Skoltech Research Statement
Alexey Frolov
August 2017

Research areas. I conduct the scientist research on the border of information theory, communications and machine learning with a special focus on application of the research results in communications, including engineering and industry problems. The main research areas are as follows:

1. Deep learning in communications. The main goal is to suggest an application of deep learning to communication problems, e.g. channel decoding. All classes of codes used in practical applications (e.g. Low-density parity-check (LDPC) codes or Polar codes) are designed such, that low complexity decoders (e.g. Belief Propagation) can be used to decode them. In 1983 Kim and Pearl introduced the belief propagation algorithm to solve statistical inference problems. Only later it was realized by MacKay, McEliece, and Cheng as well as Frey and Kschischang, that the decoding algorithm for LDPC and Turbo codes is in fact an instance of belief propagation. LDPC and Polar codes are not optimal and often suffer from error floor behavior. There are much better codes in the class of linear codes but we do not know, how to decode them with small complexity – this is exactly the place where we want to use deep learning techniques. There were already attemps to construct neural network (NN) learning-based decoders in literature, here we face with so-called curse of dimensionality problem: even for a short code of length N = 100 bits and rate R = 0.5, 2^{50} different codewords exist, which are far too many to fully train any NN in practice. In our opinion, the only way deal with the problem is to combine deep learning methods with existing decoding methods, i.e. we suggest using the decoding rules while we connect the nodes in the NN. Our initial experiments showed, that this approach works much better, than classical decoders and we can train the NN with only zero codeword (belongs to any linear code). At the same time, there is a gap in between this approach and (optimal) maximum likelihood decoding. In this project, we want to reduce this gap. This topic is new but it already attracts a lot of attention and is of theoretical (IEEE ITSoc, ComSoc) and practical interest. *We plan to collaborate with Huawei Technologies Co. Ltd. (Russian Research Center) and already participated in several meetings.*
2. Coding for distributed and cloud storage systems. Distributed and cloud storage systems have reached such a massive scale that recovery from several node failures is now part of regular operation of the system rather than a rare exception. To support reliable storage, error correcting codes are used. The simplest and to date the most frequently employed solution is to replicate the data several times. For example, Apache Hadoop, an open source software for distributed storage, uses a default method of 3-way replication. Another common solution, based on Reed-Solomon (RS) codes, provides stronger protection for the same or smaller storage overhead. For instance, the file systems of Facebook and Google use RS codes. One of the new tasks faced by such systems, but not addressed by current solutions, is recovery from a single node failure. Studies show that, although several concurrent failures are possible, and therefore the system should be able to protect against them, the most common scenario is the failure of a single node. Therefore, constructing codes that optimize the repair of a single node becomes an important problem for system developers. Another important property that codes for high-loaded distributed and cloud storage systems must require is availability. Availability plays an important role in supporting high throughput of the system. In our previous papers (see references [1, 11, 12] in CV) we improved the achievability bounds on the codes with locality and availability. We plan to construct explicit constructions of such codes with parameters close to the bounds. We plan to work in collaboration with A. Barg (Full Prof., University of Maryland, USA), Eitan Yaakobi (Assistant Prof., Technion, Israel) Antonia Wachter-

Zeh (Assistant Prof., Munich Technical University, Germany) and Itzhak Tamo (Assistant Prof., Tel-Aviv University, Israel).

3. Development of LDPC codes and their applications to future 5G wireless networks. 3GPP standardization was already started and we plan to participate in it (in collaboration with Huawei) and suggest new constructions of LDPC codes for URLLC (ultra-reliable low-latency communications). I have a huge background in LDPC codes, see references [2, 3, 7, 8, 13, 14, 20] from CV. We have already fulfilled a *Huawei cooperation project* and suggested code constructions for eMBB (extended mobile broadband) scenario. Here we want to focus on URLLC, i.e. short codes with very small ($\sim 10^{-5}$) probability of wrong decoding. Most of the latest achievements in the development of communications are based on information-theoretic principles that show how to transmit long data packets effectively (i.e. with use of long codes). Until recently, the transmission of short data packets has not been one of the major tasks in the development of wireless networks. At the same time, the traffic generated by machine type communications typically consists of short packets. We plan to develop new techniques for short code construction.
4. Development of non-orthogonal multiple access (NOMA) schemes for massive Internet of Things. Current wireless networks are designed to serve human users. The main goal of such networks is to increase the spectral efficiency (the transmission rate per 1 Hz) to transmit huge amounts of data, e.g. files or video streams. Next generation of wireless networks is facing a new challenge in the form of machine-type communication: billions of new devices (dozens per person) with dramatically different traffic patterns are expected to go live in the next decade. Indeed, according to Cisco, in 2015 the number of devices involved in network communication has exceeded the world population twofold and continues to grow exponentially. Existing wireless networks are inefficient in this case as to send a message the device should perform a complicated random access procedure, which leads to big delay. We want to address the key challenge – design of theoretical fundamentals and new technology for enabling massive random access. I have numerous publications on multiple access channels (see [4-6, 9, 10, 15-19]). We have submitted an NGP (Next Generation Project) application and will work on this topic *in collaboration with Yury Polyanskiy (Associate Prof., MIT)*. Developed technology is of great interest for such start-ups as Strizh and nWave (UK), which develop Internet of Things solutions. This research could be financially supported by Huawei Technologies Co. Ltd. within the frame of their interest and participation in 5G program.
5. Coding for fiber optic lines. *Approved cooperation project in between Huawei and Skoltech*. Currently two solutions are used:
   a. BCH or BCH-based turbo product codes with hard decision decoder. The length varies from 2000 to 65000 bits. These codes allow to achieve output bit error rate (BER) = 1e-15 for input BER in the range (1e-3; 1e-2).
   b. Convolutional generalized LDPC codes (GLDPC-CC) with soft decoding. These codes allow to achieve output BER = 1e-15 for input BER = 3e-2.

   GLPDC-CC have excellent error-correcting capabilities but the complexity is rather big. At the same time the complexity of encoding/decoding of BCH codes is very small, but we lose performance in this case. Project goal is to find a coding scheme in between GLDPC-CC and BCH codes, which is able to achieve output BER = 1e-15 for input BER in a range (1e-2; 2-2.5e-2);

Research goals. Our group is strongly motivated to provide a high-quality research in the areas mentioned above as well as the applications in the fields of expertise of the colleagues we are working with. Among them we should essentially mention A. Barg (Full Prof., University of Maryland, USA), Yury Polyanskiy (Associate Prof., MIT), Eitan Yaakobi (Assistant Prof., Technion, Israel), Itzhak Tamo (Assistant Prof., Tel-Aviv University, Israel), coding theory and communications groups (Huawei Technologies Co. Ltd). We would be happy to extend our cooperation.

Research awards.

- Russian Government Award in Science and Technology for Young Scientists 2016 «Research and development of protocols for prospective wireless networks».
- Moscow Government Award for Young Scientists 2013 «Development of multiple access techniques, which increase reliability of multimedia data delivery in real time».
- Russian President Scholarship 2013 «Coded modulation based on low-complexity decoded codes for fifth and sixth generation wireless systems».

Research grants.

- Russian Science Foundation
  - 14-50-00150 – Digital technologies and their applications (2015 – 2018). Lead Researcher.
- RFBR
  - 12-07-31035 – Signal-code construction development for fifth-generation telecommunication systems (2012-2013). Lead Researcher.
  - 13-01-12458 – Problems of constructability and complexity in information theory (2013-2015). Researcher.
  - 14-01-93108 – Discrete compressed sensing problem and error-correcting codes (2014-2015). Researcher.
  - 14-07-31197 – High-rate code constructions development for 5th- and 6th-generation telecommunication systems (2014-2015). Lead Researcher.