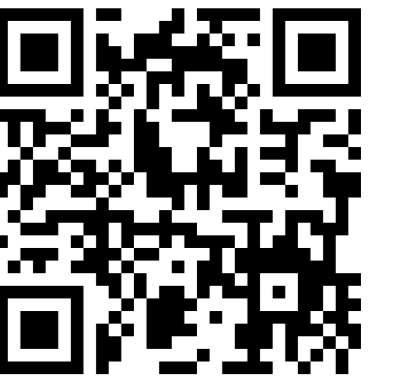


Summary

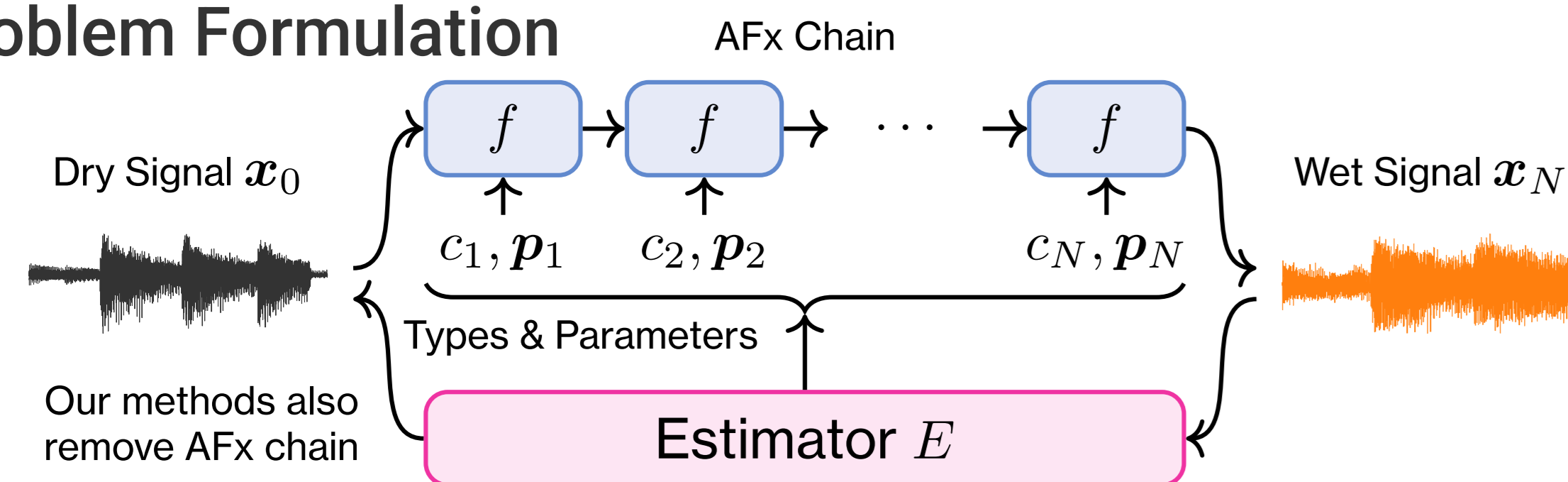
- A new approach for audio effect (AFx) chain estimation: (1) DNN-based prediction of the dry signal and all or part of the AFx configuration, (2) Search based on wet signal reconstruction
- AFx removal (dry signal estimation) in prediction stage enables complementing or improving the predictions via reconstruction similarity as an objective function
- Methods based on the proposed approach outperformed the method solely based on the prediction-based approach
- Best task division in our experiment: (1) prediction of type combination, (2) search-based order & parameter estimation



Demo

1. Introduction

Problem Formulation



- AFx chain application: $x_N = f_{c_N, p_N} \circ \dots \circ f_{c_1, p_1}(x_0) = F_{C, P}(x_0)$
 - AFx chain: $F_{C, P}$, Types: $C = (c_1, \dots, c_N)$, Param.: $P = (p_1, \dots, p_N)$
- AFx chain estimation: $(\hat{C}, \hat{P}, \hat{x}_0) = E(x_N)$
 - Estimate the configuration (types and parameters) of applied AFx and the dry signal from a wet signal

Background

- AFx play an essential role in the sound design of speech and music
- Achieving a desired sound design using diverse and complex AFx has required expertise in both technical and artistic domains
- ▶ Automatic estimation enables efficient learning and reuse of the sound design techniques from existing audio content

2. Related Works

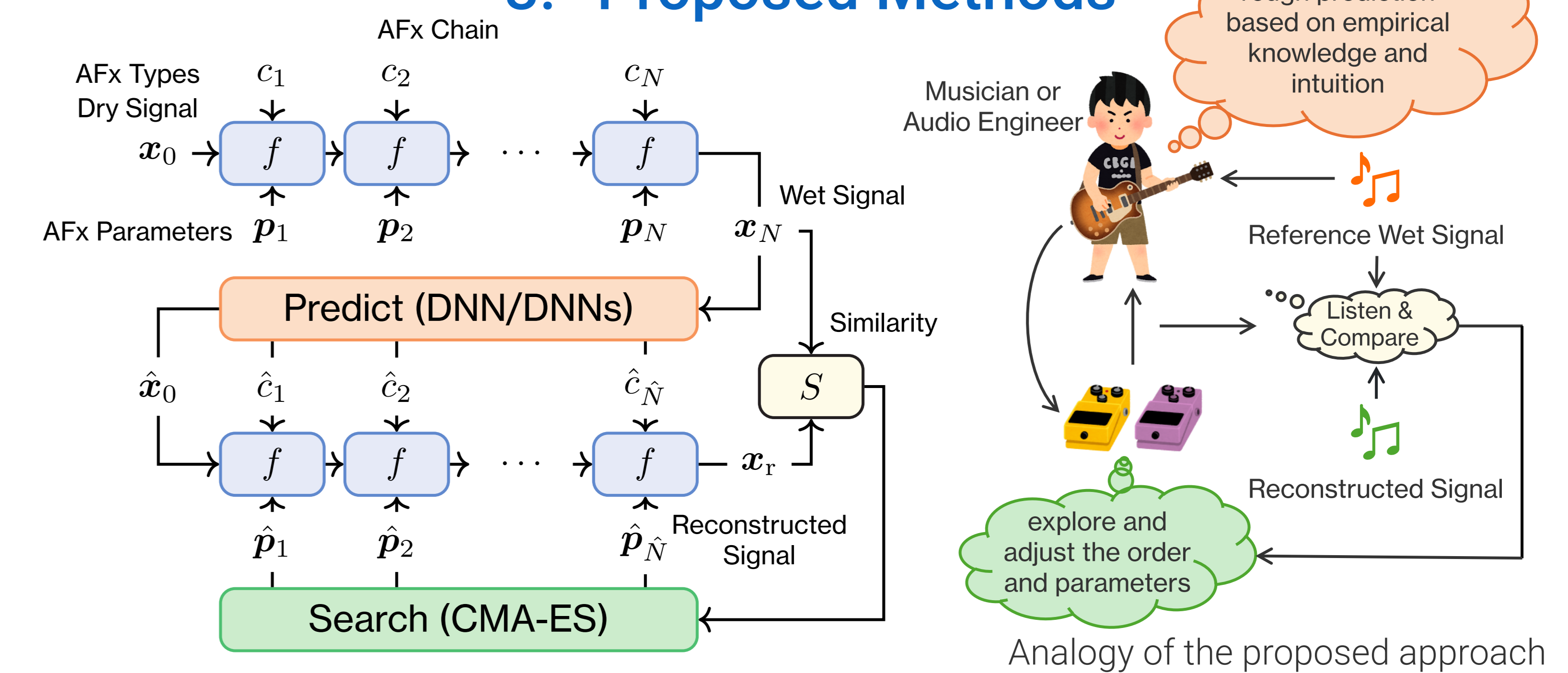
Prediction-Based Approaches

- Pre-trained models (e.g. DNNs) predict the AFx config. from an unseen wet signal
- These models are trained with the error against the data (dry signals, wet signals, AFx config.)
- Employed for estimation [Hinrichs+2022, Wada+2025], removal [Rice+2023], joint estimation of dry signal and config. [Take+2024]
- ▶ Few methods can estimate the full config. of chains, including their order and parameters

Search-Based Approaches

- Reconstructed signal is generated by applying AFx to the dry signal
- AFx config. is optimized via reconstruction similarity (the sim. between the reference wet signal and a reconstructed wet signal)
- Employed for style transfer [Steinmetz+2024]
- ▶ Dry signal required for estimation

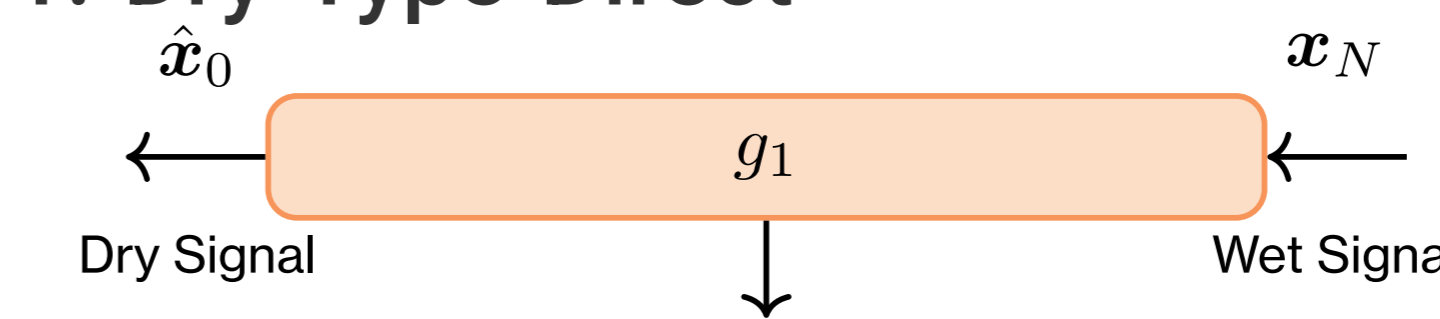
3. Proposed Methods



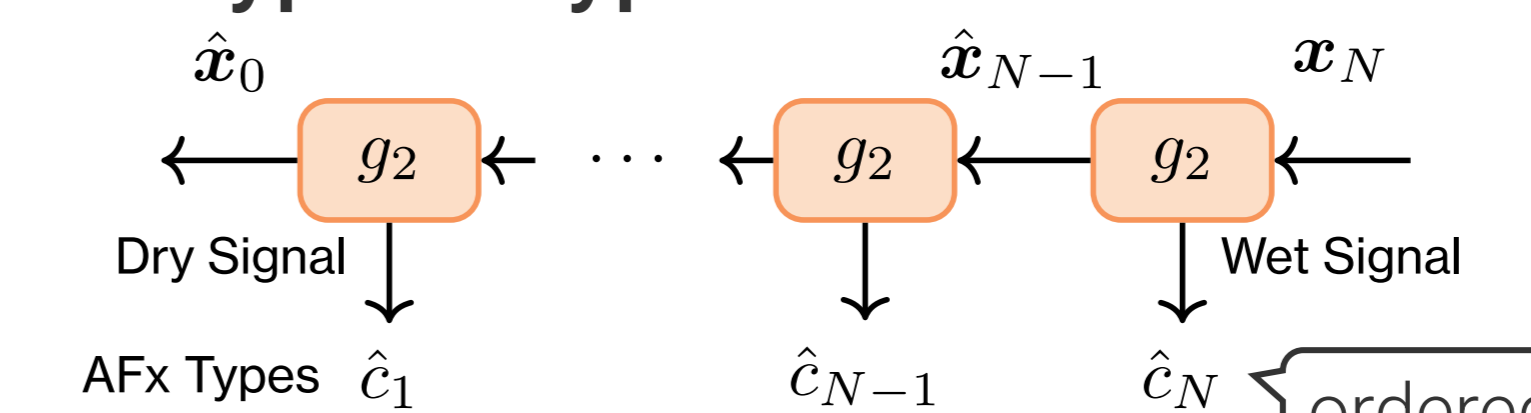
- A new approach that integrates prediction- and search-based approaches
- The division of tasks between prediction and search stage is an essential design choice in the two-stage approach
- ▶ Three settings of task division for comparison

Prediction with Deep Neural Networks

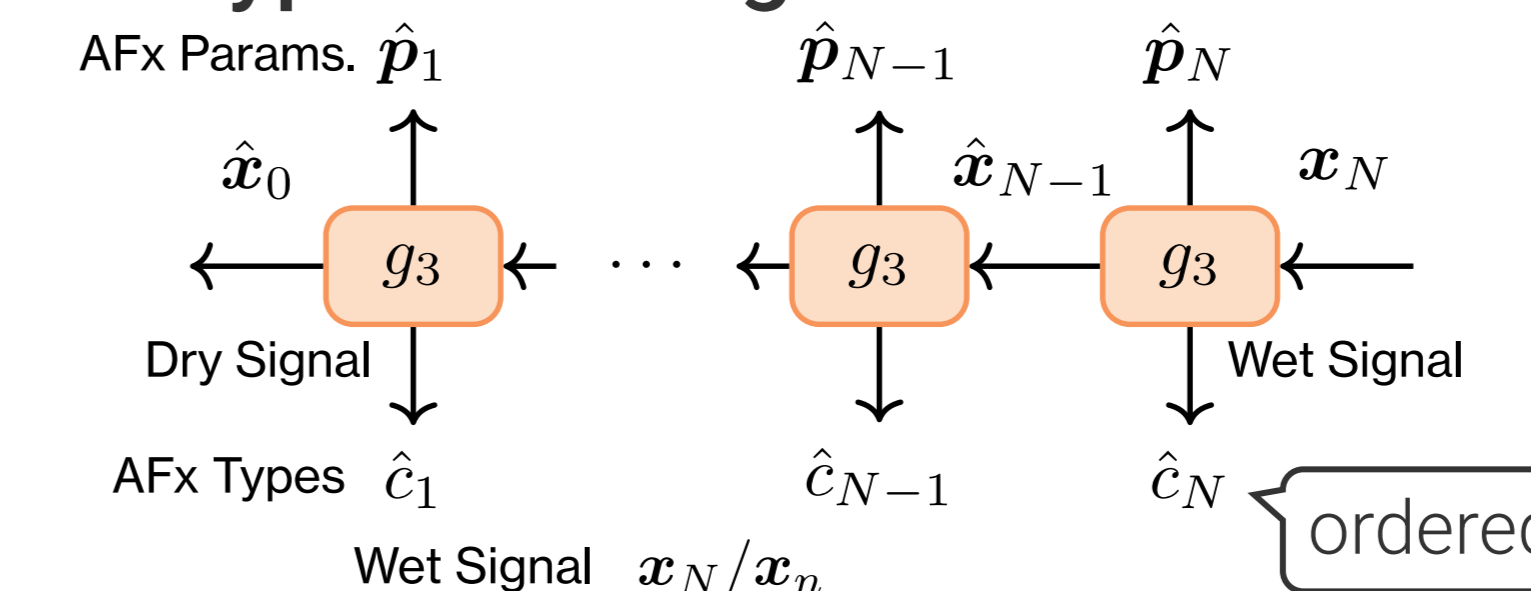
1. Dry-Type-Direct



2. Bypass-Type-Iter



3. Bypass-Config-Iter



Search Based on Wet Signal Reconstruction

$$\hat{P} = \arg \max_P S(F_{\hat{C}, P}(\hat{x}_0), x_N)$$

Reconstructed Signal
S: Similarity (SI-SDR)

- Optimizing AFx parameters to maximize reconstr. similarity between the reconstr. signal and the wet signal
- Black-box optimization (CMA-ES) for non-differentiable AFx chain
- Two-stage search in Dry-Type-Direct
 1. Ordering via exhaustive search
 2. Refinement of Parameters

4. Experimental Evaluation

Dataset

- Dry signals: musical excerpts played on the guitar from existing datasets, 10.0s x 2231 chunks
- Wet signals: Applying pedalboard AFx to dry signals, 15 possible chains using 3 types, randomized parameters, resulting in 205h

Evaluation Results

Methods	AFx Type Chain Classification			AFx Chain Removal		
	Macro F1	LD	EMA	Methods	SI-SDR	MR-STFT
Dry-Type-Direct + Search	0.958	0.313	0.774	Dry-Type-Direct	13.96	0.813
Bypass-Type-Iter	0.949	0.369	0.723	Bypass-Type-Iter	14.95	0.898
Bypass-Config-Iter	0.942	0.408	0.702	Bypass-Config-Iter	14.88	0.902

- Dry-Type-Direct+Search achieves best performance in most metrics, likely by avoiding iterative error accumulation

Types	Variable Parameters
Chorus	depth, feedback, mix
Distortion	drive_db
Reverb	room_size, damping, wet_level

Wet Signal Reconstruction

Methods	SI-SDR	MR-STFT
Bypass-Config-Iter (Baseline)	18.18	0.465
Dry-Type-Direct + Search	23.07	0.340
Bypass-Type-Iter + Search	22.68	0.361
Bypass-Config-Iter + Search	22.64	0.366

- Reconst. w/ ground-truth dry signal (independent of effect removal)

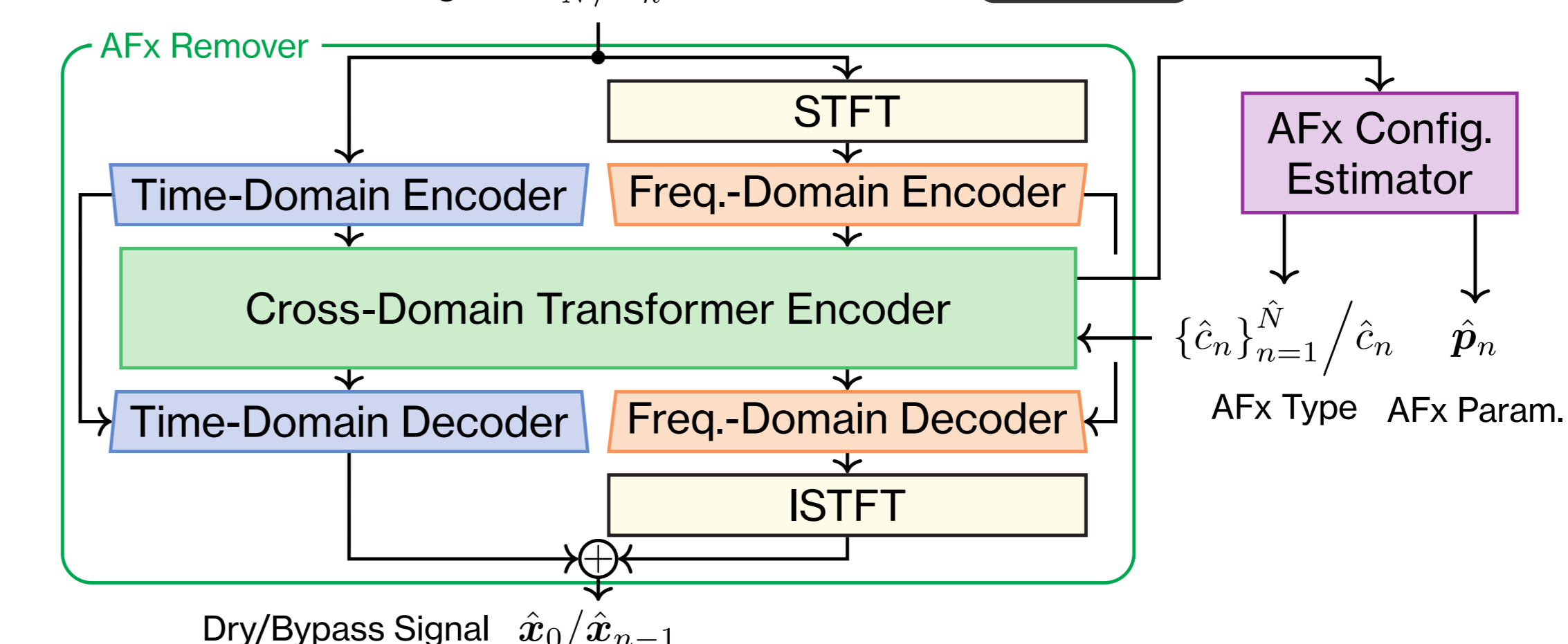
5. Limitations

The Restricted Diversity of AFx

- Fixed chorus rate due to a steep and hard-to-optimize landscape
- Evaluation scope restricted to only three AFx types

Potential Overfitting in Search Stage

- Tuning search scale to avoid overfitting to errors in estimated dry signals and improve accuracy



The architecture of DNNs g_1, g_2, g_3 follows SunAFXiNet [Take+2024], which is an extension of HTDemucs [Rouard+2023]