Machine Learning for RAN: Delusion or Salvation?

Slawomir Stanczak

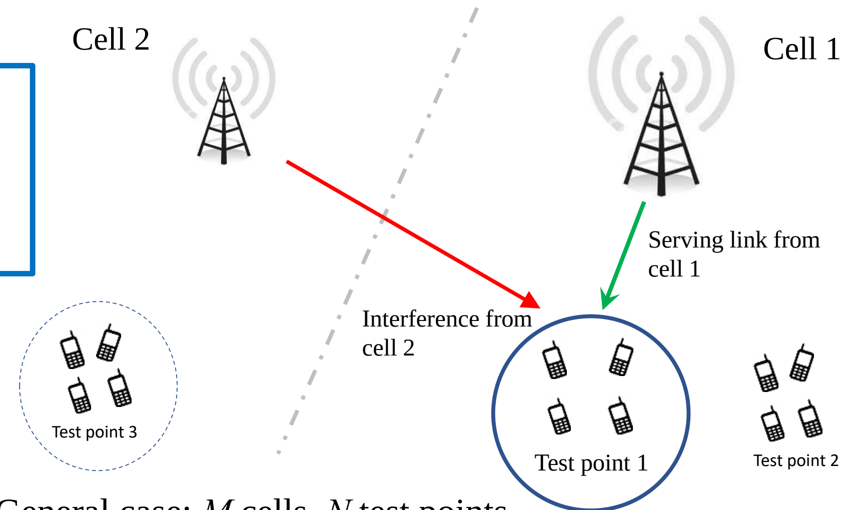
Why ML for Communications (=MLC)?

- **Entry points for ML-based improvements**
  1. high complexity (bad models)
  2. inefficient computation (limited resources)
  3. slow convergence (low latency applications)

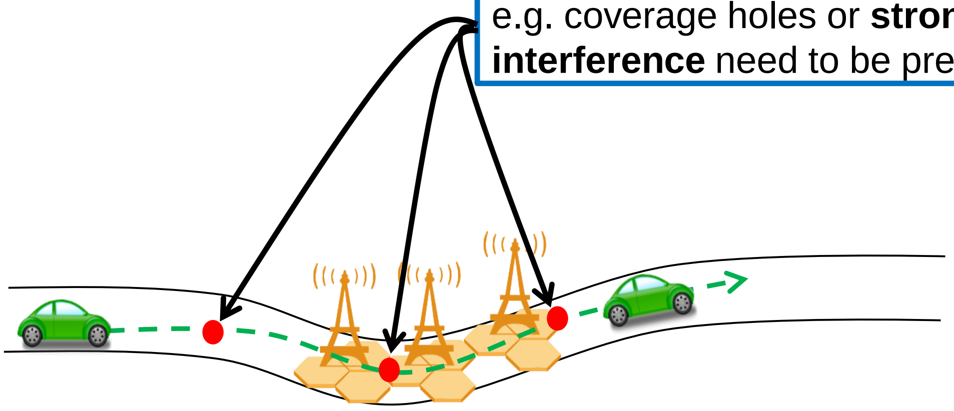
- **Potential benefits**
  1. enable to cope with increased complexity
  2. enhance efficiency
  3. facilitate cognitive network management
  4. **provide robust predictions**

Load Learning

Problem
What are users' rates as a function of the **load at each base station?**



General case: M cells, N test points



e.g. coverage holes or **strong inter-cell interference** need to be predicted

Reliable **rate-load mapping estimates/predictions** are key to reliable QoS predictions

- R. L. G. Cavalcante, Y. Shen, S. Stańczak, "Elementary Properties of Positive Concave Mappings with Applications to Network Planning and Optimization," IEEE Trans. Signal Processing, vol. 64, no. 7, pp. 1774-1783, April 2016

Performance Improvement due to Predicitons

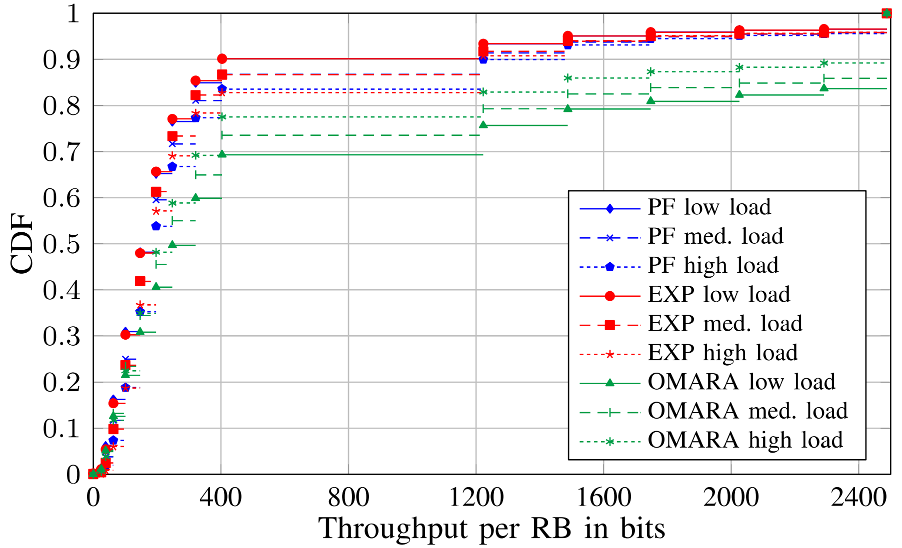


Fig. 3. CDF of the throughput per allocated RB

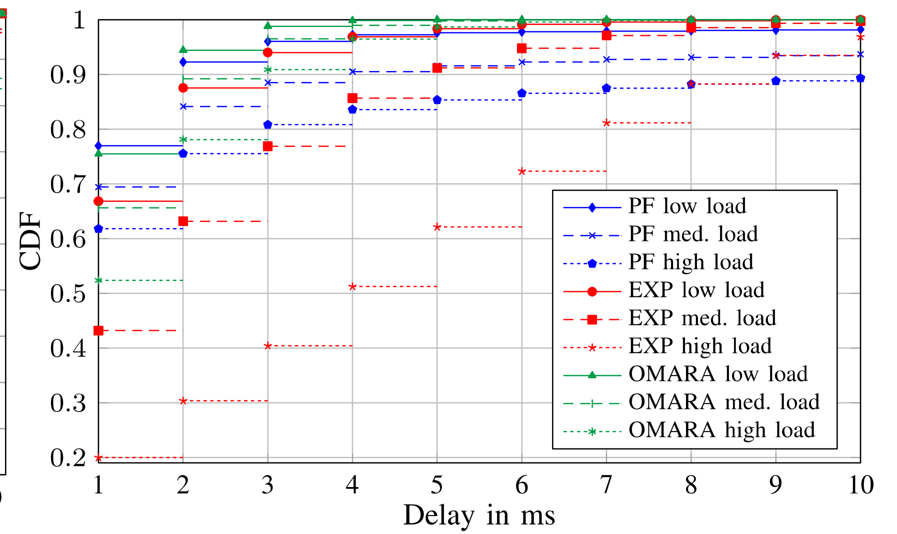


Fig. 4. DC packet delay CDF for successfully transmitted packets

- D. Külzer, S. Stańczak, M. Botsov, "Novel QoS Control Framework for Automotive Safety-Related and Infotainment Servcies," submitted, Nov. 2019

Classic vs. ML Approach (inspired by David Wipf)

Problem

Find a load vector x^* given users' rates θ and network configuration

Classic approach (model-driven)

- Modeling $f_\theta(x)$
- Simplification: $\hat{f}_\theta(x)$
- Human-designed algoirthm

  input: θ

  while <some condition is met>

  $x^{(n+1)} = T(x^{(n)})$; end

  output: $\hat{x} = x^*$

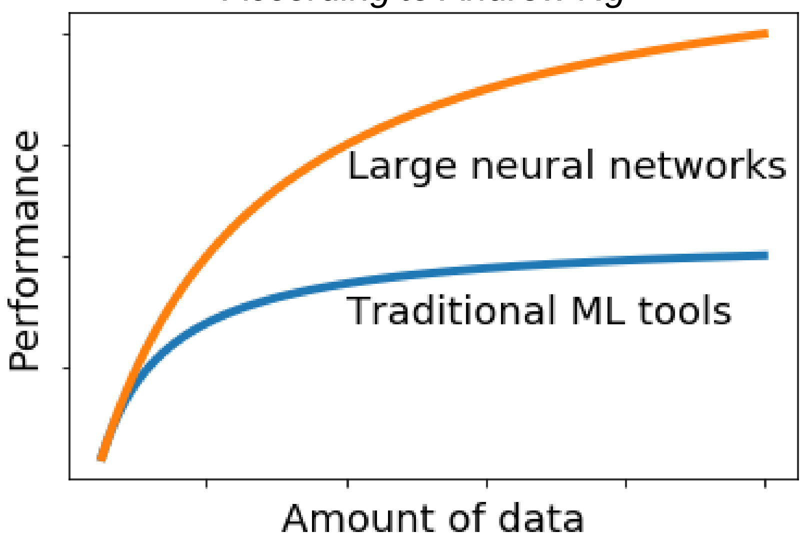
ML approach (data-driven)

- Choose a function set $\{g_\omega\}_{\omega \in \Omega}$
- Learn ω offline (e.g. from data)
- $\hat{x} = g_\omega(\theta) \approx x^*$

➔ **ML is much more than neural networks!**

Which Tools for MLC?

According to Andrew Ng



Key issues:

- Energy efficiency neglected
- Domain knowledge ignored
  ➔ Function properties not preserved
- Choice of performance metrics
- Amount of training data

Lower layers (PHY/MAC)	Higher layers
Collection of training data is limited • Fast time-varying channels and interference • **Short stationarity interval (V2X: 10-40ms)** • Distributed data • Limitations on computational power/energy	Huge datasets are available but • Incomplete data (missing measurements for long periods) • Erroneous data (e.g. software bugs) • Misaligned data (different times) • Time series (i.i.d. unrealistic)

Load Learning (cont.)

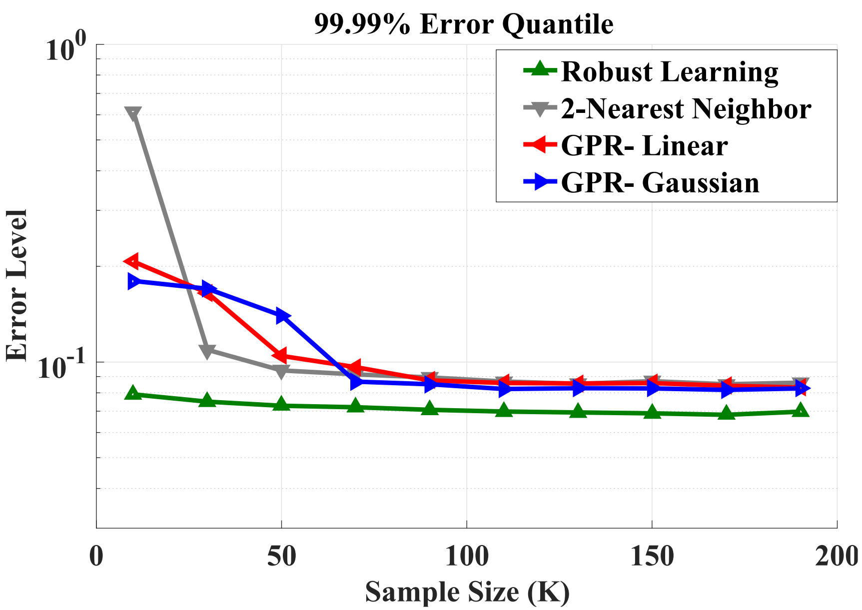
Challenge: The rate-load mapping (RLM) is highly dynamic and nonlinear owing to interference
➔ **training must be short**
➔ **important properties must be preserved**

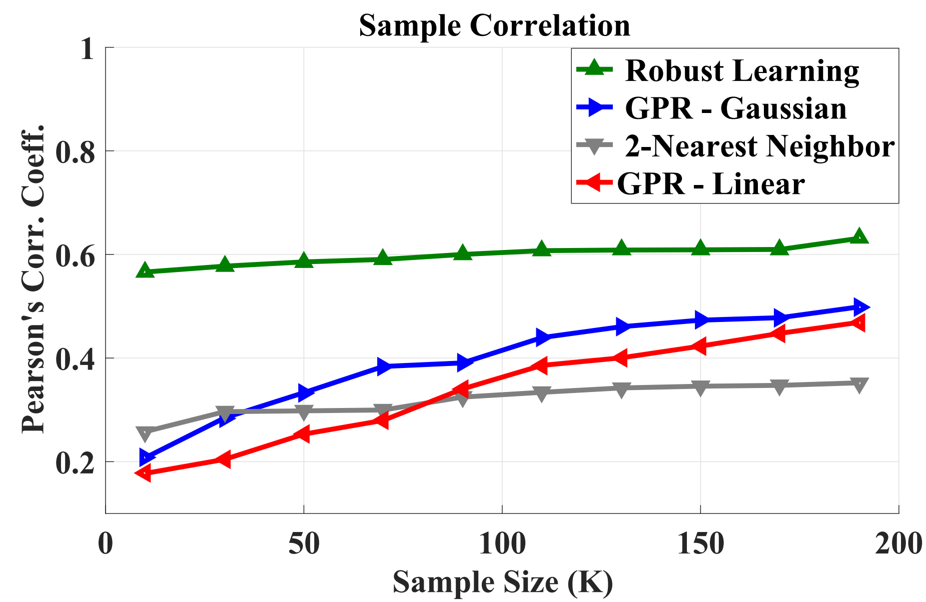
- Model-based approaches require too much a priori information
  But we should not ignore models

- The RLM has a rich structure (e.g., **monotonicity and Lipschitz**)
  They are hard to exploit in typical machine learning tools

Robust Online Load Learning

Hybrid-driven robust methods under uncertainty (e.g., few training samples)





- D. A. Awan, R. L. G. Cavalcante, and S. Stańczak, "A robust machine learning method for cell-load approximation in wireless networks," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018

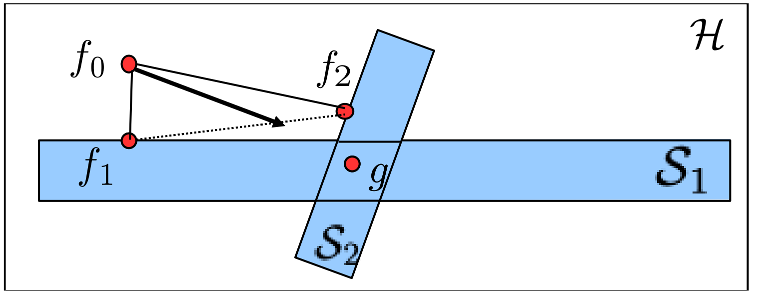
Demands on MLC

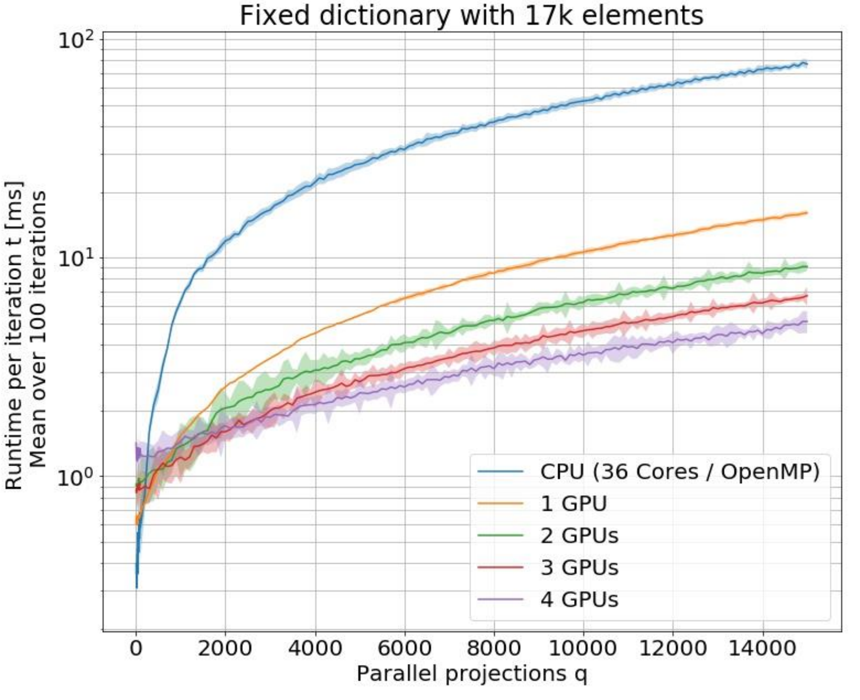
- ***Robust online*** ML with good tracking capabilities
  ➔ ML with small (uncertain) data sets
- Exploit ***domain knowledge*** (e.g. models, correlations, AoA)
  ➔ Hybrid-driven ML approaches (e.g. use production data)
  ➔ Learn features that change slowly over frequency, time...
  ➔ Preserve important function properties
- ***Distributed learning*** under communication constraints
  ➔ New functional architectures for Big Data analytics
- Low-complexity, ***low-latency implementation***
  ➔ New algorithms, massive parallelization

Learning in (Reproducing Kernel) Hilbert Spaces

Use projection methods in RKHS:

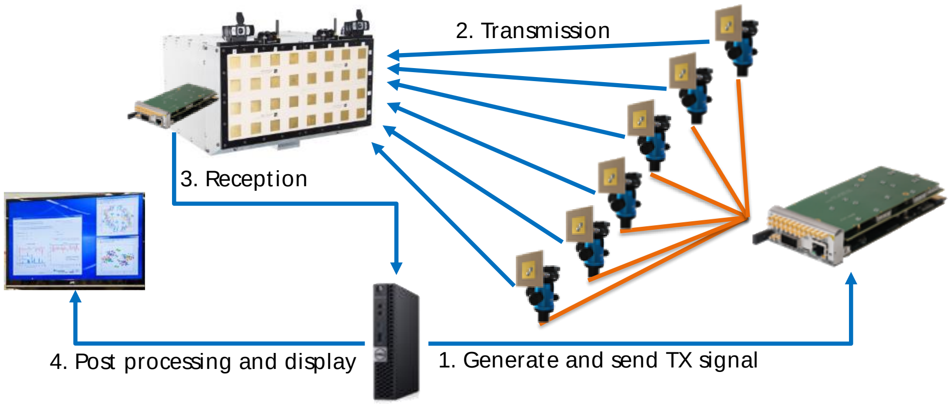
- ➔ Easy to exploit side information
- ➔ Initial fast speed
- ➔ Low complexity
- ➔ Convergence guarantees
- ➔ Massive parallelization via APSM for **fast learning on GPUs**

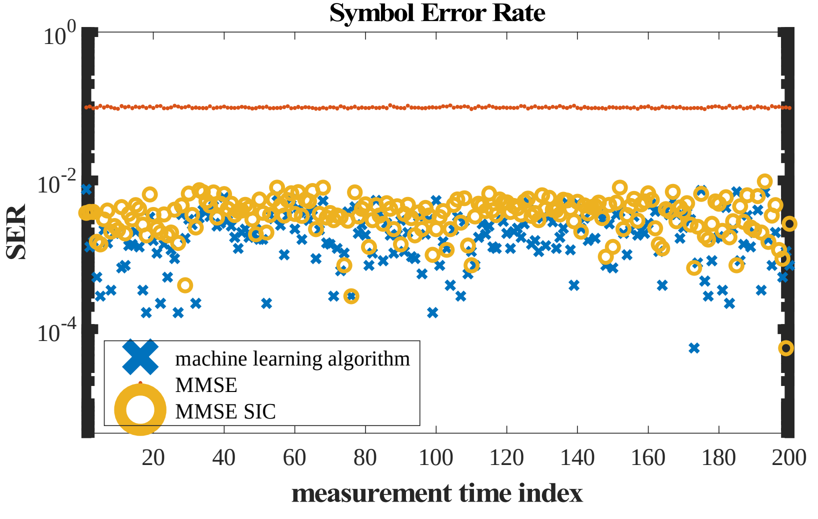


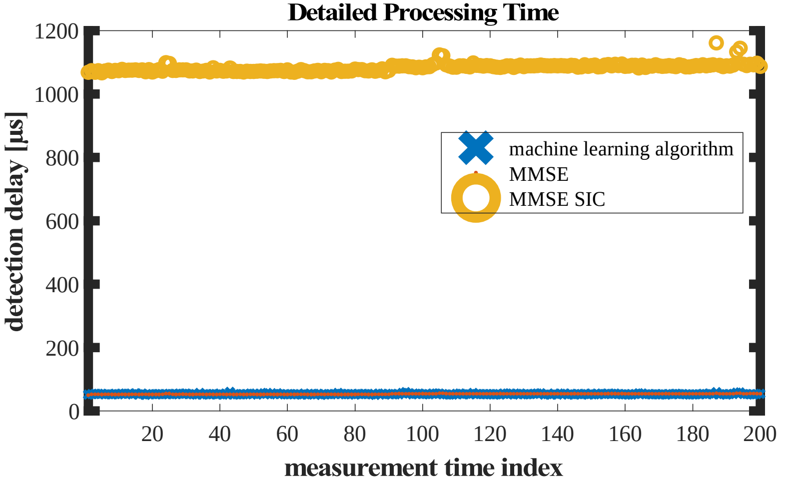


- I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," Numerical Functional Analysis and Optimization, vol. 25, no. 7/8, pp. 593–617, 2004.

Learning-based Reception for 5G NOMA







- D. A. Awan, R. L. G. Cavalcante, M. Yukawa, and S. Stańczak, "Detection for 5G-NOMA: An Online Adaptive Machine Learning Approach," in Proc. IEEE International Conference on Communications (ICC), May 2018
- D. A. Awan, R.L.G. Cavalcante, M. Yukawa, and S. Stanczak. Adaptive Learning for Symbol Detection: A Reproducing Kernel Hilbert Space Approach. Wiley, 2019. to appear.
- M. Mehlhose et.al., “Machine Learning-Based Adaptive Receive Filtering: Proof-of-Concept on an SDR Platform” submitted Oct. 2019

Can we design better neural networks?

Sparsity in Communication Systems

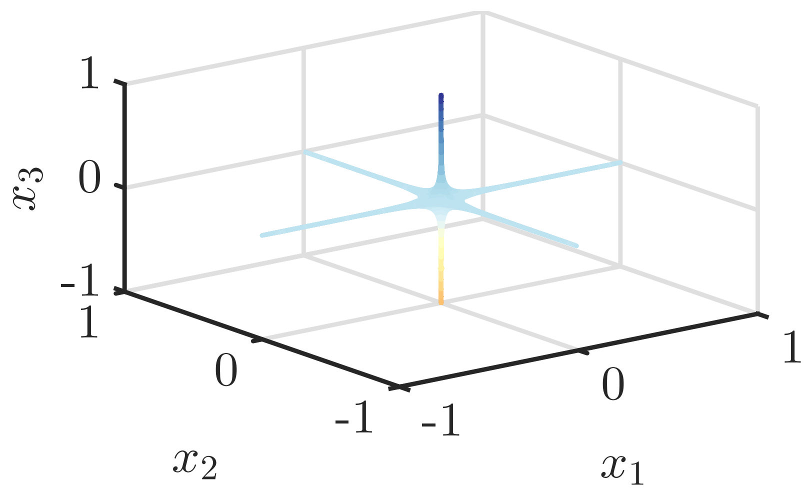
- Sparsity in the data (soft sparsity)
- Sparsity in the channel (soft sparsity)
- Sparsity in the user activity (hard sparsity)
- Sparsity in the network flow (hard sparsity)

We aren't likely to get a 1000X improvement in compute with the traditional, pure hardware improvements, or even better software and communication to put more chips together. It will need co-design of algorithms and compute e.g. can we create a model with a 1000X more parameters, but using only 10X more compute? I believe sparse models that address this issue and systems that can take advantage of these constraints will make a big difference.

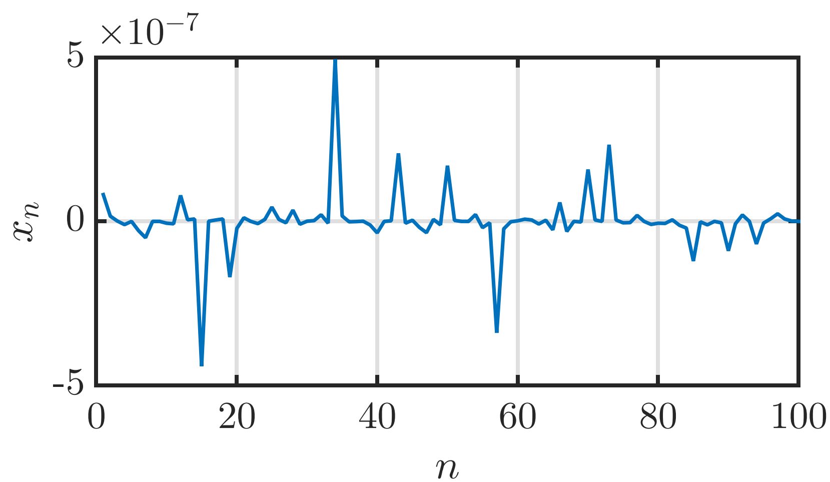
Rajat Monga, Google Brain, Lead Developer of TensorFlow

Sparsity in Communication Systems

- We can use $\mathcal{B}_p$-balls to model sparse signals $\mathcal{B}_p = \{\mathbf{x} : \sum_{i=1}^{N} |x_i|^{p_i} \leq 1\}$



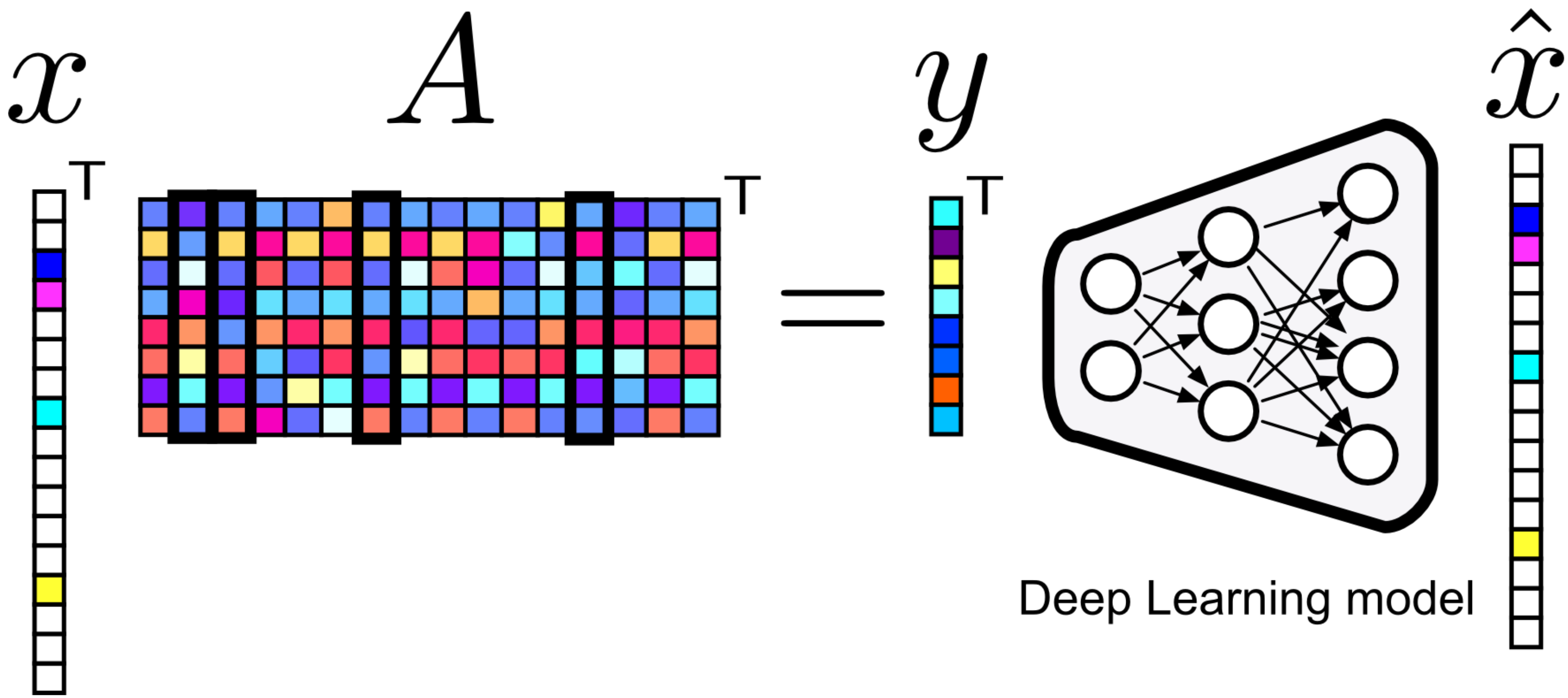
(a) $\mathcal{B}_{\boldsymbol{p}}$ for $\boldsymbol{p} = 0.25 \cdot \mathbf{1}$, $N = 3$



(b) realization of $p_{\mathbf{x}}$ for $\boldsymbol{p} = 0.25 \cdot \mathbf{1}$, $N = 100$

- S. Limmer and S. Stanczak, "Towards optimal nonlinearities for sparse recovery using higher-order statistics," 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), Vietri sul Mare, 2016, pp. 1-6.

Sparse Recovery via a Deep Neural Network

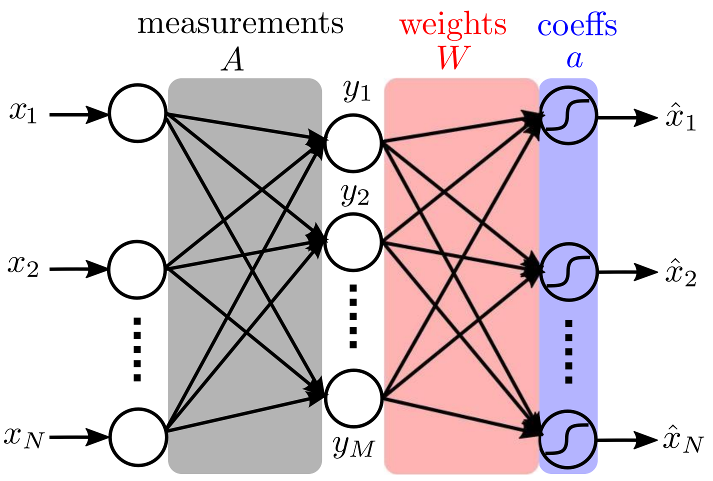


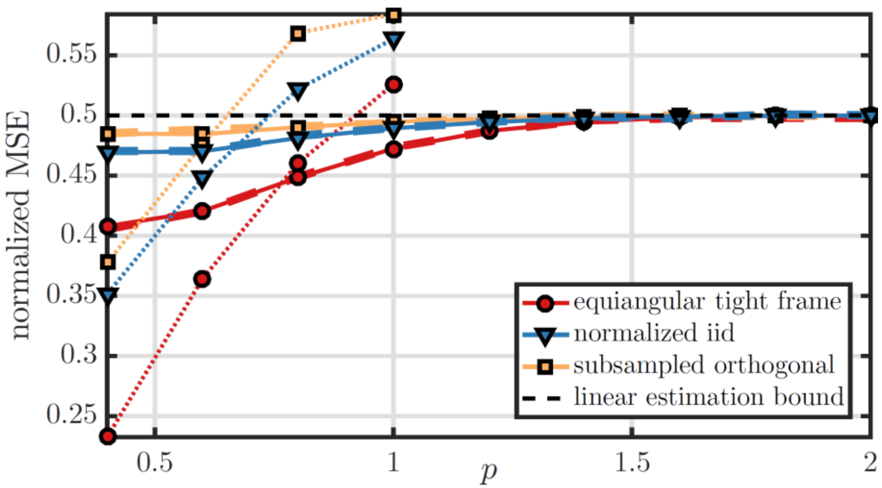
- CS methods are ***not*** suitable for low-latency applications
- Training must be short
  ➔ **Design** a good DNN for sparse recovery

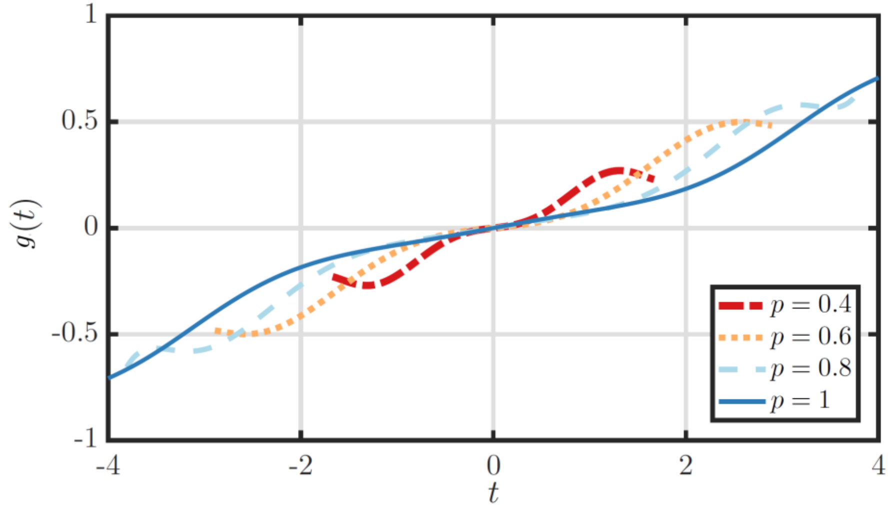
Optimization for MMSE Recovery

$\boldsymbol{A} \in \mathbb{R}^{3\times 6}$, activation σ with polynomial degree 9

- proposed (solid/dashed): linear estimator $+\epsilon \rightarrow$ online feasible
- LASSO (dotted): many iterations $\rightarrow$ online infeasible



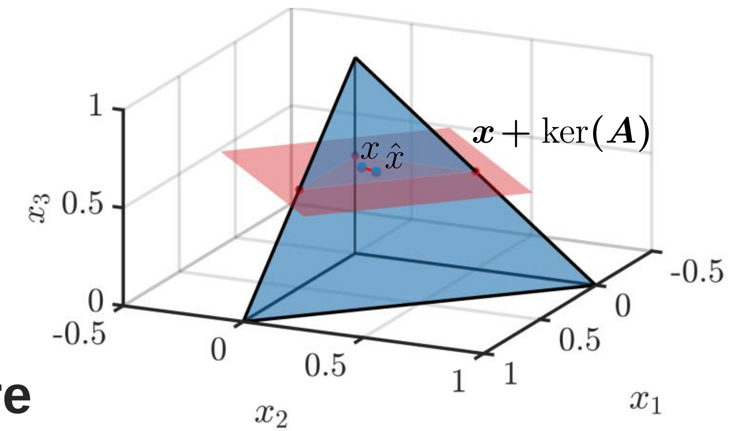


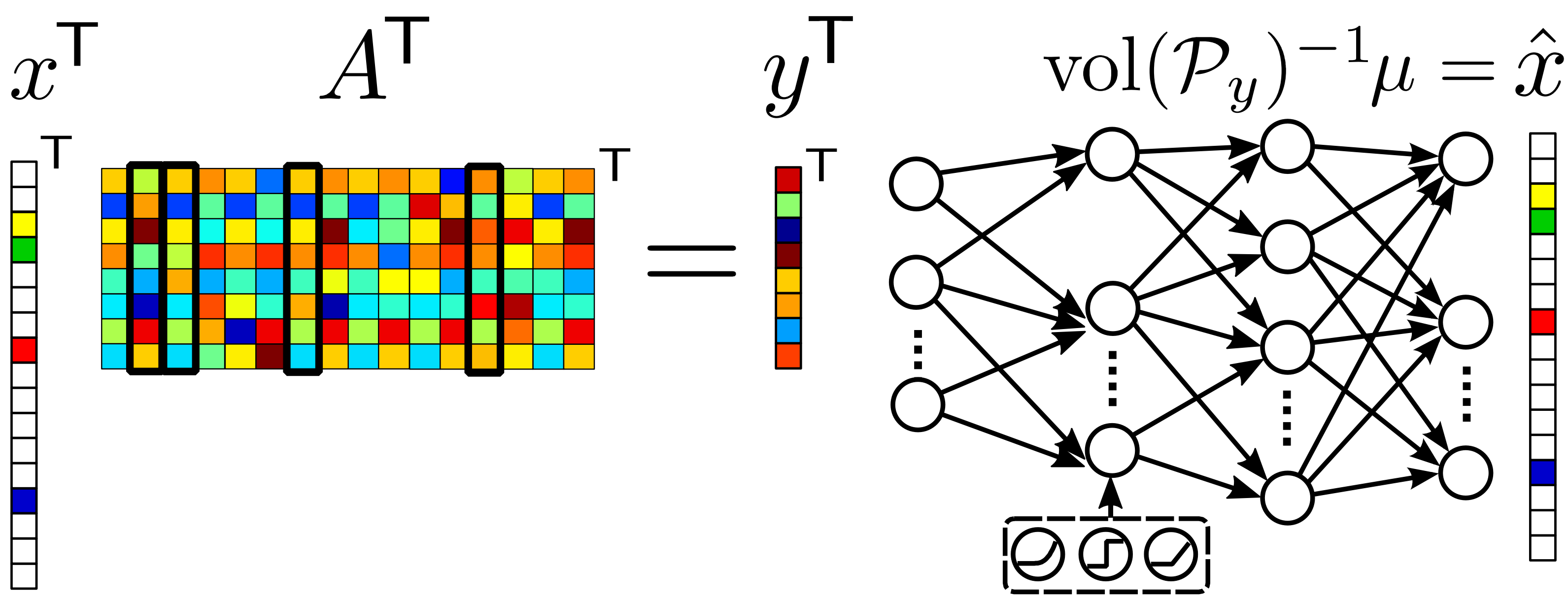


- S. Limmer and S. Stanczak, "Towards optimal nonlinearities for sparse recovery using higher-order statistics," 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), Vietri sul Mare, 2016, pp. 1-6.

Designing DNNs via Laplace Techniques

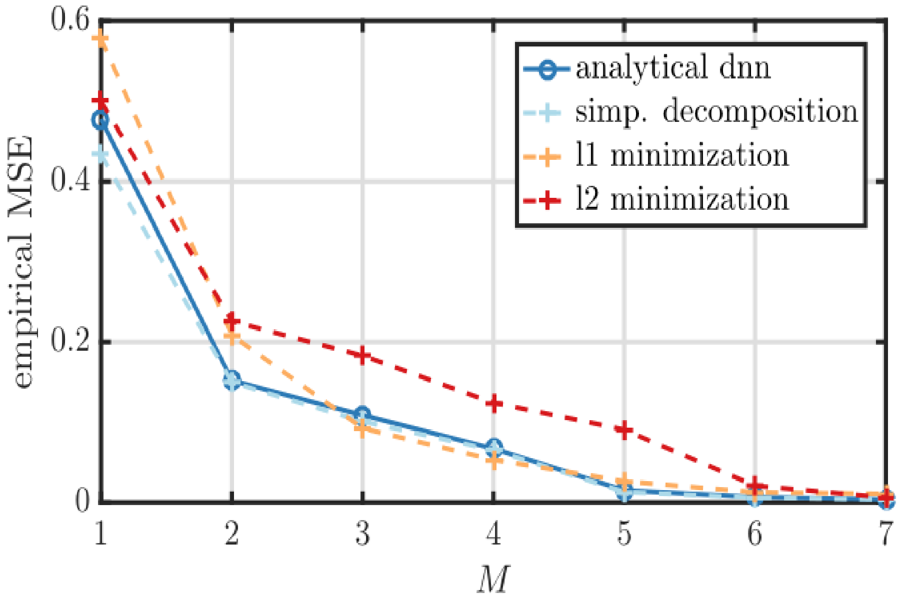
- Input uniformly distributed on
$$\mathcal{B}_1 = \{\mathbf{x} \geq 0 : \sum_{i=1}^{N} x_i \leq 1\}$$
- The conditional MMSE estimator is a polytope centroid under certain conditions.
  ➔ Volume and moment computation
- **Implementable using the DL architecture**

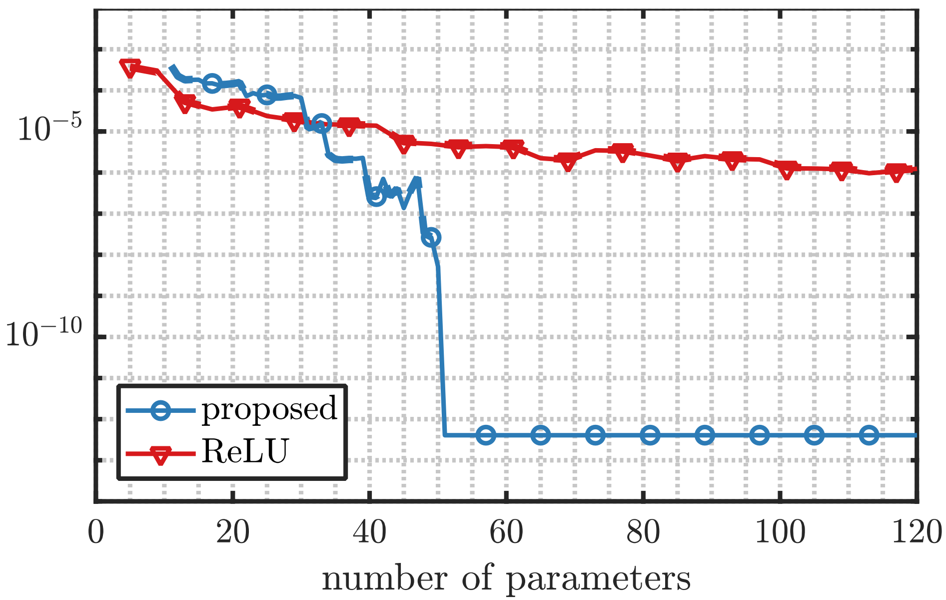




Numerical Experiments with Training

Real data





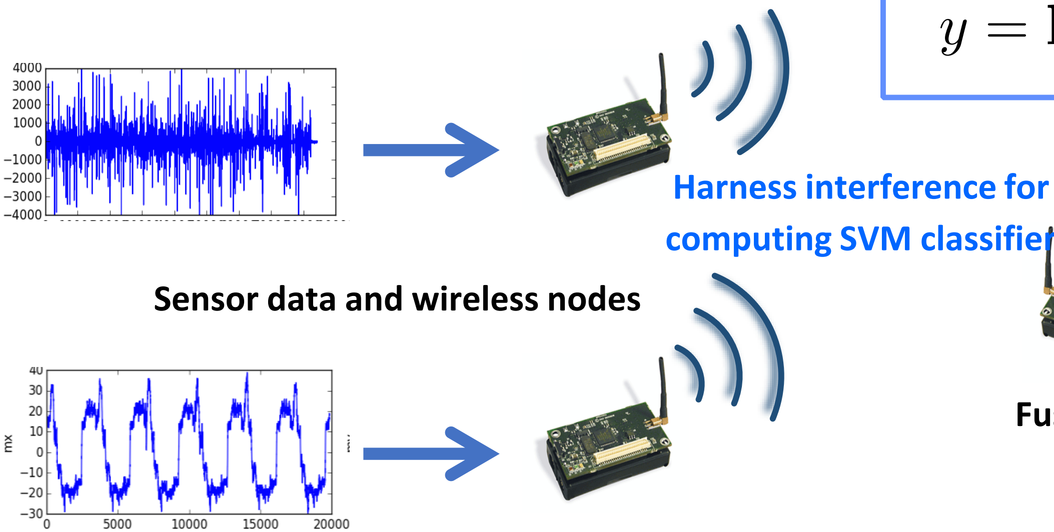
- S. Limmer and S. Stanczak, “A neural architecture for Bayesian compressive sensing via Laplace techniques“, IEEE Trans. On Signal Processing, Nov. 2018

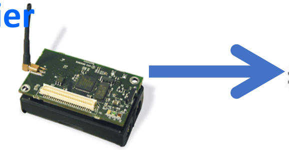
Take-away Message

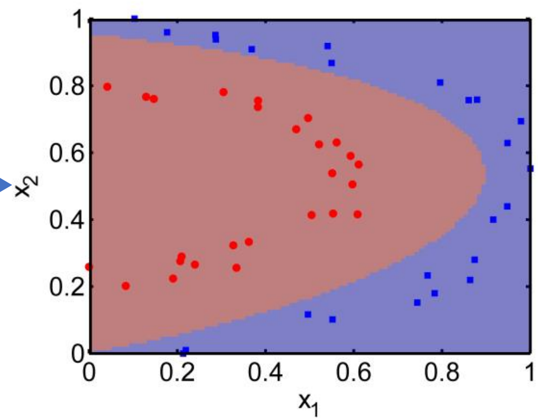
- ML/AI might be a “salvation” for industrial communication
- But there is a strong need for robust online ML methods
  → Exploit domain knowledge: Hybrid-driven distributed ML
  → Learn feature insensitive to frequency bands, phases …
- No time and data for extensive training of DNN
  → Design good NN architectures for a given task

Exploiting „Interference“ for Learning

$$y = \mathbf{h}^T \mathbf{x} + n \Rightarrow \psi\left(\sum_{k=1}^{K} \phi(x_k)\right)$$







- K. Ralinovski, M. Goldenbaum and S. Stanczak, Energy-efficient Classification for Anomaly Deteciton: The Wireless Channel as a Helper, IEEE ICC, 2016
- S. Limmer, J. Mohammadi, S. Stanczak, “A Simple Algorithm for Approximation by Nomographic Functions”, 53rd Annual Allerton Conference on Communication, Control, and Computing, 2015
- M. Raceala-Motoc and S. Limmer and I. Bjelakovic and S. Stanczak (2018). Distributed Machine Learning in the Context of Function Computation over Wireless Networks. 52nd Asilomar Conference on Signals, Systems and Computers 2018,
- Bjelakovic, M. Frey and S. Stanczak (2019). Distributed Approximation of Functions over Fast Fading Channels with Applications to Distributed Learning and the Max-Consensus Problem. 57th Annual Allerton Conference on Communication, Control, and Computing, 24-27 Sept. 2019 in Urbana, IL, USA

References

- M. Kasparick, R. L. G. Cavalcante, S. Valentin, S. Stańczak, and M. Yukawa, "Kernel-Based Adaptive Online Reconstruction of Coverage Maps with Side Information," IEEE Transactions on Vehicular Technology, vol. 65, no. 7, pp. 5461-5473, July 2016
- Z.Utkovski, P. Agostini, M.Frey, I.Bjelakovic, and S. Stanczak. Learning radio maps for physical-layer security in the radio access. In IEEE International Workshop on Signal Pro- cessing Advances in Wireless Communications (SPAWC), Cannes, France, July 2-5 2019. (invited).
- M.A. Gutierrez-Estevez, R.L.G. Cavalcante, and S. Stanczak. Nonparametric radio maps reconstruction via elastic net regularization with multi-kernels. In IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2018.
- R. L. G. Cavalcante, M. Kasparick, and S. Stańczak, "Max-min utility optimization in load coupled interference networks," IEEE Trans. Wireless Comm., vol. 16, no. 2, pp. 705-716, Feb. 2017
- D. Schäufele, et.al. "Tensor Completion for Radio Map Reconstruction using Low Rank and Smoothness", SPAWC, June 2019
- R. L. G. Cavalcante, Y. Shen, S. Stańczak, "Elementary Properties of Positive Concave Mappings with Applications to Network Planning and Optimization," IEEE Trans. Signal Processing, vol. 64, no. 7, pp. 1774-1783, April 2016
- R.L.G. Cavalcante, Q. Liao, and S. Stanczak. Connections between spectral properties of asymptotic mappings and solutions to wireless network problems. IEEE Trans. on Signal Processing, 2019. (accepted)
- D. A. Awan, R. L. G. Cavalcante, and S. Stańczak, "A robust machine learning method for cell-load approximation in wireless networks," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018
- D. A. Awan, R.L.G. Cavalcante, M. Yukawa, and S. Stanczak. Adaptive Learning for Symbol Detection: A Reproducing Kernel Hilbert Space Approach. Wiley, 2019. to appear.
- D. A. Awan, R. L. G. Cavalcante, M. Yukawa, and S. Stańczak, "Detection for 5G-NOMA: An Online Adaptive Machine Learning Approach," in Proc. IEEE International Conference on Communications (ICC), May 2018
- L. Miretti, R. L. G. Cavalcante, and S. Stańczak, "Downlink channel spatial covariance estimation in realistic FDD massive MIMO systems," in Proc. IEEE GlobalSIP 2018 (https://arxiv.org/abs/1804.04892)
- R. L. G. Cavalcante, L. Miretti, and S. Stańczak, "Error bounds for FDD massive MIMO channel covariance conversion with set-theoretic methods," in Proc. IEEE Global Telecommunications Conference (GLOBECOM), Dec. 2018 (https://arxiv.org/abs/1804.08461)
- J. Fink, D. Schaeufele, M. Kasparick, R. L.G. Cavalcante, and S. Stanczak. Cooperative localization by set-theoretic estimation. In Workshop on Smart Antennas (WSA), Vienna, Austria, April 24-26 2019.
- R. Ismayilov et.al. "Power and Beam Optimization for Uplink Millimeter-Wave Hotspot Communication Systems,"IEEE WCNC April 2019.
- R.L.G. Cavalcante, S. Stanczak, J. Zhang, and H. Zhuang. Low complexity iterative algorithms for power estimation in ultra-dense load coupled networks. IEEE Trans. on Signal Processing, 64(22):6058–6070, May 2016.
- S. Limmer and S. Stanczak, "Towards optimal nonlinearities for sparse recovery using higher-order statistics," 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), Vietri sul Mare, 2016, pp. 1-6.
- S. Limmer and S. Stanczak, "A neural architecture for Bayesian compressive sensing via Laplace techniques", IEEE Trans. On Signal Processing, Nov. 2018