From Evolution to (ML?) Revolution in Mobile Networking

Slawomir Stanczak

Fraunhofer
Heinrich Hertz Institute

The Actual Revolution May Yet Come

- 5G is mainly a revolution in business models
- Going beyond 5G may bring the actual revolution in mobile networking enabled by (RAN) **virtualization & AI/ML**

→ "true" E2E network slicing (including **vRAN**)
→ network functions executed on a GP hardware
→ in-memory computation
→ higher granularity & flexibility
→ **private, campus & regional networks**

→ new role of operators and vendors & new players, new business models & emergence of new services

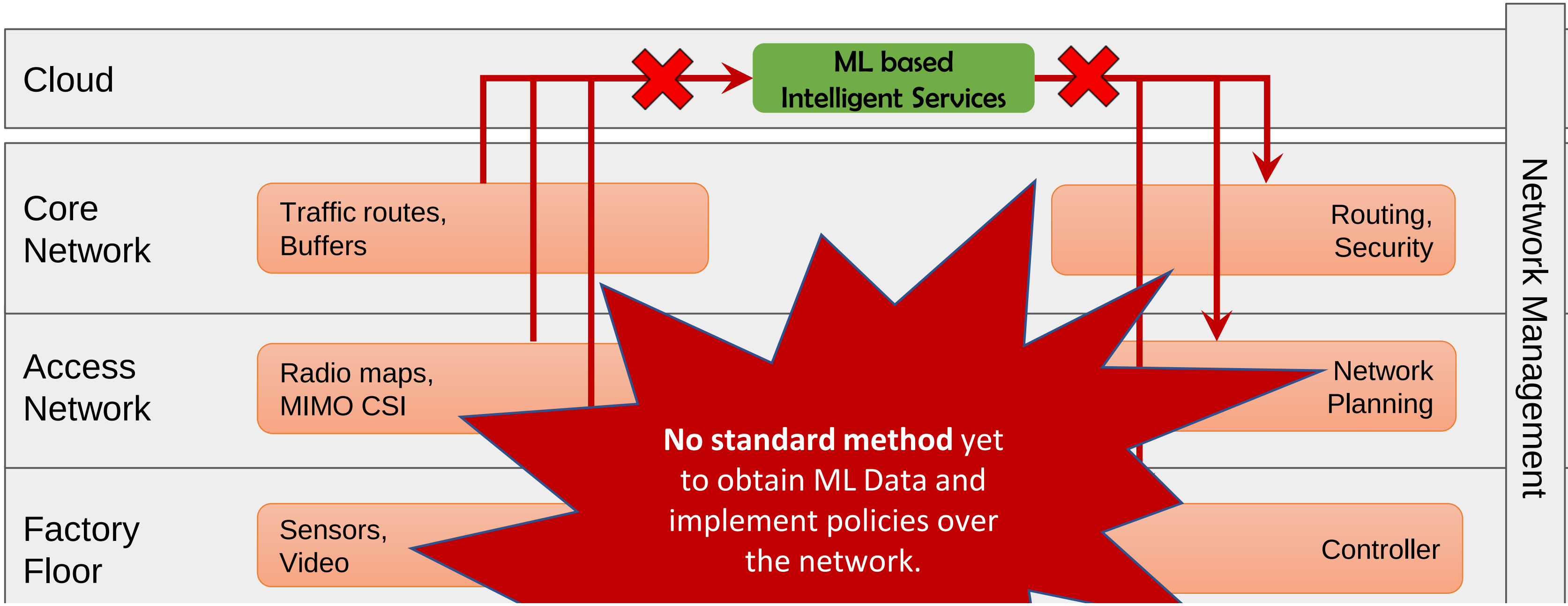
Campus Networks for Industry

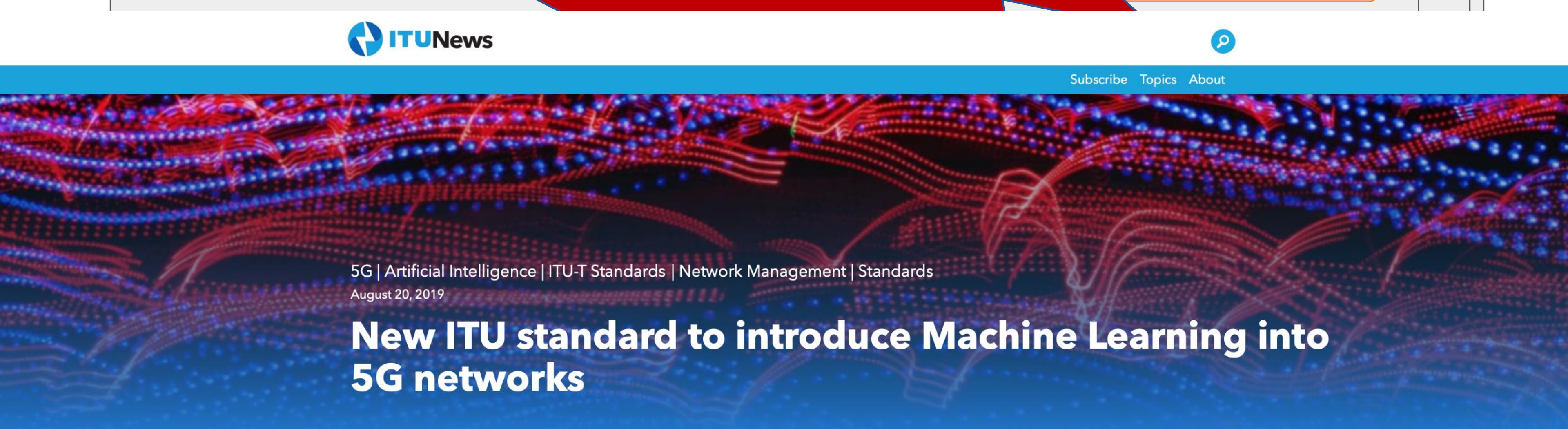
“The largest data records are not generated by companies in the Internet industry such as Google and Facebook, but by production technology systems“

McKinsey

ML based applications in Current Networks

Example: Factory Vertical



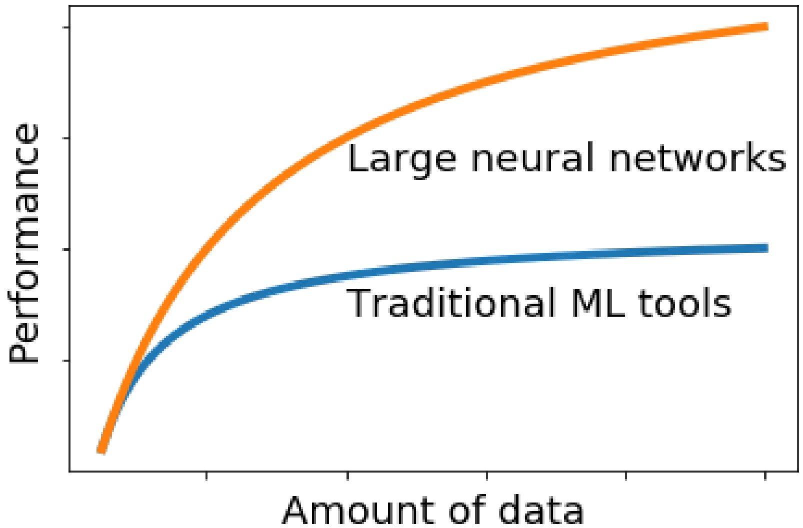


Why ML for Communications (=MLC)?

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

- **Potential benefits**
  - manageable complexity (e.g. via autoconfiguration)
  - higher efficiency (e.g. reduce # measurements)
  - fast decisions (e.g. parallelization & online learning)
  - robust predictions → anticipate rather than react

Tools for MLC



Key issues:

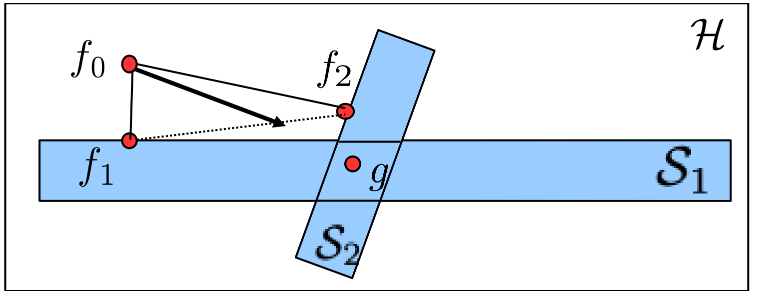
- Energy efficiency neglected
- Domain knowledge ignored
  ➔ Function properties not preserved
- Choice of performance metrics
- Amount of training data

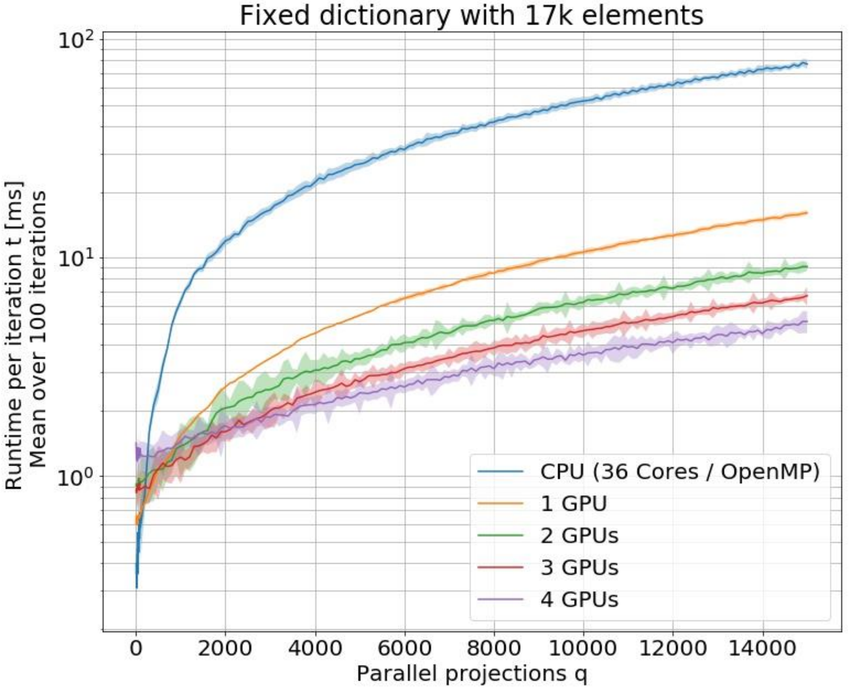
Lower layers (PHY/MAC)	**Higher layers**
Collection of training data is limited	Huge datasets are available but
• Fast time-varying channels and interference	• Incomplete data (missing measurements for long periods)
• **Short stationarity interval (V2X: 10-40ms)**	• Erroneous data (e.g. software bugs)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

Learning in (Reproducing Kernel) Hilbert Spaces

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**





ML/AI for Beyond 5G RAN

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

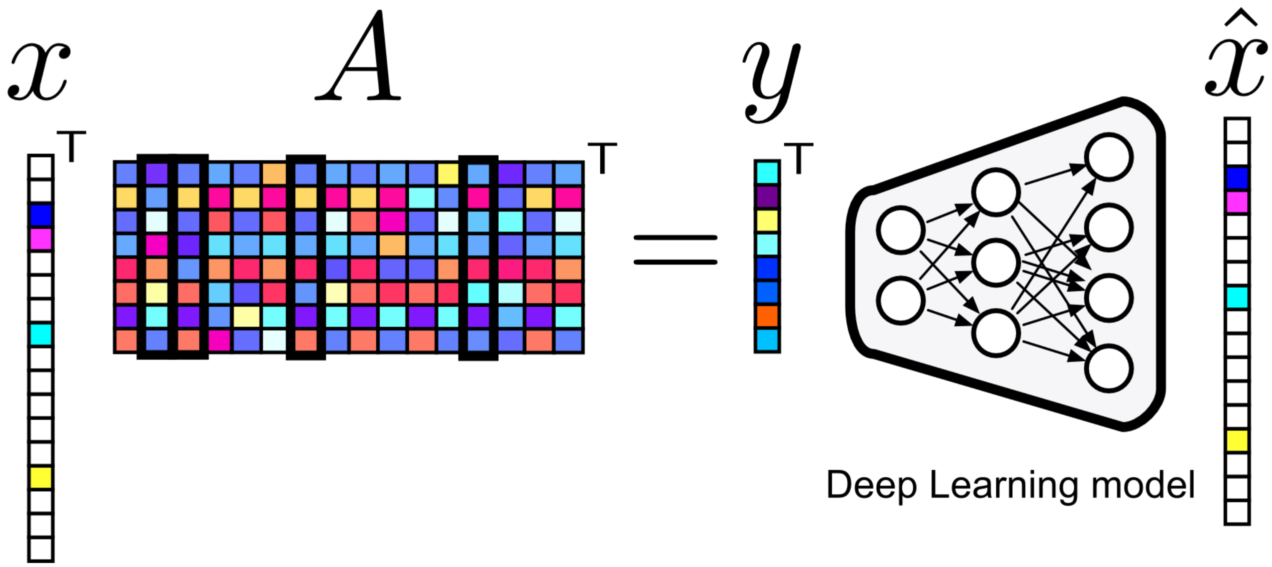
Sparsity in Communication Systems

- 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)

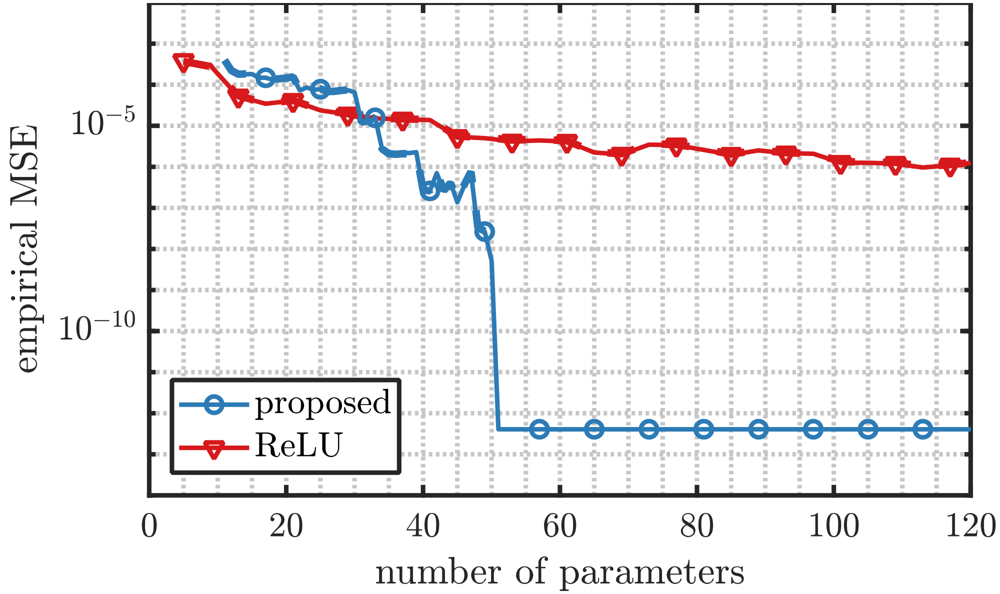
We aren't likely to get a 1000X improvement in compute with the traditional, pure hardware improvements, or even better software and communication to put more chips together. It will need co-design of algorithms and compute e.g. can we create a model with a 1000X more parameters, but using only 10X more 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

Sparse Recovery via a Deep Neural Network



- Training must be short
  ➔ **Design** a good DNN for sparse recovery and fast training



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