STEFAN BOSSE[1,2]

[1]Fachbereich Mathematik und Informatik

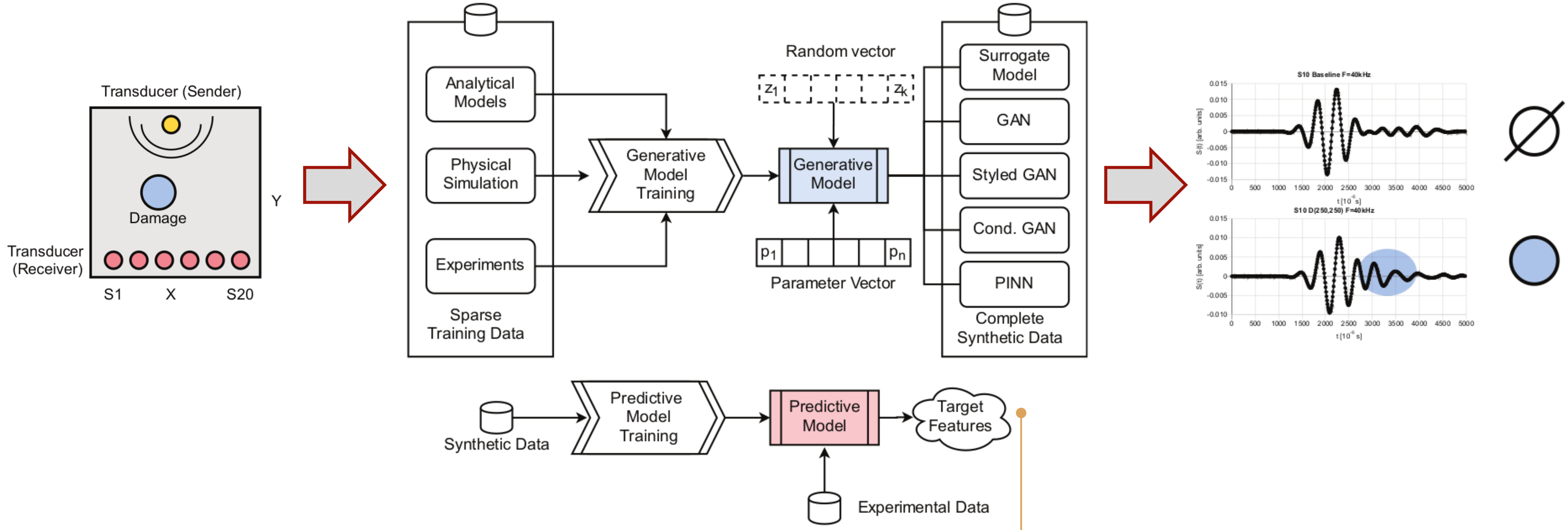
Universität Bremen

sbosse@uni-bremen.de www.edu-9.de

[2]Fakultät Maschinenbau

Universität Siegen

Universität Bremen



DATA-DRIVEN PARAMETERIZABLE GENERATIVE ADVERSARIAL NETWORKS FOR SYNTHETIC DATA AUGMENTATION OF GUIDED ULTRASONIC WAVE SENSOR SIGNALS

Stefan Bosse

EWSHM 2024 - 12.06.2024

CONTENT

01 **Introduction**
Data-driven SHM Modeling
Constraints
Experimental versa Synthetic Data

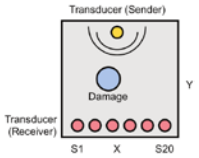
02 **Methodoloy (1)**
Training of ML Models with Synthetic Data and Prediction with Real Measured Data

03 **Methodology (2)**
Synthetic Data by Generative Models and Simulation

04 **Generative Model Zoo**
CNN, GAN, Styled GAN, and more

05 **Results**
The bad and the good news – Better than expected, worser than required!

06 **Conclusions**
Identified Challenges
What is possible? To do..
Outlook, Roadmap

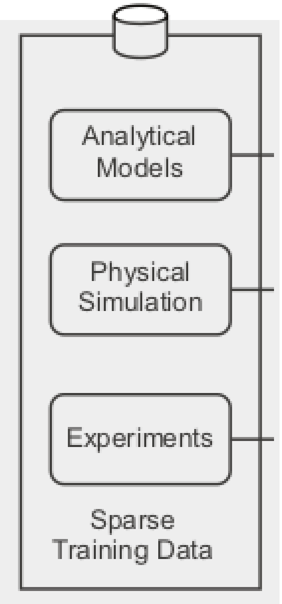


INTRODUCTION: NON-DESTRUCTIVE TESTING USING GUW

- **Detection of hidden damages, defects, and impurities (e.g., pores, cracks, delaminations) is still a challenge using GUW**!
- Example Impact Damage in multi-layer materials and laminates: Combination of different damages , i.e., cracks, delaminiation, change of material and layer thickness, damages in different layers, and so on.
- Time-resolved GUW signals are a superposition of different wave-damage interactions! Damage features are hard to be isolated from the base-line signal.
- Data-driven Modeling of damage predictor models depends strongly on measured training data.

Primary Goal. Automated Damage, Defect, and Impurity Detection in Materials and Structures including Composites using Data-driven Damage Predictor Models and GUW Signals.

INTRODUCTION: DATA



.

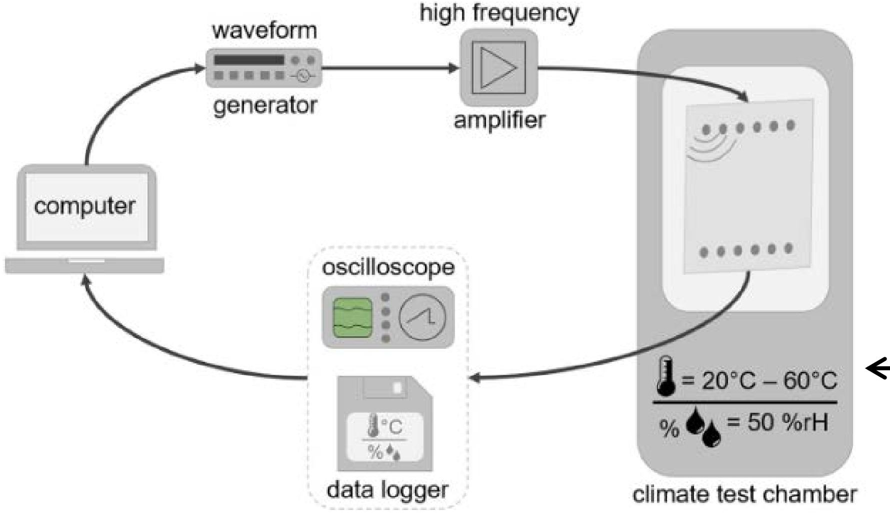
- **Data-driven Modeling of damage predictor models depends strongly on training data**
- But Data Space is sparse with respect to:
  - Geometrical Properties
  - Environmental Properties
  - Sensors, Transducers, Pitch Signal Properties
  - Damage Properties

Secondary Goal. Synthetic Data Generation constrained and controlled by Real Data from Experiments, Training of ML Models using Synthetic Data only.

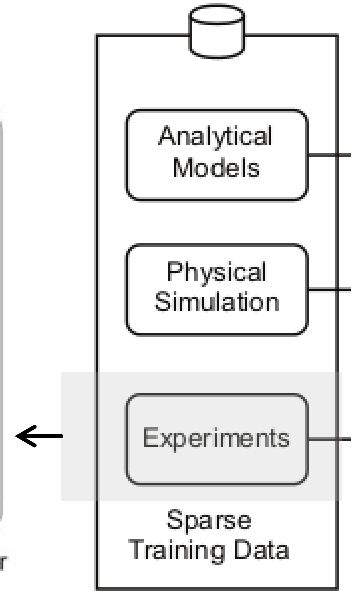
INTRODUCTION: DATA

.

- But **experimental data** space is sparse:
  - Only few transducers and positions
  - Only few measuring paths
  - Only few or no variation of environmental parameters (temperature, humidity, tense and so on)
  - Only few damage cases (position, strength, size and so on) // 1 Impact Damage : 1 Specimen!
  - Limited reproducibility (drift, changes of specimens, environmental parameters, sensors… over time)



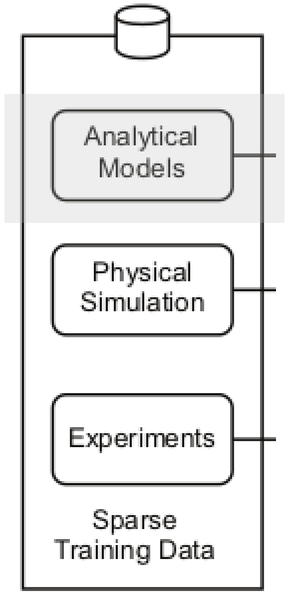
Moll et al., 2019



Secondary Goal. Synthetic Data Generation constrained and controlled by Real Data from Experiments, Training of ML Models using Synthetic Data only.

Universität Bremen

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial x^2}-\frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2}=0$$



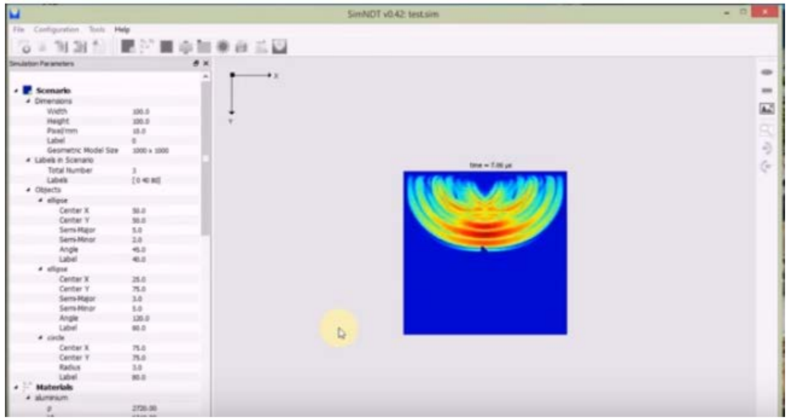
INTRODUCTION: DATA

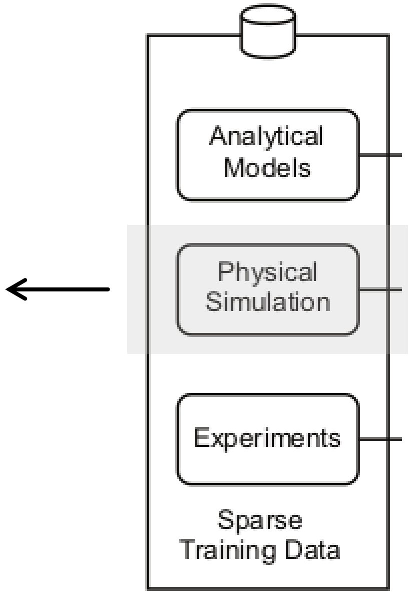
.

- But **analytical data** space is sparse:
  - Oversimplification of Physical and Material Models, commonly only macroscopic aggregates
  - 3-dim damage-wave interaction is hard to be modeled, especially in multi-material and multi-layer structures
  - Environmental parameters can be considered partially
  - Variation requires Monte Carlo simulation techniques
  - Boundary reflections are commonly not considered

Secondary Goal. Synthetic Data Generation constrained and controlled by Real Data from Experiments and Physical Models, Training of ML Models using Synthetic Data only.

INTRODUCTION: DATA





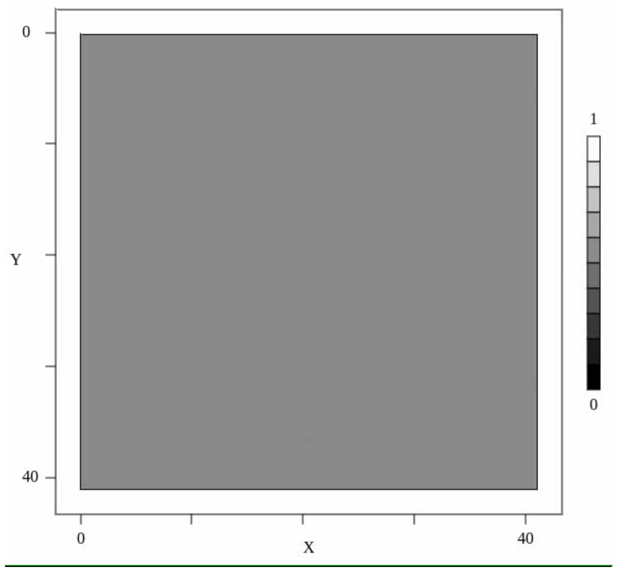
.

- But **simulation data** space is sparse:
  - Oversimplification of Physical and Material Models, but visco-elastic finite integration technique is promising
  - 3-dim damage-wave interaction is hard to be simulated, especially in multi-material and multi-layer structures
  - Environmental parameters can be considered partially
  - Variation requires Monte Carlo simulation techniques, difficult to be handled in a time-discrete simulation
  - Boundary reflections are commonly considered

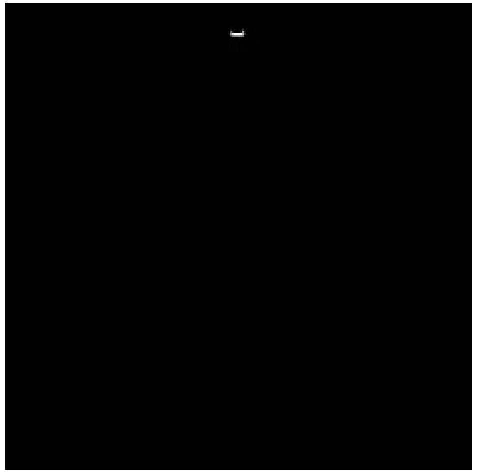
Secondary Goal. Synthetic Data Generation constrained and controlled by Real Data from Experiments and Physical Models, Training of ML Models using Synthetic Data only.

INTRODUCTION:
THE REALITY GAP

Air-coupled US Scan
FML Plate



2D Simulation
Aluminum Plate



.

- There is always a reality gap between **real measurements** and
  - **Simulation!**
  - **Analytical Modeling!**
  - **Synthetic Data Augmentation using Experimental Data?**
  - **Synthetic Data Generation using Random Process-driven Generative Models?**

Secondary Goal. Synthetic Data Generation constrained and controlled by Real Data from Experiments and Physical Models, Closing the Reality Gap.

Universität
Bremen

SIMULATION

.

- Used in this work to create a reference signal base to train and test generative models (**ground truth data**)
- The signals should be as simple as possible without complex patterns to enable comparison of generated and simulated signals (are they real or silly fakes?)
- Two classes of “simulated” GUW signals:
  - Pitch signals generated by ground-truth mathematical function (windowed sine waves)
  - Catch signals computed with a 2-dim visco-elastic finite integration method simulator (based on SimNDT)

SIMULATION: ANALYTICAL SIGNAL MODEL

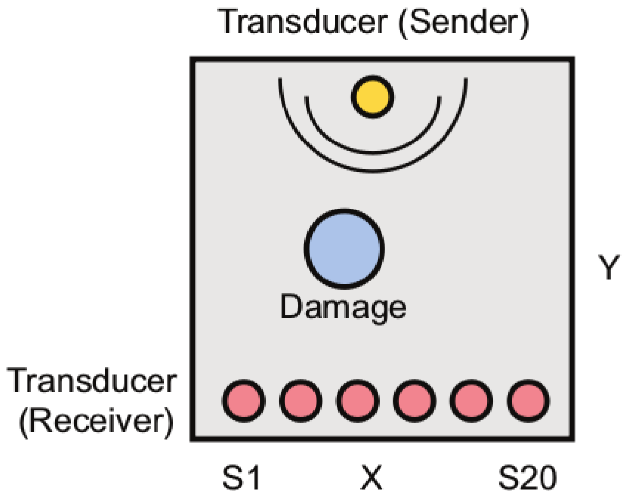
- Pitch (Sender) Signal
  - Base signal: Sine wave S
  - Mask: Gaussian window function M
  - Generator function: G

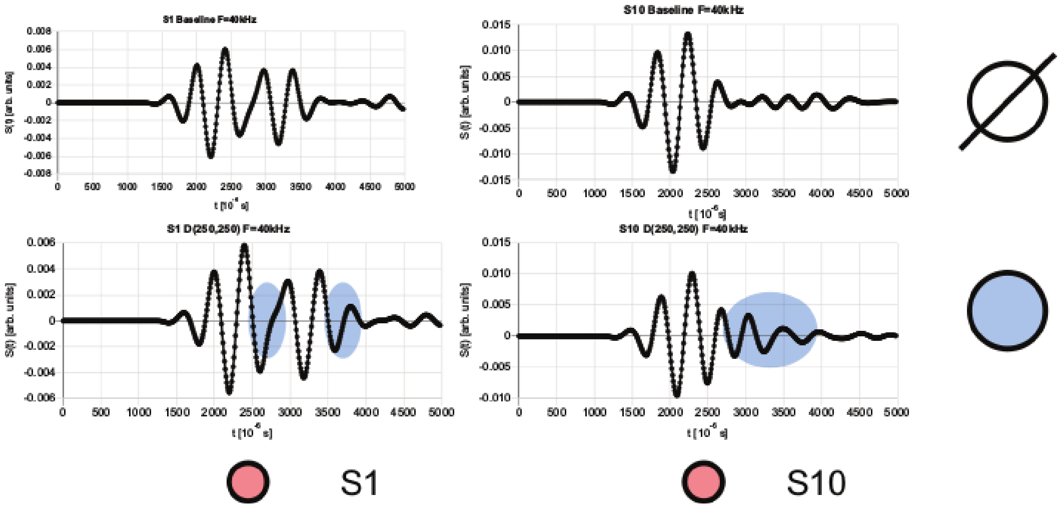
$$
\begin{gathered}
\vec{I} = (1, 2, .., N) \\
\vec{S} = \sin\left(\frac{\vec{I}}{P} 2\pi\right) \\
\sigma = kW \\
\vec{J} = \frac{\vec{I} - O}{0.5N} 2\sigma \\
\vec{M} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\vec{J}^2}{2\sigma^2}} \\
\vec{G} = \vec{S}\vec{M}
\end{gathered}
\tag{1}
$$

Universität Bremen

SIMULATION: VISCO-ELASTIC WAVE PROPAGATION MODEL

- Based on SimNDT OpenCL/GPU Solver (2-dim Finite Integration Method)
  https://github.com/mmolero/SimNDT http://git.edu-9.de/sbosse/SimNDT2
- 1 Sender Transducer, Array of 20 Receiver Transducers, Host: Aluminum, Defect: Hole (Air)

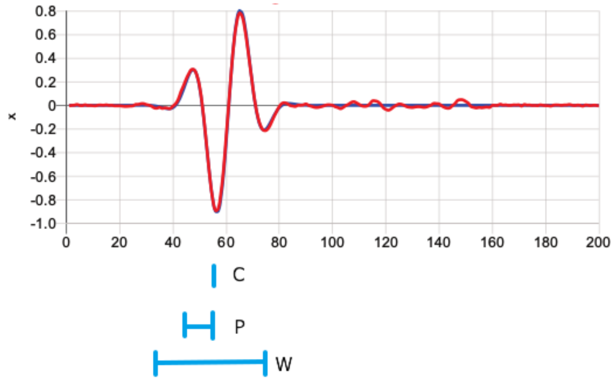


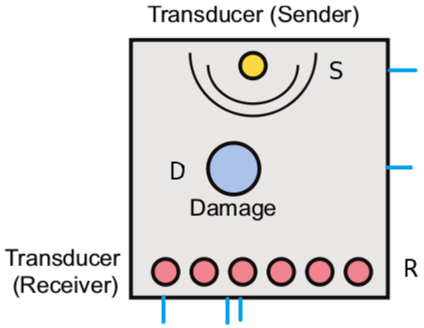


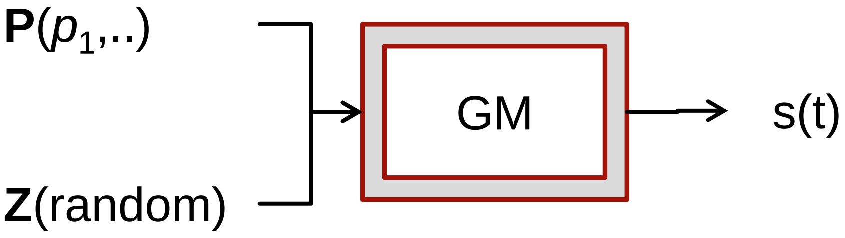
GENERATIVE MODELS

Goal. A parametrizable Generative Model with a random process but constrained by experimental (or simulated) data (GUW signals).

- **Input**: Model Parameters, e.g., for a
  Pitch Signal Model: **P**=(*p*,*w*,*c*), i.e., *p*:period of the sine wave, *w*: width (number of cycles), *c*:center position on time axis
  Catch Signal Model: **P**=(*p*,*w*,*C,E,D,R,S*), i.e., *E*: environment/temperature, *D*:damage location, *S*: sender position, *R*: receiver position, *M*: material parameters?
- **Output**: GUW signal without linear dependency to training data (example) signals!



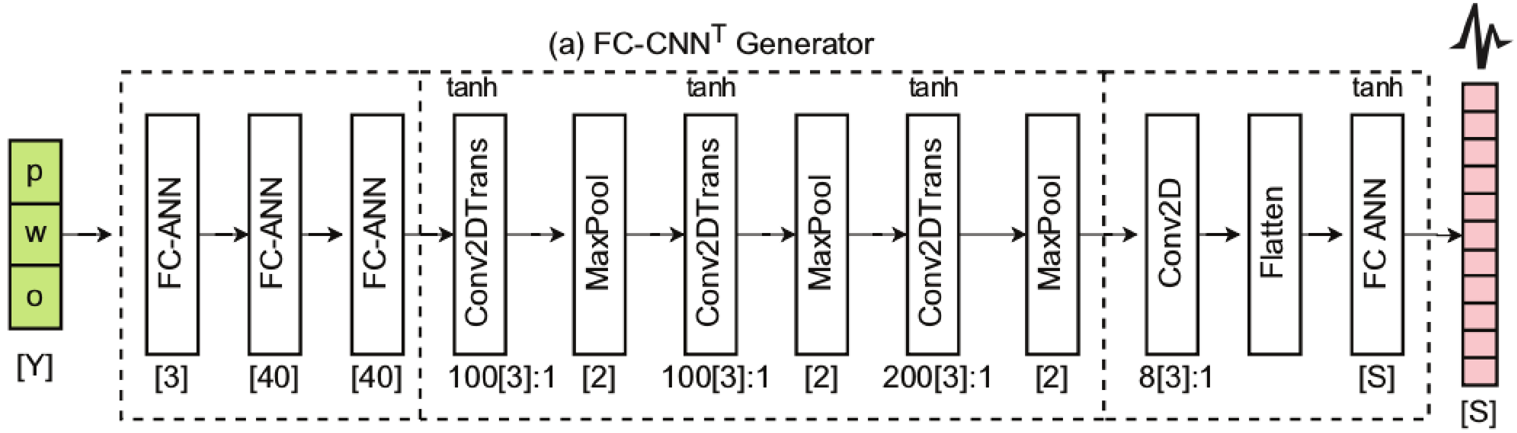


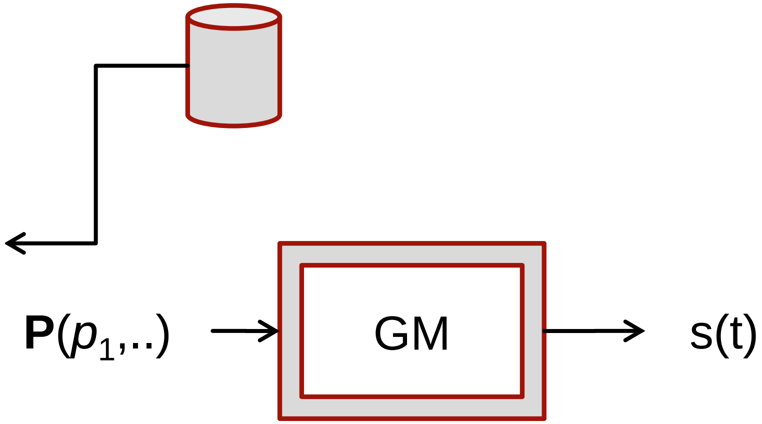


The entry level

GENERATIVE MODELS: SURROGATE FC-CNN^T MODEL

- **Question 1**: Is there is a Parametrizable Generative Model trained with example data capable to represent the entire parameter space?
  - Question 1.1: Is interpolation between points (examples) in the parameter space possible?
  - Question 1.2: Is extrapolation beyond convex hull points (examples) in the parameter space possible?
- **Input**: Model Parameters
- **Training Data**: Pitch signal from analytical model
- **Output**: GUW Signal
- **Architecture**: 3 Fully Connected Neuronal layers, 3 Transposed Convolutional/Max-Pooling layers, 1 Convolutional and 1 final FC layer

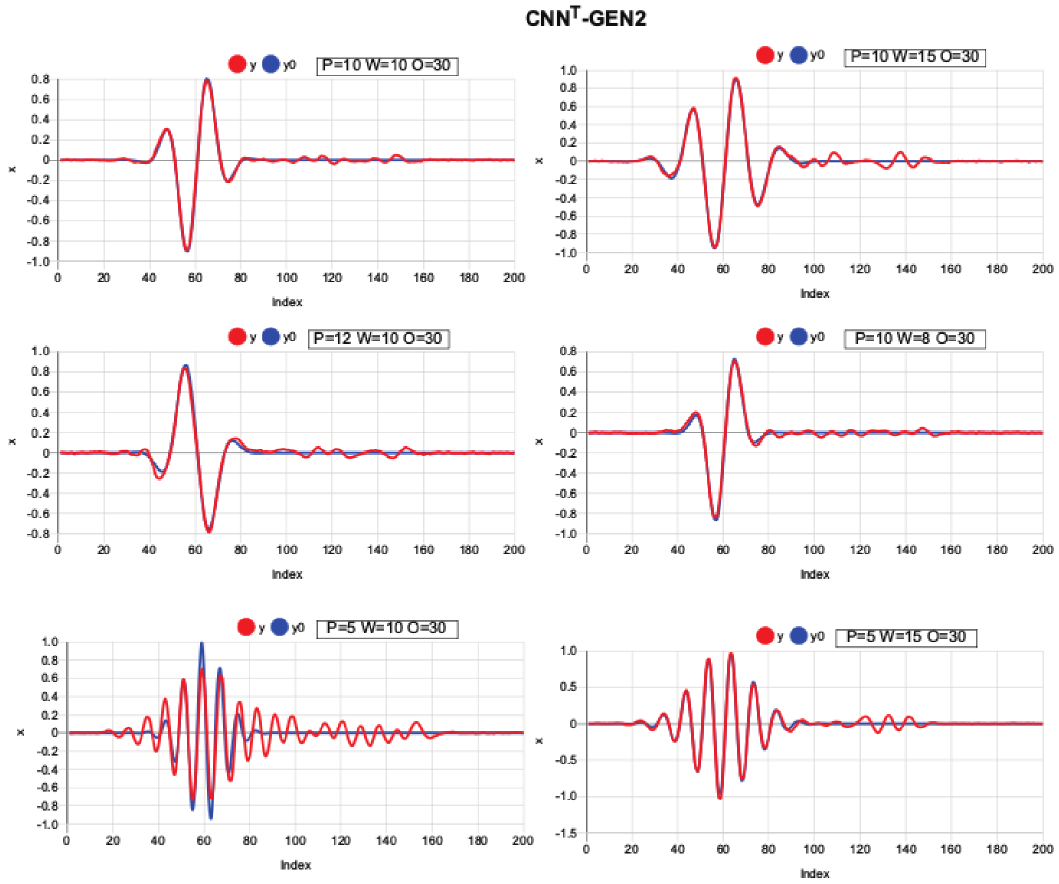




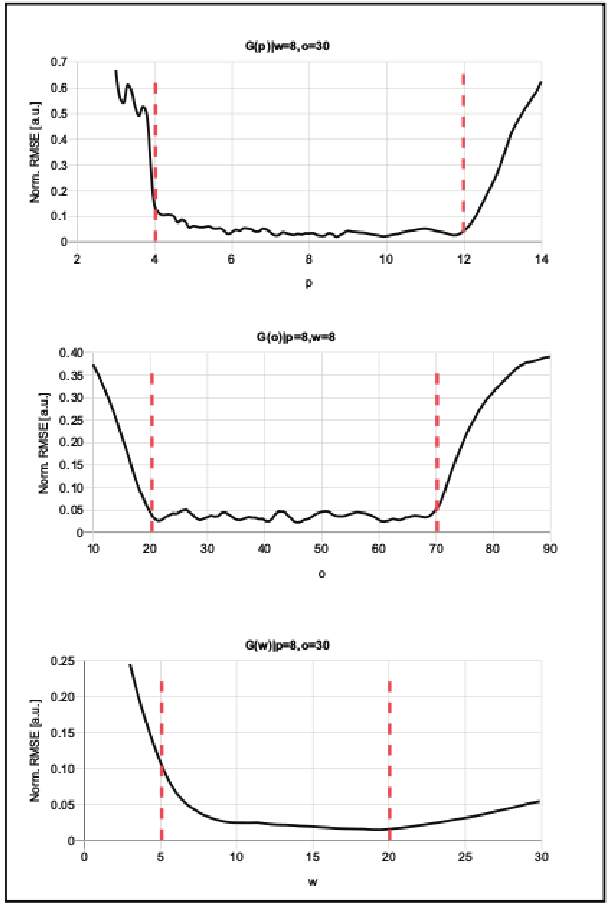
GENERATIVE MODELS: SURROGATE FC-CNN^T MODEL

- **Answer 1**: Yes
- Answer 1.1: Yes
- Answer 1.2: No
- Physical correctness: Yes
- Control parameters: Linear independent

> The model diverges/explodes quickly if we leave the convex parameter hull! The surrogate model not learned the analytical model!



(a)



(b)

The basic level

GENERATIVE ADVERSARIAL NETWORK (GAN)

- **What is that?** A random-process driven generative model
- **Use caes (typically)?** Generation of <u>**fancy**</u> synthetic data learned from real data
- **How does it works?** The generator is trained by a discriminator. The discriminator only decides if a sample is real or fake. The generator is trained to generate fakes (fooling us), not reals.
- **What do they require**? Large amount of real data! We don’t have this in measuring sciences…
- **Are the results physical correct?** No control over at all. Answered by this talk?

GENERATIVE ADVERSARIAL NETWORK (GAN)

- Works that uses GAN for signal generation…

My Favorite

sensors (2022) MDPI

Article

Diagnostic-Quality Guided Wave Signals Synthesized Using Generative Adversarial Neural Networks

Mateusz Heesch [1], Michał Dziendzikowski [2], Krzysztof Mendrok [1] and Ziemowit Dworakowski [1,*]

Time-Domain Signal Synthesis with Style-Based Generative Adversarial Networks Applied to Guided Waves

(2021)

Mateusz Heesch, Krzysztof Mendrok, and Ziemowit Dworakowski

biology MDPI

Article

SynSigGAN: Generative Adversarial Networks for Synthetic Biomedical Signal Generation

Debapriya Hazra [†] and Yung-Cheol Byun [*,†]

(2020)

SCIENTIFIC REPORTS

OPEN

Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network

(2019)

Fei Zhu [1,2], Fei Ye [1], Yuchen Fu [3], Quan Liu [1] & Bairong Shen [4]

Generative Adversarial Network for Radar Signal Generation

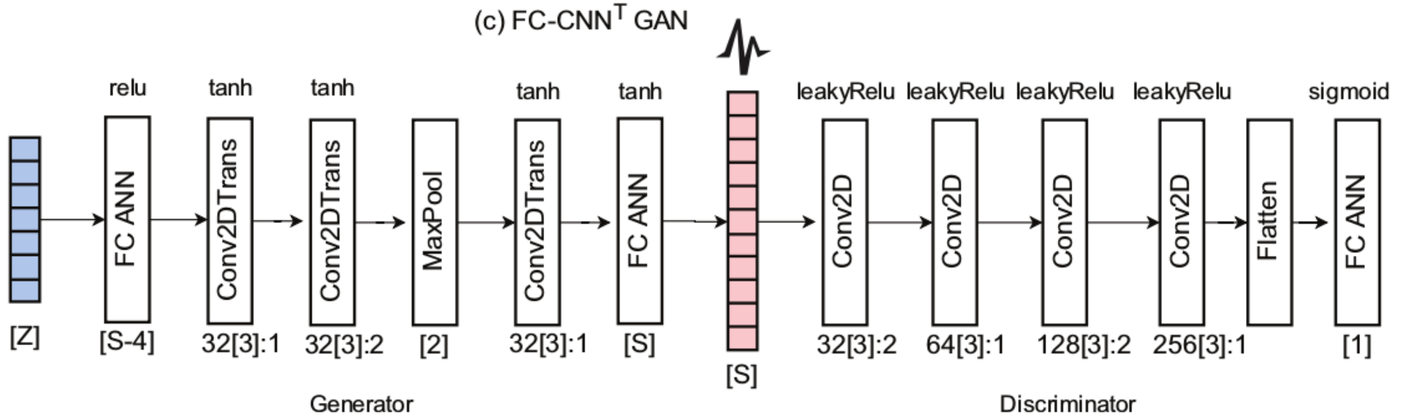
(2020)

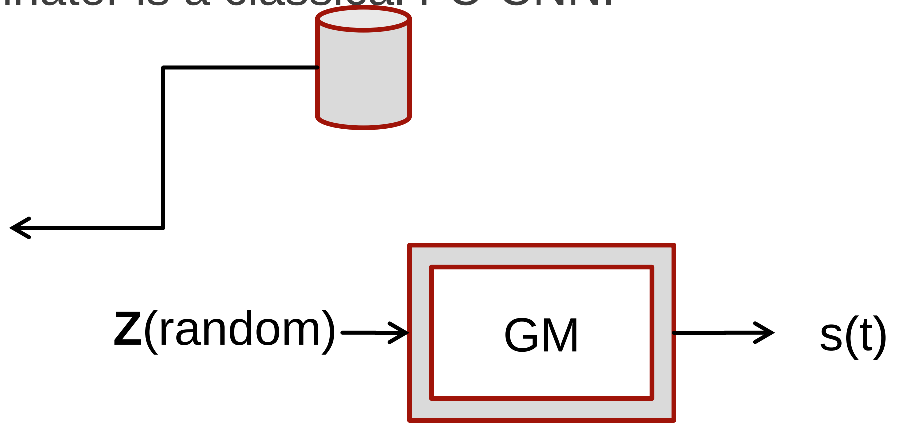
Thomas Truong Svetlana Yanushkevich

The next level

GENERATIVE MODELS: RANDOM GAN FC-CNN^T MODEL

- **Question 2**: Can a random-process driven Generative Model trained with example data produce realistic (“physical correct”) signal data?
- **Input**: Random vector
- **Training Data**: Pitch signal from analytical model
- **Output**: GUW signal
- **Architecture**: Generative Adversarial Network (GAN), Two Models, Generator and Discriminator. Generator consists of transposed Convolution layers, Discriminator is a classical FC-CNN.

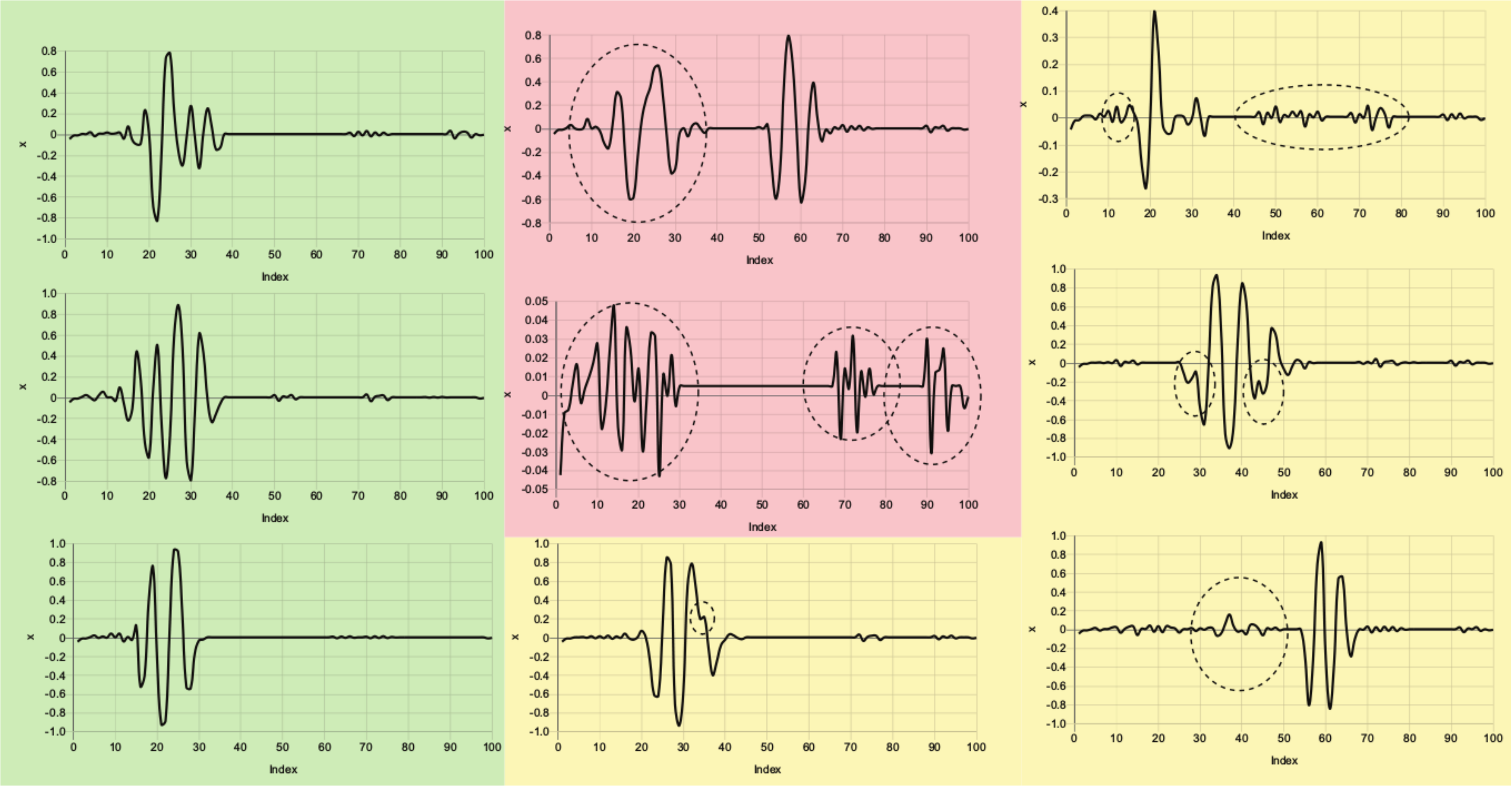




GENERATIVE MODELS: RANDOM GAN FC-CNN^T MODEL

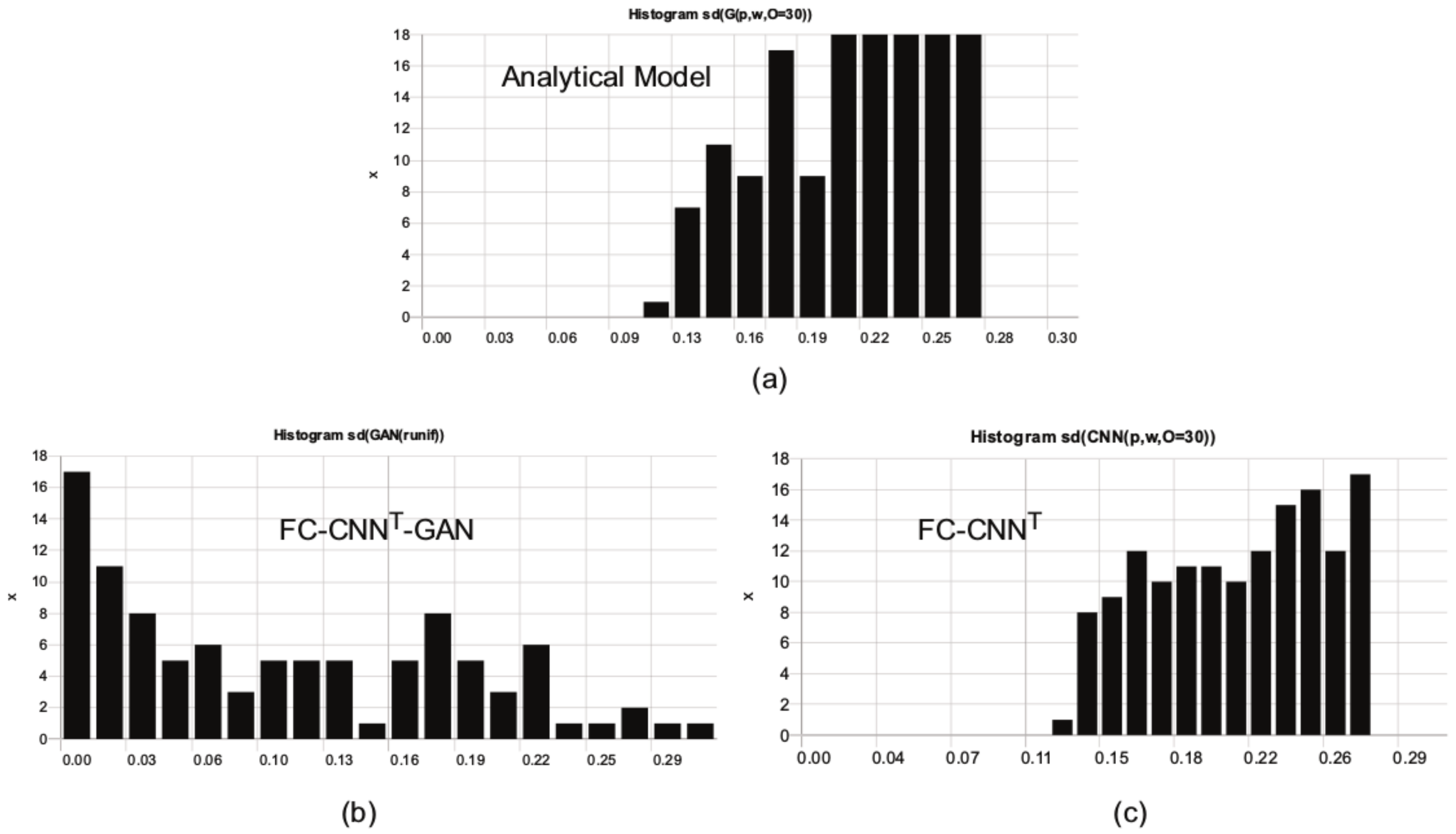
- **Answer 2**: No (not satisfying)
- Physical correctness: No
- Control parameters: No

The GAN produces artifacts and distortions.
The random GAN model not learned the analytical model!
Reason: No direct feedback of training data to the generator.



GENERATIVE MODELS: RANDOM GAN FC-CNN^T MODEL

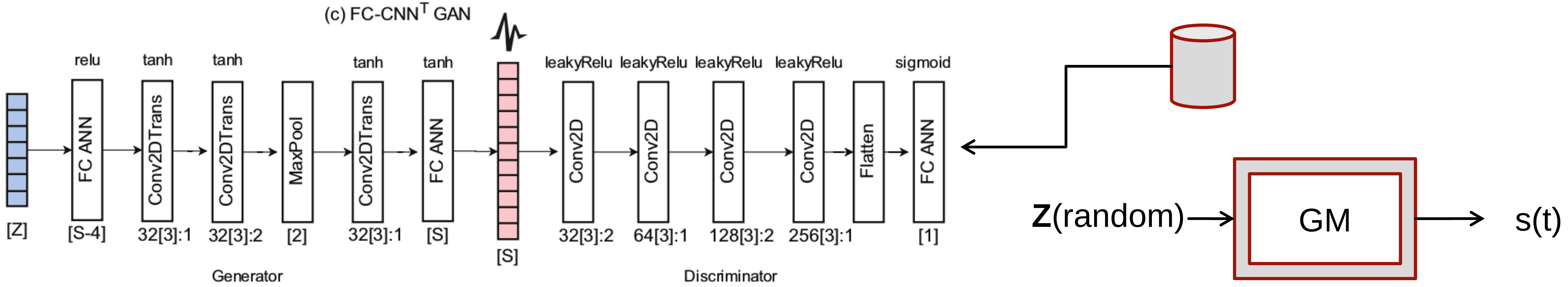
- Comparison of the histogram distribution of the standard deviation of pitch signals generated
- (a) the analytical model
- (b) the random controlled FC-CNN^T-GAN model trained with the data from the analytical model
- (c) the surrogate FC-CNN^T model
- All histograms are in the range [0,0.3]



The adventure level

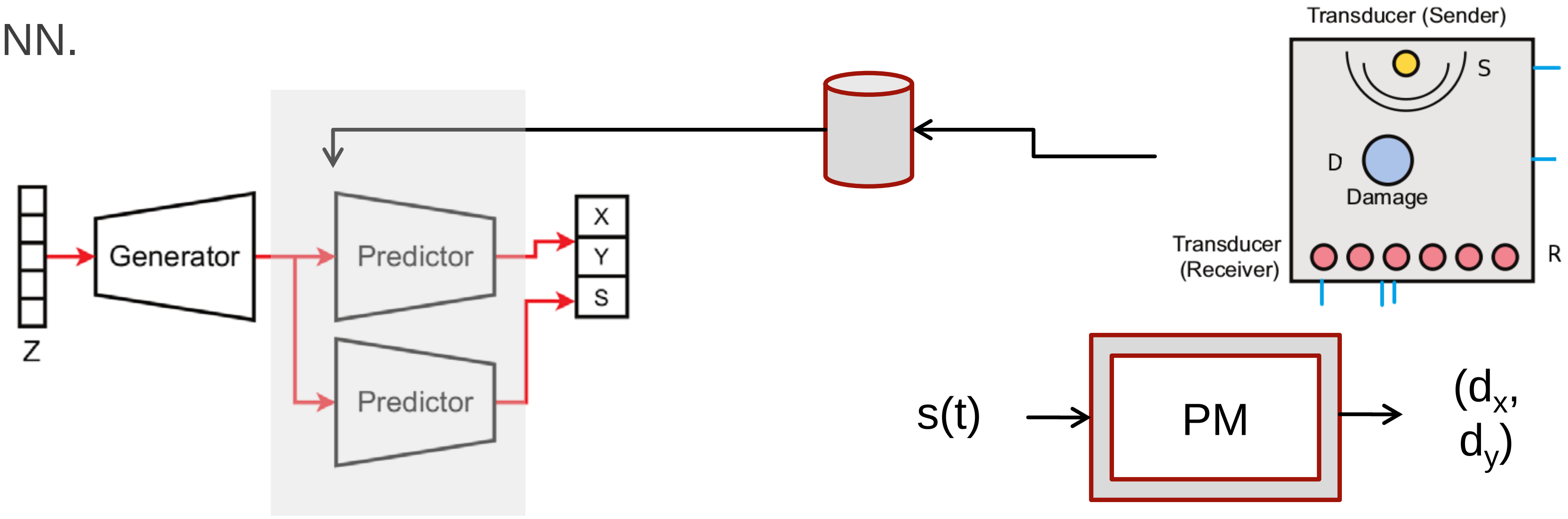
GENERATIVE GUW MODELS: RANDOM GAN FC-CNN^T MODEL

- **Question 3**: Can a random-process driven Generative Model trained with example data produce realistic (“physical correct”) catch signal data?
- **Input**: Random vector
- **Training Data**: GUW signals from 2-dim simulation, different damage and sensor positions
- **Output**: GUW signal vector (200 points)
- **Architecture**: Generative Adversarial Network (GAN), Two Models, Generator and Discriminator. Generator consists of transposed Convolution layers, Discriminator is a classical CNN.



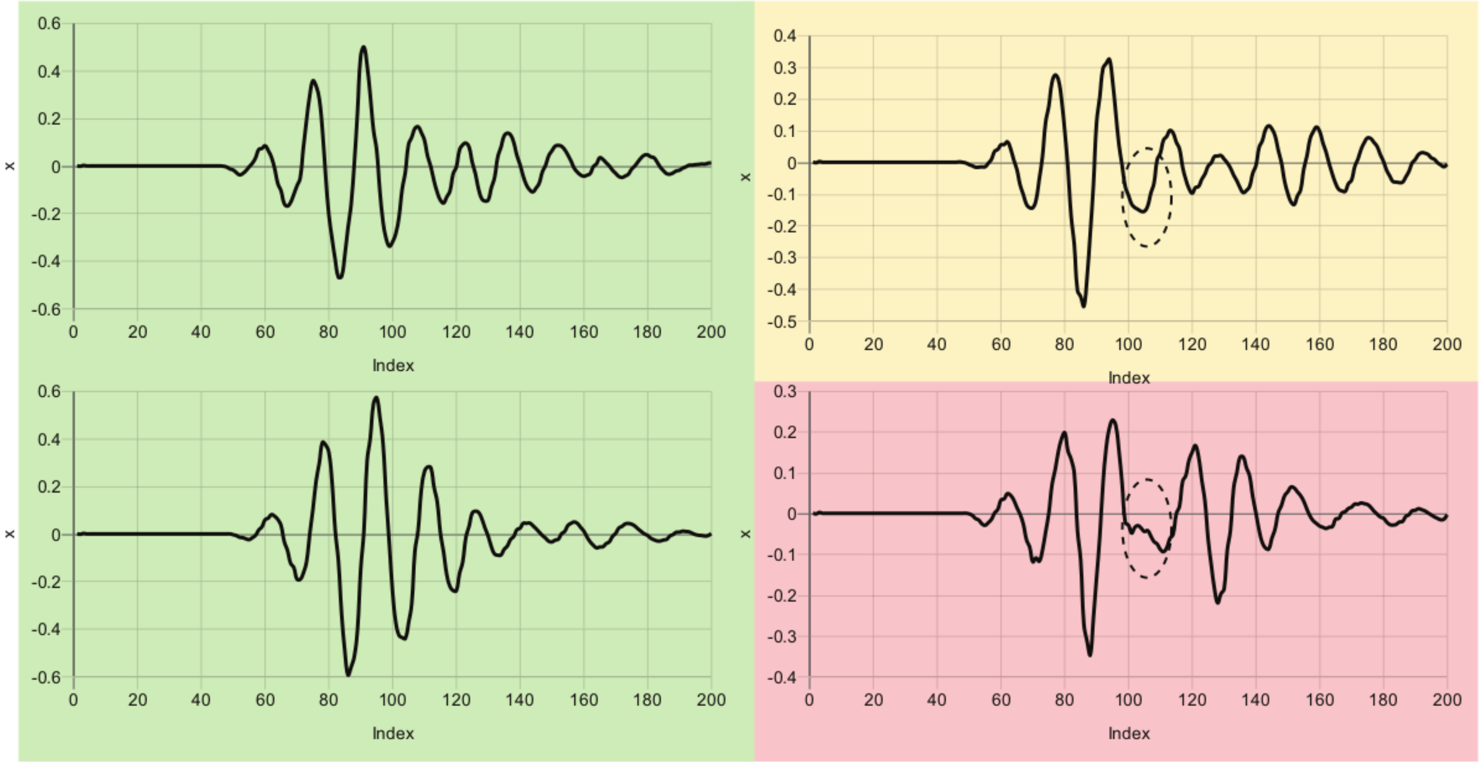
GENERATIVE GUW MODELS: RANDOM GAN FC-CNN^T MODEL

- **Question 3.1**: How can we evaluate the GUW catch signals with respect to “physical” correctens?
- **Answer 3.1**: By using a CNN regression predictor model!
- **Input**: GUW signal vector (200 points)
- **Output**: Two Models: 1. Regression of damage location (0/0 means no damage), 2. Sensor position
- **Architecture**: Classical CNN.



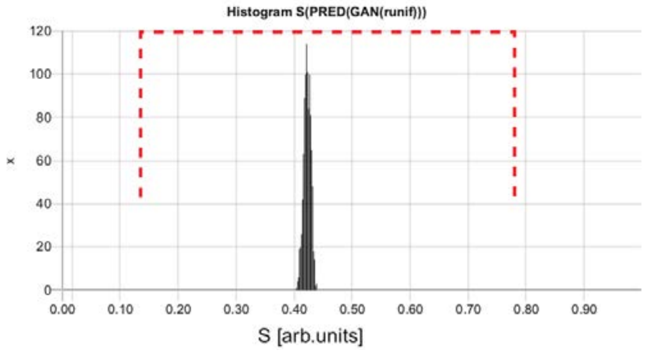
GENERATIVE GUW MODELS: RANDOM GAN FC-CNN^T MODEL

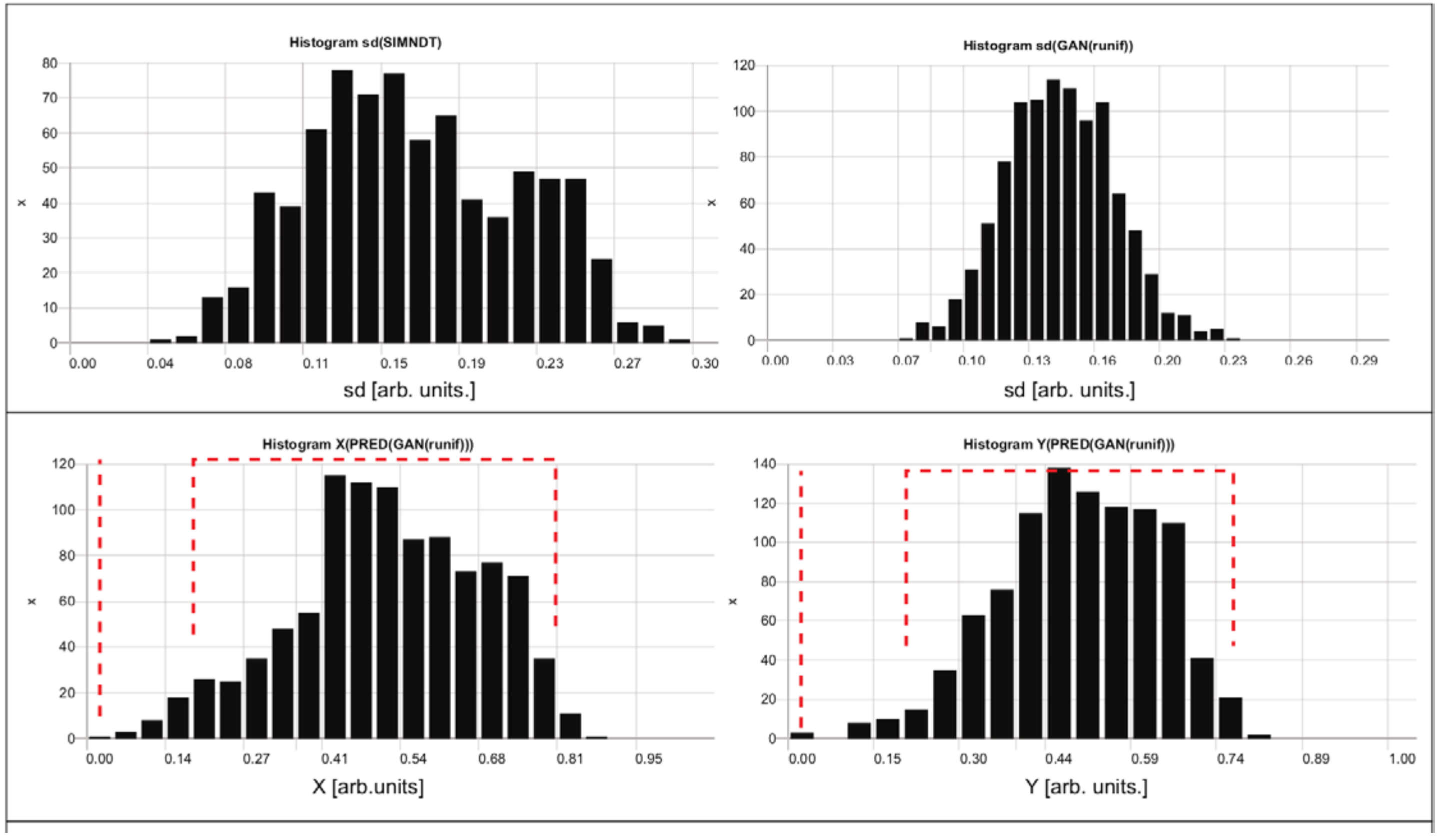
- **Answer 3**: Maybe, but again with artifacts and some physical incorrect signals
- **Question 3.2**: What is about variance with respect to a parameter space (damage and sensor positions)?



GENERATIVE GUW MODELS: RANDOM GAN FC-CNN^{T} MODEL

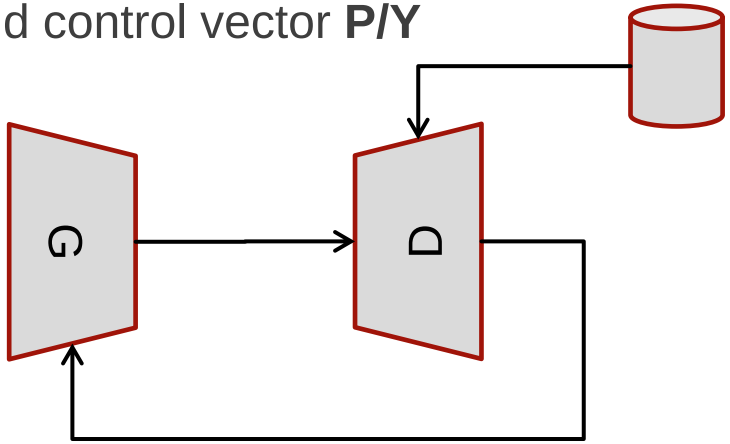
- **Answer 3.2**: Promising,
- Distribution of standard deviation of signal is comparable to ground truth data, but
- X/Y damage position show centering!
- Sensor position variance is zero!

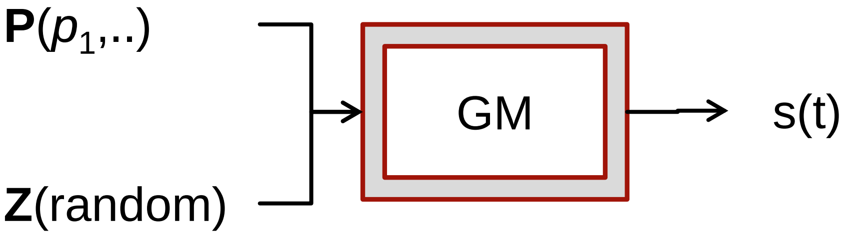




GENERATIVE GUW MODELS: CONTROLLABLE RANDOM GAN FC-CNN^T MODEL

- **Question 4**: Can a random-process be constrained by a parameter vector that drives a Generative Model trained with example data producing realistic ("physical correct") catch signal data?
- **Input**: Random vector, Parameter vector (aka. Style vector)
- **Training Data**: GUW signals
- **Output**: GUW signal vector
- **Architecture**: Generative Adversarial Network (GAN), Two Models, Generator and Discriminator. Generator consists of transposed Convolution layers, Discriminator is a classical CNN.
- There are different sub-architectures to merge the random **Z** vector and control vector **P/Y**



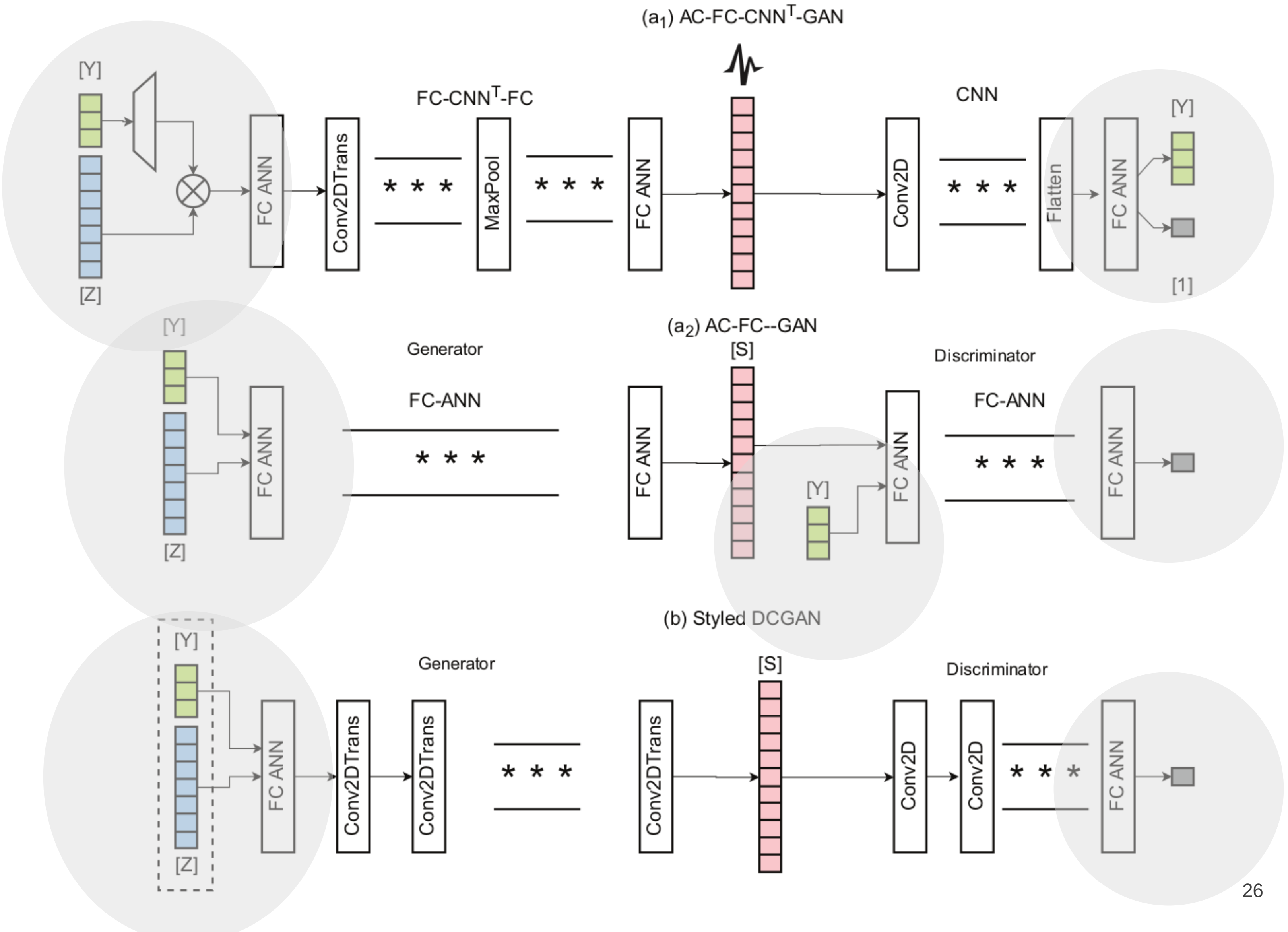


The gold level

GENERATIVE GUW MODELS: CONTROLLABLE RANDOM GAN FC-CNN^T MODEL

- **Answer 4**: Maybe
- Input and output architectures of Z/Y(P) can differ

The style vector is commonly entangled, i.e., control of individual paremeters (e.g., damage location) is not possible!



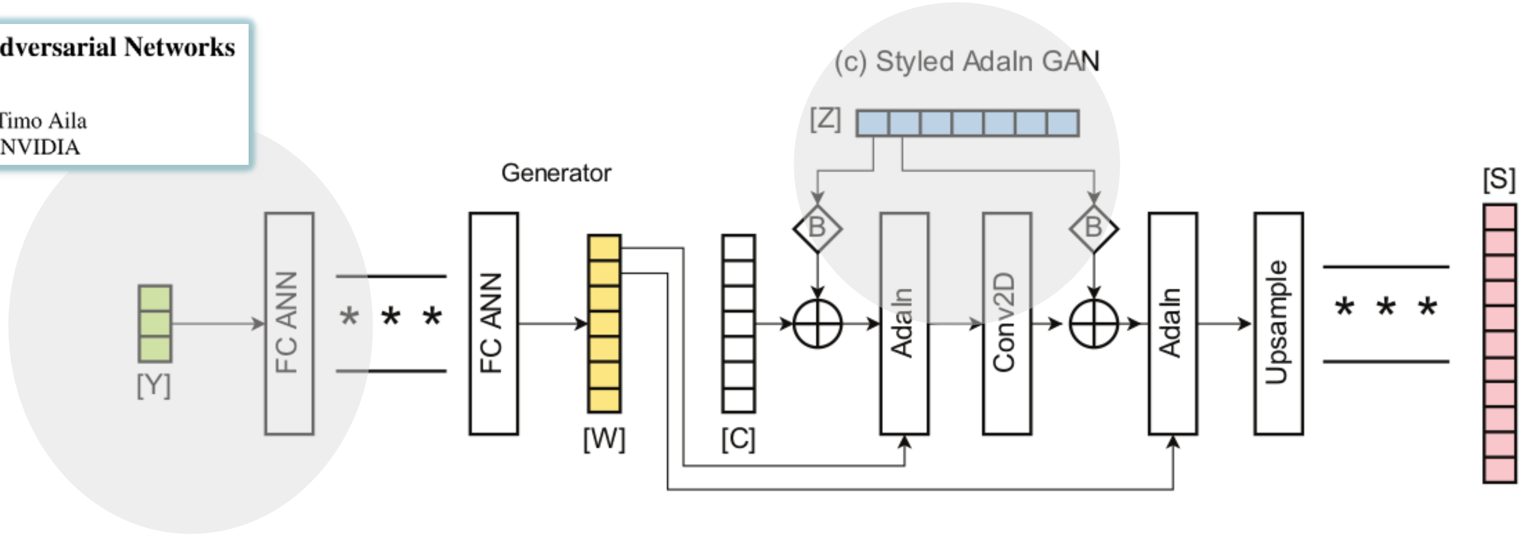
Universität Bremen

The gold level

GENERATIVE GUW MODELS: STYLED GAN FC-CNN MODEL

A Style-Based Generator Architecture for Generative Adversarial Networks

(2019) Tero Karras NVIDIA | Samuli Laine NVIDIA | Timo Aila NVIDIA



- **Styled GAN** architecture introduced by NVIDIA developers
- Adaptive Instance Normalization GAN splitting the generator into a mapping and synthesis network
- The mapping network (multi-layer FC ANN) maps the style vector Z on an intermediate latent vector w, which is the input for multiple **adaptive instance normalization layers** controlling the generation process.
- The uniform random distributed vector (Gaussian noise) is added partially to the output of multiple convolutional layers, instead passing the entire random vector to the input of the first convolutional layer

Universität Bremen

CONCLUSIONS

Don't trust data-driven modells!

Data

- **Single- and Multi-Path Guides Ultrasonic Wave Signals**
- **Data and feature variance is always limited from experiments**!
- Parameter space is sparse
- **Synthetic Signal Data** from computation based on experimental data

Methods

- Parametrizable Surrogate Generative Model
- Random-process driven Generative Model
- Generative Adversarial Networks
- Styled Generative Adversarial Networks
- Convolutional Neural Networks
- Transposed Convolutional Neural Networks

Results

- **Interpolation within the trained parameter space is possible**
- **Extrapolation fails**
- **No GAN model provides signals with physical correctness (other authors never tested this)**
- **Styled GANs are promising to get control over the generation process, but parameter vector is entangled (not independent).**

THANK YOU

Data-driven Parameterizable Generative Adversarial Networks for Synthetic Data Augmentation of Guided Ultrasonic Wave Sensor Signals

Stefan Bosse

sbosse@uni-bremen.de

www.edu-9.de

Universität Bremen