Machine Learning for 5G

Slawomir Stanczak

Joint work with R.L.G. Cavalcante, M. Kasparick, S. Limmer and L. Miretti

slawomir.stanczak@hhi.fraunhofer.de
Mobile networks - From the first to the fifth generation

1G
- Analog voice services (car telephony)
  Revolution: Analog → Digital

2G
- Digital voice services + SMS (globally available)
  Evolution: Integration of data services (mobile web)

3G
- Voice (circuit switched) + data (packet switched)
  Evolution: everything IP based + higher data rates

4G
- IP based (voice, email, web, audio/video streaming)
  Evolution & Revolution (Integration of vertical industry)

5G
- Control & Management (IoT, Tactile Internet, ML/AI)
Revolutionary Leap of Tactile Internet

Connecting machines into control loops at humanoid reaction times of milliseconds and less

Source: ITU TechWatch Report: The Tactile Internet
TACTILE INTERNET: Paradigm Shift from Transmititing Information for Humans to Networked Control

Source: www.domeoproducts.com

Source: http://blogs.voanews.com/

slawomir.stanczak@hhi.fraunhofer.de
TACTILE INTERNET: New KPIs:
Interaction Speed and Functional Safety & Security

source: kuka-robotics.com
www.youtube.com/watch?v=tIJME8-au8

source: www.prismagroup.it
Tactile Internet for Production and Logistics

5G for Industry 4.0

- Reliable remote machine operation
- Predictive maintenance
- High precision positioning
- Industrial edge cloud
- Truck-to-X Communication
- Augmented worker/workspace
- Machine and process monitoring
- Secure remote access

Source: Bosch
TACTILE INTERNET: New KPIs:
Interaction Speed and Functional Safety & Security

Source: courtesy of the US Department of Transportation
Collaborative Driving

Driver assistance with AR of potentially dangerous objects and situations

Source: ITU TechWatch Report: The Tactile Internet

slawomir.stanczak@hhi.fraunhofer.de
Tactile Networked Mobility

5G in Networked Driving

• Platooning
• Crossing traffic
• Collaborative driving
• Remote driving
• AR-based driver assistance
• Complete street perception
• Analysis of vehicle & driver conditions

Source: ITU TechWatch Report: The Tactile Internet
Source: The US Department of Transportation
4G and 5G compared

Strong interdependencies between different requirements

NB: Downlink metrics shown
# Industrial Communication Requirements

<table>
<thead>
<tr>
<th></th>
<th>Diagnoses &amp; Maintenance</th>
<th>Discrete Manufacturing</th>
<th>Warehouses &amp; Logistic</th>
<th>Process Automation</th>
<th>Augmented Reality</th>
<th>Safety</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Latency</strong></td>
<td>Generally</td>
<td>Condition Monitoring</td>
<td>Generally</td>
<td>AGV</td>
<td>Cranes &amp; Hoists</td>
<td></td>
</tr>
<tr>
<td>(Sensor → Controller →</td>
<td>&gt; 20ms</td>
<td>100 ms</td>
<td>1ms – 12ms</td>
<td>&gt;50ms</td>
<td>50ms – Xs</td>
<td>10ms</td>
</tr>
<tr>
<td>Actuator)</td>
<td></td>
<td></td>
<td>250μs – 1ms</td>
<td>15ms – 20ms</td>
<td>10ms</td>
<td>10ms</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>1 – 10⁻⁴</td>
<td>1 – 10⁻⁵</td>
<td>1 – 10⁻⁴</td>
<td>&gt; 1 – 10⁻⁴</td>
<td>1 – 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>(wrt. to „successful</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transmission within</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>latency bound)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data rate</strong></td>
<td>kbit/s – Mbit/s</td>
<td>kbit/s – Mbit/s</td>
<td>kbit/s – Mbit/s</td>
<td>kbit/s – Mbit/s</td>
<td>kbit/s – Mbit/s</td>
<td>kbit/s</td>
</tr>
<tr>
<td><strong>Packet size</strong></td>
<td>&gt; 200 Byte</td>
<td>20 – 50 Byte</td>
<td>&lt; 300 Byte</td>
<td>&lt; 80 Byte</td>
<td>&gt; 200 Byte</td>
<td>&lt; 20 Byte</td>
</tr>
<tr>
<td>**Distance (between</td>
<td>&lt;100m</td>
<td>100m – 1km</td>
<td>&lt; 200m</td>
<td>100m – 1km</td>
<td>&lt;100m</td>
<td>&lt;30m</td>
</tr>
<tr>
<td>communication devices)</td>
<td></td>
<td></td>
<td>~ 2m</td>
<td>~30m</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 100m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Motion speed</strong></td>
<td>0 m/s</td>
<td>&lt;10 m/s</td>
<td>&lt;40 m/s</td>
<td>Generally</td>
<td>&lt;3 m/s</td>
<td>&lt;10 m/s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;10 m/s</td>
<td>&lt;5 m/s</td>
<td>none, else</td>
<td></td>
<td></td>
</tr>
<tr>
<td>**Latency critical</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>mobility support**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Device density</strong></td>
<td>0.33 – 3 m⁻²</td>
<td>10 – 20m⁻²</td>
<td>0.33 – 3m⁻²</td>
<td>&gt; 0.1 m⁻²</td>
<td>10000 / Factory</td>
<td>&gt; 0.03 –</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;5 m⁻²</td>
<td>~0.1 m⁻²</td>
<td>&gt; 0.02m⁻²</td>
<td>&gt; 0.03 –</td>
</tr>
<tr>
<td><strong>Energy efficiency</strong></td>
<td>n/a</td>
<td>10 years</td>
<td>n/a</td>
<td>&lt;8h</td>
<td>10 years</td>
<td>1 day</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Localization accuracy</strong></td>
<td>&lt;50cm</td>
<td>&lt;50cm</td>
<td>&lt;5cm</td>
<td>&lt;10cm</td>
<td>&lt;50cm</td>
<td>&lt;50cm</td>
</tr>
</tbody>
</table>

---

[2] VDE Positionspapier "Funktechnologien für Industrie 4.0"

Focus of [KOI] Fraunhofer Heinrich Hertz Institute

slawomir.stanczak@hhi.fraunhofer.de
5G Enablers

- Increase of network density ➔ New small-cell architectures
- Increase of spectral efficiency ➔ Massive MIMO (massive number of antennas)
- Increase of the available bandwidth ➔ Millimeter-Wave Technology
Network Densification

Multi-Layer Mobile Networks with Small Cell Deployments

- Multiple nested layers
- Top: Macro cells
  - Coverage, connectivity and mobility
  - Control, management and interoperability
- Downwards: smaller cells
  - Higher energy efficiency
  - Higher frequencies
- Bottom:
  - D2D communication
  - IoT

Massively increased complexity
Future Mobile Architecture for Industrial Communication

Intelligent High Accuracy Indoor Positioning (HAIP)

Unlicensed and sub-licensed band

Private 4G/5G local wireless access points

Integrated High Accuracy Indoor Positioning (HAIP)

Edge-Cloud

New IoT authentication mechanisms

Interface to local applications (ERP/MES/PLM/CIM/CAx)

Automation Gateway (Operational Technology)

Private 4G/5G Base Transceiver Station including local Gateway and integrated PaaS, IoT Platform and Analytics

Public 4G/5G

Internal Enterprise Cloud

Hybrid Cloud with public share

Multi-Operator Environment

End-to-End (E2E) Industrial Slice QoS via public infrastructure

Provider Remote Operation Center

Real-time remote control and maintenance

Certificate provider for industrial communication

Public Cloud

Public 4G/5G

Private 4G/5G local wireless access points

Factory A

Factory B

Mobile Devices

x500

IC4F

INDUSTRIE 4.0

PAiCE

Fraunhofer Heinrich Hertz Institute

slawomir.stanczak@hhi.fraunhofer.de
5G Wireless Access

Massive Sensor Networks for Machine and Process Monitoring

Industrial Cloud

Edge Cloud Technology

High Accuracy Positioning

Unlicensed Wireless Access

Secure Connectivity

Ultra reliable and Low Latency

Communication for Intralogistics

PKI Solutions

Edge Cloud Technology

ML-based Data Analytics

MES/ERP Systems

Massive Sensor Networks for Machine and Process Monitoring

Slawomir.Stanczak@hh.fraunhofer.de
The largest data records are not generated by companies in the Internet industry such as Google and Facebook, but by production technology systems. McKinsey

Measuring systems in the industry require

- new rules for secure and reliable data acquisition, processing and transmission
- new measuring methods with huge amounts of data and traceable data processing
Machine Learning in Communications

These methods are widely used in traditional applications for optimization, identification, adaptation and prediction.

- **Smart Infrastructure & IoT**
  - Emergency Communications
  - Resource Management
  - Smart Home
  - Monitoring & Forecasting
  - Smart Grid

- **Communications Networks**
  - Routing
  - Traffic Identification
  - MIMO-OFDM
  - Inference Control
  - Cognitive Radio
  - Positioning
  - Power Control
  - Channel Estimation
  - Localization
  - Autonomous Systems

- **Wireless Communication**
  - PAPR Reduction
  - Power Control
  - Localization
  - Autonomous Systems
  - Satellite

- **Image & Video Communication**
  - Surveillance
  - Tracking
  - Search
  - Image & Video Coding

- **Security & Privacy**
  - Fraud Detection
  - Privacy-Preserving
  - Spam

- **Green IT**
  - Resource Management
  - Smart City

ML methods are widely used in traditional applications for optimization, identification, adaptation and prediction. These methods are also used in many other applications.
Potential Benefits of ML for 5G

• to enable us to cope with a massively increased complexity
  ➔ diminish mismatch between model and reality

• to reduce # measurements and facilitate robust decisions
  ➔ enabling massive connectivity, MIMO and mmWave

• to facilitate self-organizing networks
  ➔ cognitive network management

• to provide robust predictions
  ➔ QoS prediction, anticipatory networking
A potential future scenario

Control & user plane split

Prediction of capacity map

Use context information
Trajectory prediction  Localization

Massive MIMO

This can be a wireless link

Data transfer

slawomir.stanczak@hhi.fraunhofer.de
Demands on ML for 5G

- Robust online ML with good tracking capabilities
  - ML with small (uncertain) data sets and

- Exploit domain knowledge (e.g. models, correlations, AoA)
  - Hybrid-driven ML approaches
  - Learn features that change slowly over frequency, time...

- Distributed learning for efficient usage of scarce resources
  - New functional architectures for Big Data analytics

- Low-complexity, low-latency implementation
  - New algorithms, massive parallelization
Current View of ML tools

- The collection of training data may be limited by *physical properties of the wireless environment*
ML for Reconstruction of Capacity Maps

- Adaptive learning of long-term capacity maps
Madrid Scenario: Learning of Capacity Maps

Blue: true data rate, Red: capacity-map based prediction
Other Applications of ML for 5G

Prediction of significant changes

Traffic forecasts

Automatic network configuration

slawomir.stanczak@hhi.fraunhofer.de
Traffic forecasting is based on machine learning techniques. The algorithms incorporate prior information about the regularity of the time series and use contextual information such as information about holidays and weekends.
Energy-Saving Optimization

Traffic Forecast

Optimize the Network Configuration

Save Energy

Source: A GreenNets project deliverable

Avrg. Load [%]

Power Consumption [kW]

Time of Day

Source: A GreenNets project deliverable

Total Power Consumption

Total Power Consumption Switch-Off

Avrg. GSM Load

Avrg. UMTS Load

Avrg. LTE Load

Fraunhofer Heinrich Hertz Institute

slawomir.stanczak@hhi.fraunhofer.de
Energy-Saving Optimization

- Simulation area: 20 km x 20 km
- Number of iterations in our algorithm: 10
- Single-RAT optimization (LTE)
- Using CPLEX in the iteration of the algorithm
- Conventional laptop (Core i7 with 4GB of ram)

<table>
<thead>
<tr>
<th># Cells</th>
<th># test points</th>
<th>Considered cells per test point</th>
<th># opt var</th>
<th>Time [s]</th>
<th>Memory usage [%]</th>
<th>#active cells after optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>900</td>
<td>20.000</td>
<td>900</td>
<td>1.5 mio</td>
<td>276</td>
<td>70-80</td>
<td>440</td>
</tr>
<tr>
<td>900</td>
<td>20.000</td>
<td>10</td>
<td>0.19 mio</td>
<td>41</td>
<td>30-40</td>
<td>440</td>
</tr>
<tr>
<td>900</td>
<td>20.000</td>
<td>5</td>
<td>0.09 mio</td>
<td>29</td>
<td>30-40</td>
<td>516</td>
</tr>
</tbody>
</table>

Source: www.communicate-green.de
Machine Learning in the Physical Layer / Short-Term Learning
Example: 5G NOMA

- Conventional approach:
  1. Send pilots
  2. Estimate channels and other system parameters
  3. Construct, for example, a linear filter (receiver)
  4. Detect the symbols with the filter in step (3)

\[ f : \mathbb{C}^M \rightarrow \mathbb{C} \]

(One filter for each user)

---

slawomir.stanczak@hhi.fraunhofer.de
Example: 5G NOMA

• An ML approach:
  
  - (1) Send pilots
  
  - (2) Estimate channels and other system parameters
  
  - (2) Use the pilots to train an ML tool to detect the symbols directly (no need to obtain channel state information, for example)
  
  - (3) Detect the symbols with the ML tool
  
(One nonlinear filter for each user)

\[ f_{ML} : \mathbb{C}^M \rightarrow \mathbb{C} \]
A Fundamental Challenge

- Many unknowns: fast-time varying channels, varying number of users (inter- and intra-cell interference), changing modulation and power, etc.

- Consequences:
  1. Training samples are highly limited (hundreds). Deep neural networks often require hundreds of thousands or more samples.
  2. Training and detection have to be performed within the coherence time, or ML tools have to learn (or be given) time-invariant features.

\[ f^{(1)} : \mathbb{C}^M \rightarrow \mathbb{C} \quad f^{(2)} : \mathbb{C}^M \rightarrow \mathbb{C} \]

Time

Coherence time (e.g. 10ms)

---

slawomir.stanczak@hhi.fraunhofer.de
Example: Detection for 5G NOMA

- **Goal**: Use pilots to learn the receiver structure directly
- **Approach**: Online adaptive filter in the sum space of linear and Gaussian reproducing kernel Hilbert spaces
ML-based Solution for 5G NOMA

- Initial fast convergence and low complexity
- Easy to exploit side information and convergence guarantees


slawomir.stanczak@hhi.fraunhofer.de
Cooperative Deep Learning

Collaborative Compressive Inference for 5G
Georg Hieronimus, Steffen Limmer, and Sławomir Stańczak

Harness interference for computation

Camera Classification Compression

Decoder Fusion

Base Station with NVIDIA GPUs

NVIDIA Jetson: Supercomputers on the Edge

slawomir.stanczak@hhi.fraunhofer.de
Cooperative Deep Learning

Objectives (e.g. for industrial applications):

• High compression -> low latency
Take-away Message

• ML will help to cope with the increasing complexity of 5G

• 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 …

• Projection methods for CFP allow massive parallelization
  ➔ Low-latency implementations using GPUs possible

• No time and data for extensive training of DNN
  ➔ Design good NN architectures for a given task
ITU-T **Focus Group** on Machine Learning for Future Networks including 5G (FG-ML5G)

Established by ITU-T SG13 in November 2017 to:

- Identify relevant standardization gaps in order to improve interoperability, reliability and modularity of Machine Learning (ML) for 5G
- Draft technical reports & specifications for ML for future networks, including interfaces, network architectures, protocols, algorithms and data formats
- Analyse the impact of the adaption of ML for future networks (e.g. autonomic network control and management)

First meeting:
30 January – 2 February 2018 (Geneva);
preceded by a workshop on **ML for 5G and beyond**, on 29 January 2018

Remote participation will be provided

Chairman:
- Slawomir STANCZAK (Fraunhofer HHI, Germany)

Vice-chairmen:
- Charles Chike ASADU (University of Nigeria)
- Seongbok BAIK (KT, Republic of Korea)
- Viliam SARIAN (NIIR, Russian Federation)
- Mingjun SUN (CAICT, People's Republic of China)

“Machine learning and artificial intelligence are finding promising applications in communications networking,” says the Focus Group’s Chairman, Slawomir Stanczak of Germany’s Fraunhofer Heinrich-Hertz-Institut.

“This Focus Group will establish a basis for ITU standards experts to capitalize on machine learning in their preparations for the 5G era.”

http://news.itu.int/itu-launches-new-focus-group-study-machine-learning-5g-systems/

slawomir.stanczak@hhi.fraunhofer.de
FG ML5G: Deliverables

Three working groups:
• **WG1: Use cases, services & requirements**
  • Deliverables
    – Use cases
    – Ecosystem, terminology and services.
    – Requirements and standardization gap

• **WG2: Data formats & Machine Learning technologies**
  • Deliverables
    – ML algorithms in communication networks: categorization, terminology & implications
    – Data formats including privacy and security aspects for ML in communication networks
    – Standardization and technology gaps

• **WG3: Machine Learning-aware network architecture**
  • Deliverables
    – Analysis of communication network architectures from the viewpoint of ML
    – Description of ML-related functions, interfaces and resources for communication network architectures
    – Standardization and technology gaps
How to participate: Website, Collaboration site

• **FG ML5G website**
  Participation in FG-ML5G is free of charge and open to all.

• **Access to FG-ML5G Collaboration Site**
  to see meeting documents, agendas, reports, etc.

  **How to access:**
  1. A [TIES](#) or [Guest](#) account is required to access FG documents and subscribe to the FG-ML5G mailing lists. [Sign up here](#)
  2. Alternatively, sign in to [MyWorkspace](#) and use the "Mailing lists" feature for one-click subscription (search for "fgml5g")