.

2. Miller, K.; Bethanabhotla, D.; Caire, G.; Wolisz, A. "A Control-Theoretic Approach to Adaptive Video Streaming in Dense Wireless Networks," IEEE Transactions on Multimedia, Year: 2015, Volume: 17, Issue: 8 Pages: 1309 - 1322

   (this paper has been nominated for the best paper award in the IEEE Transactions on Multimedia (we still do not know if the paper will get the prize).

1.4 Optimization and implementation of massive MIMO schemes

Massive MIMO (multiple-input multiple-output) systems are equipped with a large number (dozens or hundreds) of antenna elements at the base station (BS) [20, 51]. They are intended to be employed in a multi-user MIMO (MU-MIMO) setting, such that the number of BS antenna elements is much larger than the number of users. Such an arrangement leads not only to very high spectral efficiency, but also to an important simplification of the signal processing: in the idealized regime of independent and isotropically distributed channel vectors, in the limit of an infinite number of BS antennas, single-user beamforming, specifically conjugate beamforming (i.e., maximum ratio combining in the receive mode, and maximum ratio transmission for the transmit mode) eliminates inter-user interference. Furthermore, the transmit power can be drastically reduced, leading to less interference and a lower energy consumption of the BS. For all these reasons, massive MIMO has received tremendous attention in the last years [52, 53, 54, 55, 56].

Massive MIMO is especially promising for systems operating at millimeter (mm-) Wave frequencies. Due to the short wavelength, very large arrays can be created with a reasonable form factor - a 100-element linear array is only about 50 cm long at a carrier frequency of 30 GHz. In light of the extremely large bandwidths that are available for commercial use (up to 7 GHz bandwidth in the 60 GHz band, and around 1 GHz at 28 and 38 GHz carrier frequency), massive MIMO systems in the mm-Wave range are ideally suited for high-capacity transmission and thus anticipated to form an important part of 5G systems. While the first commercial mm-Wave products are intended for in-home, short-range communications (e.g., for transmission of uncompressed video) [57], the potential of mm-Waves for *cellular outdoor* has recently been investigated [58, 59, 60]. Experiments have shown a coverage range of more than 200 m even in non line of sight (NLOS) situations [60]. Such long-range transmissions require high-gain adaptive antennas - something that massive MIMO implicitly provides.

For the downlink, massive MIMO systems at mm-Wave (or, for that matter, any other) frequencies require channel state information at the transmitter (CSIT), for conjugate beamforming as well as for other, more advanced, forms of MU-MIMO precoding (see [61] and references therein). In most existing papers, it has been assumed that this CSIT can be obtained from the uplink sounding signals, based on the principle of channel reciprocity [20]. However, reciprocity only holds (approximately) in Time Division Duplexing (TDD) systems, where the duplexing time is much shorter than the coherence time of the channel. In

Frequency Division Duplexing (FDD) systems, which are widely used in cellular communications, the spacing between uplink and downlink frequency is - for all practical systems - much larger than the coherence bandwidth of the channel [62]. Consequently, CSIT has to be provided through feedback - i.e., each user measures its channel vector in the downlink, and sends it to the BS in (quantized) form. Due to the large number of BS antenna elements, the overhead for this feedback can become overwhelming, and methods have to be devised for reducing this load.[2]

Joint Spatial Division and Multiplexing (JSDM) is a recent technique proposed in [63] to achieve massive-MIMO like gains for FDD systems (or, more generally, for systems that do not make explicit use of channel reciprocity), with the added advantage of a reduced requirement for CSIT[3]. The idea is to partition the user space into groups of users with *approximately similar* covariances,[4] and split the beamforming into two stages: a first stage consisting of a pre-beamformer that depends only on the second order statistics, i.e., the covariances of the user channels, and a second stage comprising a standard MU-MIMO precoder for spatial multiplexing on the effective channel obtained after pre-beamforming. The instantaneous CSIT of such an effective channel is easier to acquire thanks to the considerable dimensionality reduction produced by the pre-beamforming stage. Also, JSDM lends itself to a *hybrid beamforming* implementation, where pre-beamforming (which changes slowly in time) may be implemented in the analog RF domain, while the MU-MIMO precoding stage is implemented by standard baseband processing. This approach allows the use of a very large number of antennas with a limited number of baseband-to-RF chains; the latter depends on the number of independent data streams that we wish to send simultaneously to the users. A major challenge for massive MIMO in the mm-Wave region is the fact that the Doppler shift scales linearly with frequency, and thus the coherence time is an order of magnitude lower than that of comparable microwave systems. Thus, massive MIMO systems at mm-Wave frequencies need to be restricted to low-mobility scenarios. For comparable speeds of motion, for example, at pedestrian speeds (1 m/s), coherence times are of the order of a few ms at mm-Wave frequencies. Since (outdoor) coherence bandwidths of mm-Wave channels are similar to those of microwave channels [58, 65], the overall challenges of CSI feedback overhead are then comparable to those of higher-mobility (vehicular) microwave massive-MIMO systems. For example, a 30 GHz channel for a user moving at 1 m/s has the same coherence time and bandwidth of a 3 GHz channel for a user moving at 10 m/s. In this work, we explicitly assume the availability of perfect channel state information for simplicity (wherever required). In reality, devoting a certain amount of resource to the training phase would discount the achievable throughput by a certain factor [63].

The performance of JSDM depends on the type of channel statistics. Previous analysis was based on the one-cluster (local scattering) model, which means that the BS "sees" the incoming multi-path components (MPCs) under a very constrained angular range. This allows for an easy division of the users into sets, whose associated MPCs are disjoint in the angular domain, and can thus be separated by the pre-beamformers. However, this model does not represent many important scenarios. For example, in urban environments, high-rise buildings or street canyons can act as important "common clusters" that create spatially correlated MPCs for many users [66], [67], [68]. Another important effect, which becomes particularly relevant at mm-Wave frequencies, is *channel sparsity* - in other words, the number of significant MPCs is much lower than that for a microwave system operating in a similar environment. The low number of MPCs enables a further reduction of the CSIT that has to be fed back, and enables a new "degenerate" variant of JSDM, proposed in this paper and referred to as *Covariance-based JSDM*, that depends on the channel covariance information only. In fact, it is well known that, as long as the scattering geometry relative to a given user remains unchanged, the fading channel statistics are wide-sense stationary (WSS). In particular, this means

[2] TDD might also require feedback because accurate TDD calibration is difficult to achieve in practical hardware implementations. This is the reason why the only existing commercial standard that considers MU-MIMO downlink, IEEE 802.11ac, also prescribes explicit downlink training and quantized CSIT feedback, even though it uses TDD.

[3] An approach that exploits the same directional structure of the channel covariance matrix used by JSDM, in order to eliminate pilot contamination in a multi-cell massive MIMO setting, was proposed concurrently and independently in [64].

[4] Usually caused by the fact that the multi-path components of such users have similar angles at the BS

that the channel covariance matrix is time-invariant. In a typical scattering scenario, even if a user changes its position by several meters, the channel second order statistics remain unchanged [69, Chapter 4]. Hence, for a user moving at walking speed (1 m/s), the channel fading process is "locally" WSS over a time horizon of several seconds, spanning a very large number of symbol time slots (for example, a 20 MHz OFDM channel has symbol duration of $4\ \mu$s, corresponding to 10^6 symbols over an interval of 4s, corresponding to a user position displacement of 4m). We conclude that it is effectively possible to learn very accurately the channel covariance matrix at the transmitter side, even without requiring very fast CSIT feedback. This makes our scheme particularly interesting for mm-Waves. The main goal of our work is to apply the JSDM approach to *realistic* propagation channels inspired, inter alia, by the recent experimental observations of mm-Wave channels in an urban outdoor environment [60].

Our contributions in 2015-2016: We have made progress in the understanding of HDA massive MIMO implementation. From the signal processing viewpoint, we have proposed a number of novel schemes for estimating the dominant signal subspace (eigenspace of the channel covariance matrix) from noisy rank-deficient projections of the channel vector, measured through uplink pilots. This work is also the first contribution in the CoSIP DFG project that we have with Prof. Kutyniok of the Mathematics Department of TUB.

Then, we made significant advance in the design of a novel massive MIMO software-radio interface that we have called MMIB (massive MIMO interface board), which is able to support up to 16 RF chains with up to 40 MHz bandwidth operating between 1 and 5 GHz, in TDD mode, and route the signal samples in a perfectly synchronous way to the main FPGA card, which eventually is connected to the host server through a PCIe bus. This architecture will be able to support up to 128 antennas base station in a fairly standard Linux box, for a total cost of less than 15,000 Euros. This cost-effective testbed will host open source software provided by the Eurecom Institute (Open Air Interface), providing an LTE-TDD compatible base station able to handle commercial devices such as iPhones and iPads. Overall, our plan is to get to the pre-commercial stage, and perhaps spinoff the development as a startup, if the marked and finance conditions are favorable.

Going back to the theoretical study, we have also continued our collaboration with ETRI, a Government Research Institute in Korea, on the idea of JSDM, and in particular on the quantification of the so-called Transmit Correlation Diversity gain in massive MIMO with correlated base station antennas. Joint publications with Dr. Junyoung Nam of ETRI are in the making (submitted), and will be presented in next year report.

Main results: Our results on massive MIMO optimized in various relevant situations for TDD and FDD systems, including application to mm-Wave channels relevant to 5G systems, are included in the published papers:

1. Vishnu V. Ratnam; Andreas F. Molisch; Giuseppe Caire, "Capacity Analysis of Interlaced Clustering in a Distributed Transmission System With/Without CSIT," *IEEE Transactions on Wireless Communications* Year: 2016, Volume: 15, Issue: 4, Pages: 2629 - 2641.

2. M. Dai; B. Clerckx; D. Gesbert; G. Caire "A Rate Splitting Strategy for Massive MIMO with Imperfect CSIT," IEEE Transactions on Wireless Communications Year: 2016, Volume: PP, Issue: 99, Pages: 1 - 1.

3. K. Mahler; W. Keusgen; F. Tufvesson; T. Zemen; G. Caire "Tracking of Wideband Multipath Components in a Vehicular Communication Scenario," IEEE Transactions on Vehicular Technology, Year: 2016, Volume: PP, Issue: 99, Pages: 1 - 1.

4. Adhikary, A.; Dhillon, H.S.; Caire, G. "Massive-MIMO Meets HetNet: Interference Coordination Through Spatial Blanking," IEEE Journal on Selected Areas in Communications, Year: 2015, Volume: 33, Issue: 6 Pages: 1171 - 1186

5. Lim, Y.; Chae, C.; Caire, G. "Performance Analysis of Massive MIMO for Cell-Boundary Users," IEEE Transactions on Wireless Communications, Year: 2015, Volume: PP, Issue: 99 Pages: 1 - 1

6. Yongpeng Wu; Schober, R.; Ng, D.W.K.; Chengshan Xiao; Caire, G. "Secure Massive MIMO transmission in the presence of an active eavesdropper," Communications (ICC), 2015 IEEE International Conference on Year: 2015, Pages: 1434 - 1440

1.5 An information-theoretic approach to device-to-device communications

Power control and treating interference as Gaussian noise (TIN) is one of the most well-known, vastly employed, and yet most attractive interference management techniques, due to its low complexity, robustness to channel uncertainty, and to the fact that codes for the single-user Gaussian channel are well understood and efficiently implemented. Interestingly, it has also been shown that in some cases TIN is optimal or approximately optimal. For example, we know that TIN achieves the sum-capacity in the noisy regime of the two-user Gaussian interference channel [70, 71, 72]. In the general K-user single-antenna Gaussian interference channel, Geng *et. al* [73] have shown that, subject to certain conditions on the channel strengths, TIN achieves the optimal *Generalized Degrees of Freedom* (GDoF) region, and achieves the capacity region to within a constant gap, independent of the channel coefficients and the signal-to-noise ratio (SNR). The TIN optimality condition found in [73] is simply expressed in words as the fact that, for each user (i.e., intended transmitter-receiver pair) the desired signal strength level is no less than the sum of maximum strengths of all interfering signals from the transmitter to the other (unintended) receivers, and to the receiver from the other (unintended) transmitters, when all signal strengths are expressed in log-scale (e.g., in dB). For future reference, we indicate this condition as the "GNAJ" condition, from the initials of the authors of [73]. Under the GNAJ condition, the TIN-Achievable GDoF region (briefly referred to as "TINA region") is a convex polytope defined by the individual GDoF constraints and by the sum-GDoF inequalities corresponding to all possible ordered subsets of users. With the aid of a combinatorial tool named *potential graphs*, the K-user TINA region was characterized in [73] by $\sum_{m=2}^{K} \binom{K}{m}(m-1)! \approx (K-1)!$ constraints. More recently, it has been also shown by Sun and Jafar in [74] that, by a series of transformations of linear programs, the sum-GDoF characterization can be translated into a minimum weighted matching problem in combinatorial optimization. As such, the sum-GDoF under the GNAJ condition can be characterized as disjoint cycles partition of the interference network.

Such remarkable findings have inspired various related works, such as the TIN optimality of general X-channels [75], parallel interference networks [74], and compound interference networks [76]. In general, the TIN problem consists of two subproblems. Beyond the TINA region characterization, it is also important to find efficient methods to solve the TIN power control problem, that is, finding the (minimum) transmit powers that achieve a certain desired GDoF-tuple in the TINA region. The TIN power control problem has been open for a long time until a recent progress reported by Geng and Jafar in [76], where a simple yet elegant polynomial-time centralized iterative algorithm to find the globally optimal power allocation variables is provided. This centralized algorithm relies still on the representation by potential graphs.

One may wonder if the potential graph representation is the only path to both TINA region characterization and TIN power control problems. Further, due to the distributed nature of interference channels, decentralized power allocation algorithms are more interesting, desirable and yet challenging. In addition, it is worth noting that the GNAJ condition was only proven to be sufficient. An interesting counter-example in [73] showed that there exist partially-connected (in the sense of channel strength levels) interference channels, such that TIN achieves the optimal GDoF region and yet the GNAJ condition is not satisfied. A natural question then arises as to whether there exists a larger class of networks, including partially-connected ones, such that TIN is GDoF-optimal (i.e., TIN with power control still achieves the optimal GDoF region of the channel). These questions motivate this work.

Beyond the theoretical interest of characterizing the TINA region, we are also interested in translating these results into practical system optimization algorithms. Device-to-Device (D2D) communication is expected to play an important role in future wireless communication systems (e.g., 5G), including applications such as car-to-car, machine-to-machine, proximity-based services, and multi-hop infrastructureless mesh networks. The physical layer of D2D systems is usually modeled as a Gaussian interference channel. Under the practical constraint of treating interference as Gaussian noise for the sake of complexity and robustness,[5] a long-standing problem consists of controlling the power of the D2D links (transmit-receive pairs) in order to maximize the overall network throughput.[6] The usual approach of guaranteeing a target signal-to-interference-plus-noise ratio (SINR) to each link turns out to yield an operating point that can be arbitrarily far from optimal. This is because some bottleneck links may impose too stringent constraints to the overall network. In contrast, much better network throughput can be achieved by selecting a subset of active links in each slot and allocating positive power only to these selected links [79, 80, 81]. By scheduling the subsets of active links over time, it is possible to achieve individual throughputs such that some *network utility function* is maximized. In turn, the shape of the network utility function determines the desired fairness criterion (e.g., see [46, 82]). Link selection and scheduling has become the subject of intensive research. This problem is closely related to power control, since link selection corresponds to allocating either zero or positive power to the transmitters. For a general D2D network, this problem is non-convex and, as a matter of fact, has a combinatorial nature. For example, a well-known power control method consists of replacing $\log(1+\mathrm{SINR})$ with $\log(\mathrm{SINR})$ in the user rate expression, and using Geometric Programming (GP) [83]. However, by neglecting the "$1+$" inside the "$\log$" one has implicitly forced all links to use positive power, since assigning zero power to some links would drive the GP objective function to $-\infty$. Instead, it is known that generally much better solutions can be found by first selecting a "good" subset of active links, and then allocating (positive) power only to the selected links.

Various schemes for link selection have been proposed in the literature, e.g., [84, 85, 86, 79, 81] to name a few. For example, a large number of works is based on constructing an interference conflict graph [87], and then selecting maximal independent sets. These "maximal independent set scheduling" schemes are flawed by a fundamentally arbitrary choice of the threshold according to which two links are considered to be in conflict. Recently, a distributed link scheduling mechanism called FlashLinQ was proposed in [79]. Compared to "maximal independent set scheduling" schemes, FlashLinQ takes both signal and interference strength into account. In FlashLinQ, links are ranked in priority order and considered one by one. A candidate link is scheduled if it does not cause/receive too much interference to/from links of higher priority that have already been selected (i.e., declared active). It is also possible to enforce fairness among the links by changing the priority order at each scheduling slot, such that each link with a certain probability will be given the highest priority. More recently, inspired by the GNAJ condition in [73], the authors in [81] proposed a new distributed link scheduling mechanism (referred to as "ITLinQ") that provides sum throughput gains over FlashLinQ and yet maintains the same level of low-complexity. Instead of comparing the ratio of signal to interference strength of each new candidate link with a fixed threshold as in FlashLinQ, ITLinQ compares the interference level caused to/received from existing links with *an appropriately chosen exponent* of the signal strength of the new candidate link. It was verified by simulation in [81] that ITLinQ outperforms FlashLinQ with 28%-110% gains for a scenario where up to 4096 links can be scheduled.

Our contributions in 2015-2016: We have reformulate the TINA region problem in a reversed way,

[5] From [77] we know that this condition is essentially equivalent to imposing the use of minimum distance decoding at each receiver.

[6] Consistently with [78], we use the term "throughput" to indicate the time-averaged rate over a long sequence of scheduling time slots. In contrast, the instantaneous rate is the rate achieved in a single slot, for a given set of active users, i.e., links with positive transmit power.

from a combinatorial optimization perspective [88]. Interestingly, by first casting power allocation into an assignment problem, the globally optimal power allocation variables corresponding to any feasible GDoF tuple in the TINA region can be found by solving the equivalent assignment problem in polynomial time, either in a centralized manner (e.g., Hungarian method [89, 90]) or in a distributed one (e.g., Auction algorithm [91]). Inspired by the duality between the assignment and the maximum weighted matching problems in combinatorial optimization [92], we can express the TINA region characterization in terms of a maximum weighted matching problem. In doing so, the TINA region is significantly simplified, requiring only $2^K - 1$ constraints instead of $\approx (K-1)!$. Interestingly, such a representation also offers an interpretation of the disjoint cycle partition in [74]. By this new formulation, we show that the TINA region is a convex polytope under a novel channel strength condition that relaxes the GNAJ condition in [73]. This new condition requires that the desired signal strength of each user is no less than the *maximum difference* between the sum strength of any pair of incoming/outgoing interference signals and the strength of the link between such a pair (all in dB scale). Furthermore, together with a connectivity condition, we are able to establish the optimality of TINA region for a new class of networks. Such conditions are not included nor include the GNAJ condition [73].

We have also developed a new method for D2D link scheduling and distributed power control. As a matter of fact, for general channel strength coefficients, the maximal subset of links satisfying the GNAJ condition may not lead to the maximal (weighted) sum throughput or sum-GDoF. Our relaxed channel strength conditions provide a larger convex polytope TINA region. This provides a generally larger subset of links on which power control can be applied, resulting in generally higher weighted sum-GDoF. As a consequence, we are able to design a new distributed link scheduling and power control mechanism (named "ITLinQ+"), further fine-tuning the decision criterion of link selection. It is demonstrated by simulation that, when only link selection with constant power transmission for the selected links is used, ITLinQ+ gains 5%-20% average sum throughput improvement over ITLinQ with 1024 links, at the expense of slightly increased signaling overhead. A problem with both FlashLinQ and ITLinQ is that they do not consider power control and simply enforce constant transmit power for all the selected links. In contrast, thanks to the new understanding of the TINA region developed in this paper, we can augment ITLinQ+ with also a power control mechanism able to find the minimum power vector supporting the GDoF tuple corresponding to the selected links. When ITLinQ+ is applied together with this power control mechanism, the achieved average sum throughput is further enhanced. Most notably, the energy efficiency of ITLinQ+ can be substantially improved (e.g., 30 dB power saving to achieve 60 bit/s/Hz sum throughput for a 16-user D2D network), at the cost of additional computational complexity and some signaling overhead. In short, ITLinQ+ improves the sum throughput performance and yet requires much less energy consumption, which is desirable for battery-powered D2D communications. Notice that achieving better or equal throughput with less energy consumption is not a contradiction here, since the network is operated in an interference limited regime, such that rate is not immediately and obviously correlated to transmit power.

Main results:

1. Song-Nam Hong; Caire, G. "Beyond Scaling Laws: On the Rate Performance of Dense Device-to-Device Wireless Networks," IEEE Transactions on Information Theory, Year: 2015, Volume: 61, Issue: 9 Pages: 4735 - 4750

2. A. S. Tehrani; A. F. Molisch; G. Caire "Directional ZigZag: Neighbor Discovery with Directional Antennas," 2015 IEEE Global Communications Conference (GLOBECOM) Year: 2015, Pages: 1 - 6.

3. Yi, Xinping; Caire, Giuseppe "On the optimality of treating interference as noise: A combinatorial

optimization perspective," Information Theory (ISIT), 2015 IEEE International Symposium on Year: 2015, Pages: 1721 - 1725

4. Song-Nam Hong; Maric, I.; Hui, D.; Caire, G. "Multihop virtual full-duplex relay channels," Information Theory Workshop (ITW), 2015 IEEE Year: 2015, Pages: 1 - 5

1.6 Information theoretic security: authentication and identification

Consider an identification and authentication system with K users (see Fig. 1). In the enrollment phase, each user $w \in \{1, 2, \ldots, K\}$ generates a source sequence $X^n(w)$ and provides it to the system. Such source sequences are compressed into $\bar{M} \triangleq \{M(w) : w = 1, \ldots, K\}$ and stored into a database. The compressed user source data will be used as a reference for identification of the enrolled users. At the same time, the system produces a set of secret keys $\{S(w) : w = 1, \ldots, K\}$, also functions of the users' source sequences, which will be used as a reference for authentication of the identified user. In the identification/authentication phase, an a-priori unknown user W provides a measurement Y^n to the system. For example, this could be seen as a noisy version of its enrolled source sequence $X^n(W)$. Based on the stored database $\bar{M}$ and measurment Y^n, the user is identified as $\hat{W}$. The system also produces an estimated key $\hat{S}$. The user is successfully identified and authenticated if $(\hat{W}, \hat{S}) = (W, S(W))$.

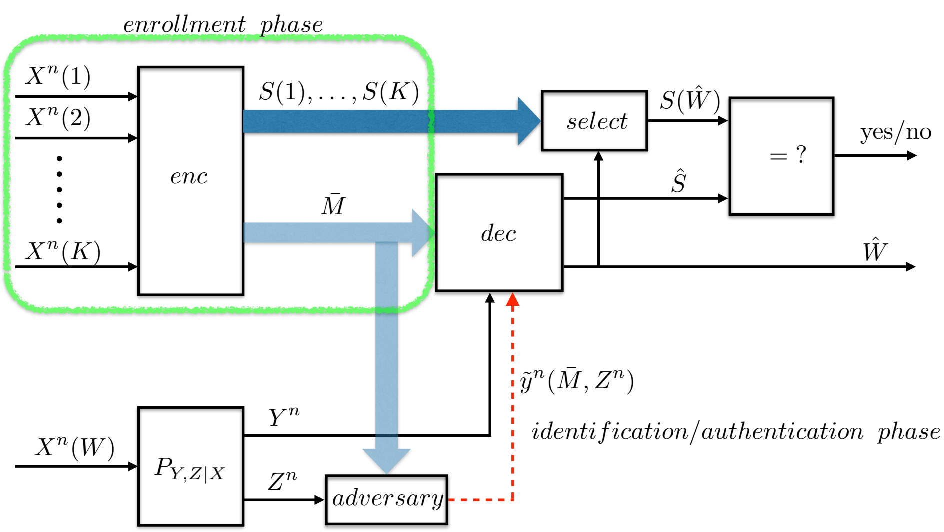

Figure 1: Identification and secret key-based authentication system with an adversary. The enrollment phase is contained in the green box. The remaining part corresponds to the identification/authentication phase. The red dashed arrow corresponds to an active adversary which replace the original user measurement Y^n with its own generated signal $\tilde{y}^n(\bar{M}, Z^n)$ in order to gain access to the system.

The system described above can be relevant in several applications including those involving access control, secure, and trustworthy communication. In database identification for access control applications, the system identifies an individual as an enrolled user and then grants the corresponding access based on authentication using the secret-key. In other words, the system first finds out which user in the database the individual corresponds to, and then verifies whether the individual is really the user he/she claims to be.

One important class of access control applications is related to using biometric data such as fingerprint, iris scans, voice, face, and DNA sequences. Unlike passwords, the biometric data inherently belong to the users and provide a convenient and seemingly more secure way for identification/authentication. However, it is crucial that privacy of the enrolled data must be protected from any inference of an adversary. The privacy risk in this case is of potentially high impact since the biometric data is typically tied to the person identity. If it is compromised, it cannot be reverted or changed easily, unlike in the case of a passwords.

In this work, we consider secret-key based identification and authentication problems in the presence of an adversary which is not part of the system but has full knowledge of the stored database data $\bar{M}$ as well as to some "on-line" side information Z^n, as shown in Fig. 1. We refer to Z^n as on-line side information since it is statistically dependent on the actual user W which is trying to be identified and authenticated. In contrast, the knowledge of $\bar{M}$ can be regarded as "off-line" side information for the adversary. Two closely related scenarios are studied:

1) The adversary is passive and is only interested in inferring the user source sequence. In this case, we wish to design a reliable identification/authentication system that achieves maximal identification rate and secret key rate while minimizing the the compression rate of the stored descriptions and the information leakage of the enrolled source sequences. In general, there exists tension between these performance metrics. For this scenario, our main contribution is a single-letter characterization of the optimal tradeoff region of the identification rate, compression rate, information leakage rate, and key rate for discrete memoryless sources.

2) The adversary is active and tries to deceive the identification/authentication system by using its own sequence $Y^n = \tilde{y}^n(\bar{M}, Z^n)$. We refer to the event where the legitimate user fails during the identification/authentication as a *false rejection*, and to the event where the system accepts the adversary as a *false acceptance*. In this case, we wish to design a secure identification/authentication system that achieves arbitrarily small false rejection probability with maximum identification rate and: i) minimizes the compression rate of each stored description, ii) minimizes the leakage rate of each enrolled sources, and iii) maximizes the error exponent of the maximum false acceptance probability (mFAP). For this scenario, our main contribution is a single-letter characterization of the optimal tradeoff between the identification rate, compression rate, information leakage rate, and mFAP exponent for discrete memoryless sources.

In order to motivate the role of key-based authentication to the possibly unfamiliar audience, we use the following naive everyday-life example. Consider the front door of a building with an intercom with multiple buttons. Each button corresponds to an apartment in the building. An intruder may wish to gain access to the building by hitting at random a button, hoping that the people inside the corresponding apartment just open the door, by identifying the intruder as friend/family just because he/she hit their button. Instead, if the intercom is also equipped with a camera and a facial recognition software, the door will be opened only if the intruder face (properly projected into some features space) generates a hashing function value that matches with the key corresponding to that apartment. Technically speaking, the optimal identification problem corresponds to K-ary hypothesis testing, which just provides the answer minimizing the average probability of wrong identification. However, the identified user needs also to be authenticated (in this case, by showing his/her face) in order to rightfully gain access to the system.

Our contributions in 2015-2016: The setting considered here includes correlated side information at the adversary, as treated here, is of practical interest since it models scenarios where the adversary can have access to noisy version of the source data. We have studied the secret key-based identification with a privacy constraint and provide a complete characterization of the identification-compression-leakage-key rate region $\mathcal{R}_1$ for discrete memoryless sources. It is shown that the layered binning scheme with rate allocation between compression and identification only on the first-layer description is optimal. The result includes many other results as special cases, one of which is the compression-leakage-key rate region for secret key-based authentication problem. Binary examples illustrating the derived tradeoffs have been developed. We have also studied a secure identification problem with a privacy constraint and provide a complete characterization of the identification-compression-leakage-mFAP exponent region $\mathcal{R}_2$ for discrete memoryless sources.

Our results show that the maximum key rate in $\mathcal{R}_1$ is equivalent to the maximum mFAP exponent in $\mathcal{R}_2$, revealing a connection between secret key rate and security of identification/authentication system.

Main results:

1. Kittichokechai, Kittipong; Caire, Giuseppe "Secret key-based authentication with a privacy constraint," Information Theory (ISIT), 2015 IEEE International Symposium on Year: 2015, Pages: 1791 - 1795.

References

[1] "White Paper: Cisco Visual Networking Index: Global Mobile DataTraffic Forecast Update, 2010 – 2015," http://www.scribd.com/doc/63529506/Cisco-White-Paper-c11-520862, Feb. 2010.

[2] F-L. Luo, *Mobile Multimedia Broadcasting Standards: Technology and Practice*, Springer Verlag, 2008.

[3] U. Reimers, "Digital video broadcasting," *Communications Magazine, IEEE*, vol. 36, no. 6, pp. 104–110, 1998.

[4] U. Ladebusch and C.A. Liss, "Terrestrial dvb (dvb-t): A broadcast technology for stationary portable and mobile use," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 183–193, 2006.

[5] O.Y. Bursalioglu, M. Fresia, G. Caire, and H.V. Poor, "Lossy multicasting over binary symmetric broadcast channels," *Signal Processing, IEEE Transactions on*, vol. 59, no. 8, pp. 3915–3929, 2011.

[6] Y. Li, E. Soljanin, and P. Spasojević, "Three schemes for wireless coded broadcast to heterogeneous users," *Physical Communication*, 2012.

[7] W.H.R. Equitz and T.M. Cover, "Successive refinement of information," *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 269–275, 1991.

[8] O.Y. Bursalioglu, M. Fresia, G. Caire, and H.V. Poor, "Lossy joint source-channel coding using raptor codes," *International Journal of Digital Multimedia Broadcasting*, vol. 2008, 2008.

[9] M.R. Chari, F. Ling, A. Mantravadi, R. Krishnamoorthi, R. Vijayan, G.K. Walker, and R. C, "Flo physical layer: An overview," *Broadcasting, IEEE Transactions on*, vol. 53, no. 1, pp. 145–160, 2007.

[10] V.K. Goyal, "Multiple description coding: Compression meets the network," *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 74–93, 2001.

[11] Y. Wang, A.R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, 2005.

[12] R. Ahlswede, "On multiple descriptions and team guessing," *Information Theory, IEEE Transactions on*, vol. 32, no. 4, pp. 543–549, 1986.

[13] Erik Nygren, Ramesh K Sitaraman, and Jennifer Sun, "The akamai network: a platform for high-performance internet applications," *ACM SIGOPS Operating Systems Review*, vol. 44, no. 3, pp. 2–19, 2010.

[14] N. Golrezaei, A.F. Molisch, and A.G. Dimakis, "Base station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Communications Magazine, in press.*, 2012.

[15] N. Golrezaei, K. Shanmugam, A. G Dimakis, A. F Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," *CoRR*, vol. abs/1109.4179, 2011.

[16] M.A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *arXiv preprint arXiv:1209.5807*, 2012.

[17] M.A. Maddah-Ali and U. Niesen, "Decentralized caching attains order-optimal memory-rate tradeoff," *arXiv preprint arXiv:1301.5848*, 2013.

[18] “Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018.,” .

[19] Hiroyuki Ishii, Yoshihisa Kishiyama, and Hideaki Takahashi, “A novel architecture for LTE-B: C-plane/U-plane split and phantom cell concept,” in *IEEE Globecom Workshops*, 2012, pp. 624–630.

[20] T.L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3590–3600, 2010.

[21] Jakob Hoydis, Stephan Ten Brink, and Mérouane Debbah, “Massive MIMO: How many antennas do we need?,” in *IEEE 49th Annual Allerton Conference on Communication, Control, and Computing*, 2011, pp. 545–550.

[22] H. Huh, G. Caire, H.C. Papadopoulos, and S.A. Ramprashad, “Achieving massive MIMO spectral efficiency with a not-so-large number of antennas,” *IEEE Trans. on Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, 2012.

[23] Jakob Hoydis, Mari Kobayashi, and Mérouane Debbah, “Green small-cell networks,” *IEEE Vehicular Technology Magazine,*, vol. 6, no. 1, pp. 37–43, 2011.

[24] V. Chandrasekhar, J. Andrews, and A. Gatherer, “Femtocell networks: a survey,” *IEEE Commun. Magazine,*, vol. 46, no. 9, pp. 59–67, 2008.

[25] “Quantenna 4×4 MIMO technology,” .

[26] “Broadcom 6×6 MIMO press release,” .

[27] T.S. Rappaport, Shu Sun, R. Mayzus, Hang Zhao, Y. Azar, K. Wang, G.N. Wong, J.K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!,” *IEEE Access*, vol. 1, pp. 335–349, 2013.

[28] Ansuman Adhikary and Giuseppe Caire, “Joint spatial division and multiplexing: Opportunistic beamforming and user grouping,” *arXiv preprint arXiv:1305.7252*, 2013.

[29] Amitabha Ghosh, Nitin Mangalvedhe, Rapeepat Ratasuk, Bishwarup Mondal, Mark Cudak, Eugene Visotsky, Timothy A Thomas, Jeffrey G Andrews, Ping Xia, Han Shin Jo, et al., “Heterogeneous cellular networks: From theory to practice,” *IEEE Communications Magazine*, vol. 50, no. 6, pp. 54–64, 2012.

[30] 3GPP technical specification group radio access network, “Further advancements for E-UTRA: LTE-Advanced feasibility studies in RAN WG4,” Tech. Rep., 3GPP TR 36.815, March 2010.

[31] Harpreet S Dhillon, Radha Krishna Ganti, François Baccelli, and Jeffrey G Andrews, “Modeling and analysis of k-tier downlink heterogeneous cellular networks,” *IEEE J. on Sel. Areas in Commun.*, vol. 30, no. 3, pp. 550–560, 2012.

[32] Mingyue Ji, Giuseppe Caire, and Andreas F Molisch, “Optimal throughput-outage trade-off in wireless one-hop caching networks,” *arXiv preprint arXiv:1302.2168*, 2013.

[33] Mingyue Ji, Giuseppe Caire, and Andreas F Molisch, “Wireless device-to-device caching networks: Basic principles and system performance,” *arXiv preprint arXiv:1305.5216*, 2013.

[34] Negin Golrezaei, Andreas F Molisch, Alexandros G Dimakis, and Giuseppe Caire, “Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution,” *Communications Magazine, IEEE*, vol. 51, no. 4, pp. 142–149, 2013.

[35] Karthikeyan Shanmugam, Negin Golrezaei, Alexandros G Dimakis, Andreas F Molisch, and Giuseppe Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," *arXiv preprint arXiv:1109.4179*, 2011.

[36] Negin Golrezaei, Karthikeyan Shanmugam, Alexandros G Dimakis, Andreas F Molisch, and Giuseppe Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 1107–1115.

[37] Negin Golrezaei, Karthikeyan Shanmugam, Alexandros G Dimakis, Andreas F Molisch, and Giuseppe Caire, "Wireless video content delivery through coded distributed caching," in *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2467–2472.

[38] Mohammad Ali Maddah-Ali and Urs Niesen, "Fundamental limits of caching," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 1077–1081.

[39] Urs Niesen and Mohammad Ali Maddah-Ali, "Coded caching with nonuniform demands," *arXiv preprint arXiv:1308.0178*, 2013.

[40] Ramtin Pedarsani, Mohammad Ali Maddah-Ali, and Urs Niesen, "Online coded caching," *arXiv preprint arXiv:1311.3646*, 2013.

[41] Mingyue Ji, Antonia M Tulino, Jaime Llorca, and Giuseppe Caire, "Order optimal coded caching-aided multicast under zipf demand distributions," *arXiv preprint arXiv:1402.4576*, 2014.

[42] Mingyue Ji, Giuseppe Caire, and Andreas F Molisch, "Fundamental limits of distributed caching in d2d wireless networks," in *Information Theory Workshop (ITW), 2013 IEEE*. IEEE, 2013, pp. 1–5.

[43] FP Kelly, "The mathematics of traffic in networks," *The Princeton Companion to Mathematics*, 2006.

[44] Y. Yi and M. Chiang, "Stochastic network utility maximisation-a tribute to Kelly's paper published in this journal a decade ago," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 421–442, 2008.

[45] M. Chiang, S.H. Low, A.R. Calderbank, and J.C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, 2007.

[46] Jeonghoon Mo and Jean Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking (ToN)*, vol. 8, no. 5, pp. 556–567, 2000.

[47] M.J. Neely, "Stochastic Network Optimization with Application to Communication and Queueing Systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[48] M.J. Neely, "Universal scheduling for networks with arbitrary traffic, channels, and mobility," in *Decision and Control (CDC), 2010 49th IEEE Conference on*. IEEE, 2010, pp. 1822–1829.

[49] Michael J Neely, "Wireless peer-to-peer scheduling in mobile networks," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2012, pp. 1–6.

[50] A. Ortega, "Variable bit-rate video coding," *Compressed Video over Networks*, pp. 343–382, 2000.

[51] F. Rusek, D. Persson, Buon Kiong Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 40–60, 2013.

[52] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Trans. on Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[53] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.

[54] T. L. Marzetta, G. Caire, M. Debbah, I. Chih-Lin, and S. K. Mohammed, ," .

[55] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, "Achieving "Massive MIMO" Spectral Efficiency with a Not-so-Large Number of Antennas," *IEEE Transactions on Wireless Communications*, , no. 9, pp. 3226–3239, September 2012.

[56] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[57] Eldad Perahia and Michelle X Gong, "Gigabit wireless lans: an overview of ieee 802.11 ac and 802.11 ad," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 15, no. 3, pp. 23–33, 2011.

[58] T.S. Rappaport, Shu Sun, R. Mayzus, Hang Zhao, Y. Azar, K. Wang, G.N. Wong, J.K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!," *Access, IEEE*, vol. 1, no. 1, pp. 335–349, May 2013.

[59] Mathew Samimi, Kevin Wang, Yaniv Azar, George N. Wong, Rimma Mayzus, Hang Zhao, Jocelyn K. Schulz, Shu Sun, Felix Gutierrez Jr., and Theodore S. Rappaport, "28 ghz angle of arrival and angle of departure analysis for outdoor cellular communications using steerable beam antennas in new york city," in *Vehicular Technology Conference Fall (VTC 2013-Fall), 2013 IEEE 74th*, 2013.

[60] Y. Azar, G.N. Wong, K. Wang, R. Mayzus, J.K. Schulz, Hang Zhao, F. Gutierrez, D. Hwang, and T.S. Rappaport, "28 ghz propagation measurements for outdoor cellular communications using steerable beam antennas in new york city," in *Communications (ICC), 2013 IEEE International Conference on*, 2013, pp. 5143–5147.

[61] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser mimo achievable rates with downlink training and channel state feedback," *Information Theory, IEEE Transactions on*, vol. 56, no. 6, pp. 2845–2866, 2010.

[62] Andreas F Molisch, *Wireless communications*, Wiley. com, 2010.

[63] A. Adhikary, Junyoung Nam, Jae-Young Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *Information Theory, IEEE Transactions on*, vol. 59, no. 10, pp. 6441–6463, 2013.

[64] Haifan Yin, David Gesbert, Miltiades Filippou, and Yingzhuang Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, vol. 31, no. 2, 2013.

[65] T.S. Rappaport, F. Gutierrez, E. Ben-Dor, J.N. Murdock, Y. Qiao, and J.I. Tamir, "Broadband Millimeter-Wave Propagation Measurements and Models Using Adaptive-Beam Antennas for Outdoor Urban Cellular Communications," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1850 – 1859, 2013.

[66] J. Fuhl, A.F. Molisch, and E. Bonek, "Unified channel model for mobile radio systems with smart antennas," *Radar, Sonar and Navigation, IEE Proceedings -*, vol. 145, no. 1, pp. 32–41, 1998.

[67] H. Asplund, A.A. Glazunov, A.F. Molisch, K.I. Pedersen, and M. Steinbauer, "The cost 259 directional channel model-part ii: Macrocells," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 12, pp. 3434–3450, 2006.

[68] M. Toeltsch, J. Laurila, K. Kalliola, A.F. Molisch, P. Vainikainen, and E. Bonek, "Statistical characterization of urban spatial radio channels," *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 3, pp. 539–549, 2002.

[69] T. S. Rappaport, *Wireless communications: principles and practice*, Prentice Hall, 1996.

[70] Xiaohu Shang, Gerhard Kramer, and Biao Chen, "A new outer bound and the noisy-interference sum–rate capacity for Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 689–699, 2009.

[71] Abolfazl S Motahari and Amir K Khandani, "Capacity bounds for the gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620–643, 2009.

[72] V Sreekanth Annapureddy and Venugopal V Veeravalli, "Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3032–3050, 2009.

[73] Chunhua Geng, Navid Naderializadeh, A Salman Avestimehr, and Syed A Jafar, "On the optimality of treating interference as noise," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1753–1767, Apr. 2015.

[74] Hua Sun and Syed A Jafar, "On the optimality of treating interference as noise for K user parallel Gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1911–1930, Apr. 2016.

[75] Chunhua Geng, Hua Sun, and Syed A Jafar, "On the optimality of treating interference as noise: General message sets," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 3722–3736, Jul. 2015.

[76] Chunhua Geng and Syed A. Jafar, "On the optimality of treating interference as noise: Compound interference networks," *arXiv:1412.2711*, Dec. 2014.

[77] Amos Lapidoth, "Nearest neighbor decoding for additive non-gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, 1996.

[78] D.N.C. Tse and S.V. Hanly, "Multiaccess fading channels. I. polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov 1998.

[79] Xinzhou Wu, Saurabha Tavildar, Sanjay Shakkottai, Tom Richardson, Junyi Li, Rajiv Laroia, and Aleksandar Jovicic, "FlashLinQ: A synchronous distributed scheduler for peer-to-peer ad hoc networks," *IEEE/ACM Trans. Networking*, vol. 21, no. 4, pp. 1215–1228, 2013.

[80] A Gjendemsj, David Gesbert, Geir E Oien, and Saad G Kiani, "Binary power control for sum rate maximization over multiple interfering links," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3164–3173, 2008.

[81] N. Naderializadeh and A.S. Avestimehr, "ITLinQ: A new approach for spectrum sharing in device-to-device communication systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1139–1151, June 2014.

[82] Leonidas Georgiadis, Michael J Neely, and Leandros Tassiulas, *Resource allocation and cross-layer control in wireless networks*, Now Publishers Inc, 2006.

[83] Mung Chiang, Chee Wei Tan, Daniel P Palomar, Daniel O'Neill, and David Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, 2007.

[84] Bruce Hajek and Galen Sasaki, "Link scheduling in polynomial time," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 910–917, 1988.

[85] Steven A Borbash and Anthony Ephremides, "Wireless link scheduling with power control and SINR constraints," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5106–5111, 2006.

[86] Weizhao Wang, Yu Wang, Xiang-Yang Li, Wen-Zhan Song, and Ophir Frieder, "Efficient interference-aware TDMA link scheduling for static wireless networks," in *Proc. ACM MobiCom*, Los Angeles, CA, USA, Sept. 2006, ACM, pp. 262–273.

[87] Kamal Jain, Jitendra Padhye, Venkata N Padmanabhan, and Lili Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless networks*, vol. 11, no. 4, pp. 471–487, 2005.

[88] Alexander Schrijver, *Combinatorial optimization: polyhedra and efficiency*, vol. 24, Springer, 2003.

[89] Harold W Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[90] James Munkres, "Algorithms for the assignment and transportation problems," *J. SIAM*, vol. 5, no. 1, pp. 32–38, 1957.

[91] Gabrielle Demange, David Gale, and Marilda Sotomayor, "Multi-item auctions," *The Journal of Political Economy*, pp. 863–872, 1986.

[92] Rainer Burkard, Mauro Dell'Amico, and Silvano Martello, *Assignment Problems, Revised Reprint*, vol. 106, SIAM, 2012.