Hyper Recurrent Neural Network: Condition Mechanisms for Black-box Audio Effect Modeling

Yen-Tung Yeh[1] Wen-Yi Hsiao[2] Yi-Hsuan Yang[1]

[1] National Taiwan University [2] Independent researcher

Introduction

Virtual Analog Modeling

- **Goal**: Emulate the exact behavior of the analog device in the digital world

- Typical methods:

Black-box Grey-box White-box

Circuit knowledge

- We use neural networks for black-box modeling in our work

Modeling Types

- Snapshot modeling versus Full modeling
  - Need to consider the condition or not

- Snapshot modeling

$$y[n] = f(x[n])$$

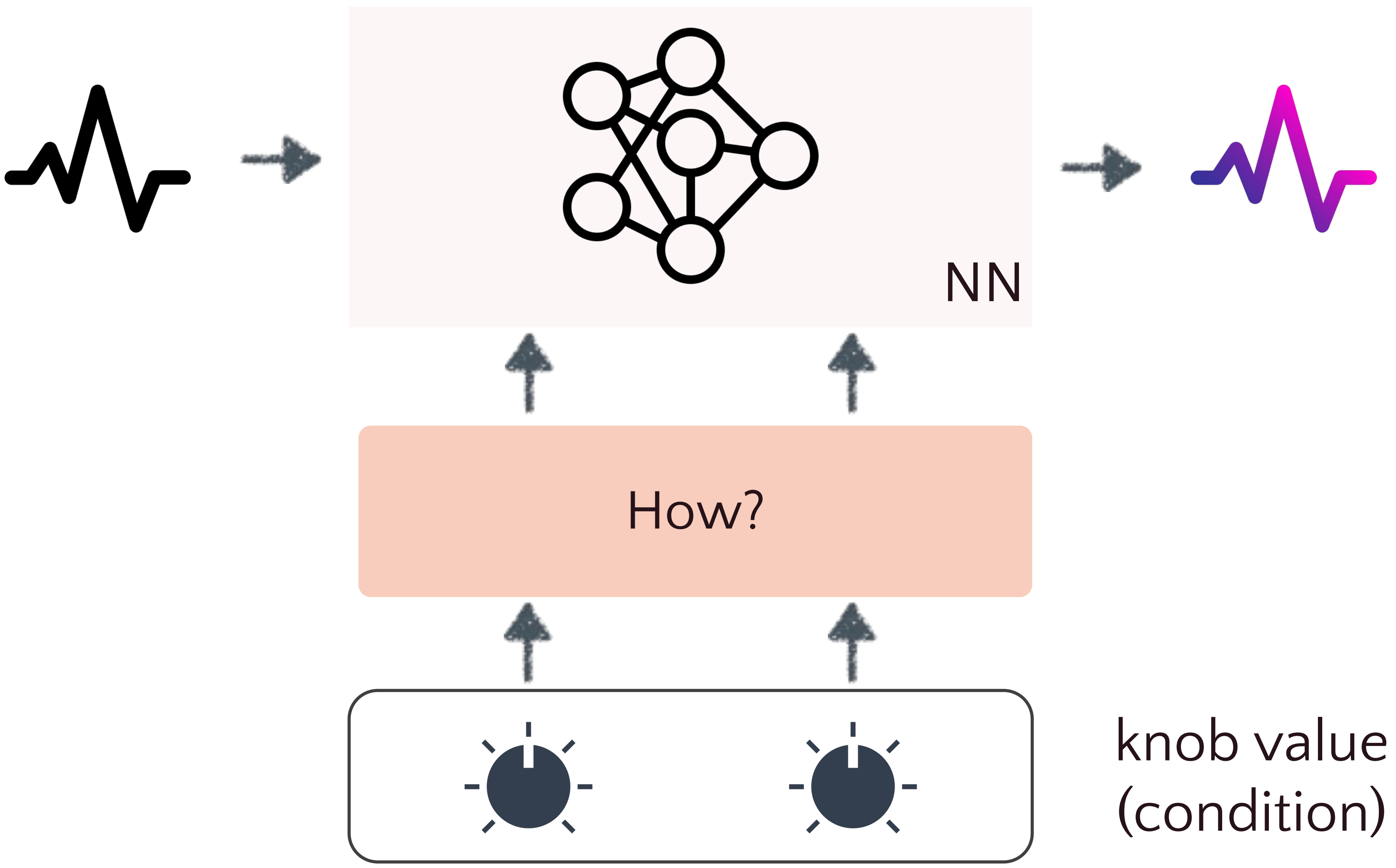
- Full modeling

$$y[n] = f(x[n], c)$$

y[n]: Output signal
x[n]: Input signal
f: Nonlinear function
c: Condition signal

Core Questions

- How to inject the conditioning value to neural networks?



Prior work

CNN-based

- Concatenation
  (Damskägg & Välimäki 2019)
- FiLM
  (Steinmetz & Reiss 2022)

RNN-based

- Concatenation
  (Wright & Välimäki 2019)

- Is there an advanced conditioning method for RNN-based architecture?

Damskägg & Välimäki "Deep learning for tube amplifier emulation" in ICASSP 2019
Steinmetz & Reiss "Efficient neural networks for real-time modeling of analog dynamic range compression" in JAES 2022
Wright & Välimäki "Real-time black-box modelling with recurrent neural networks," in DAFx19.

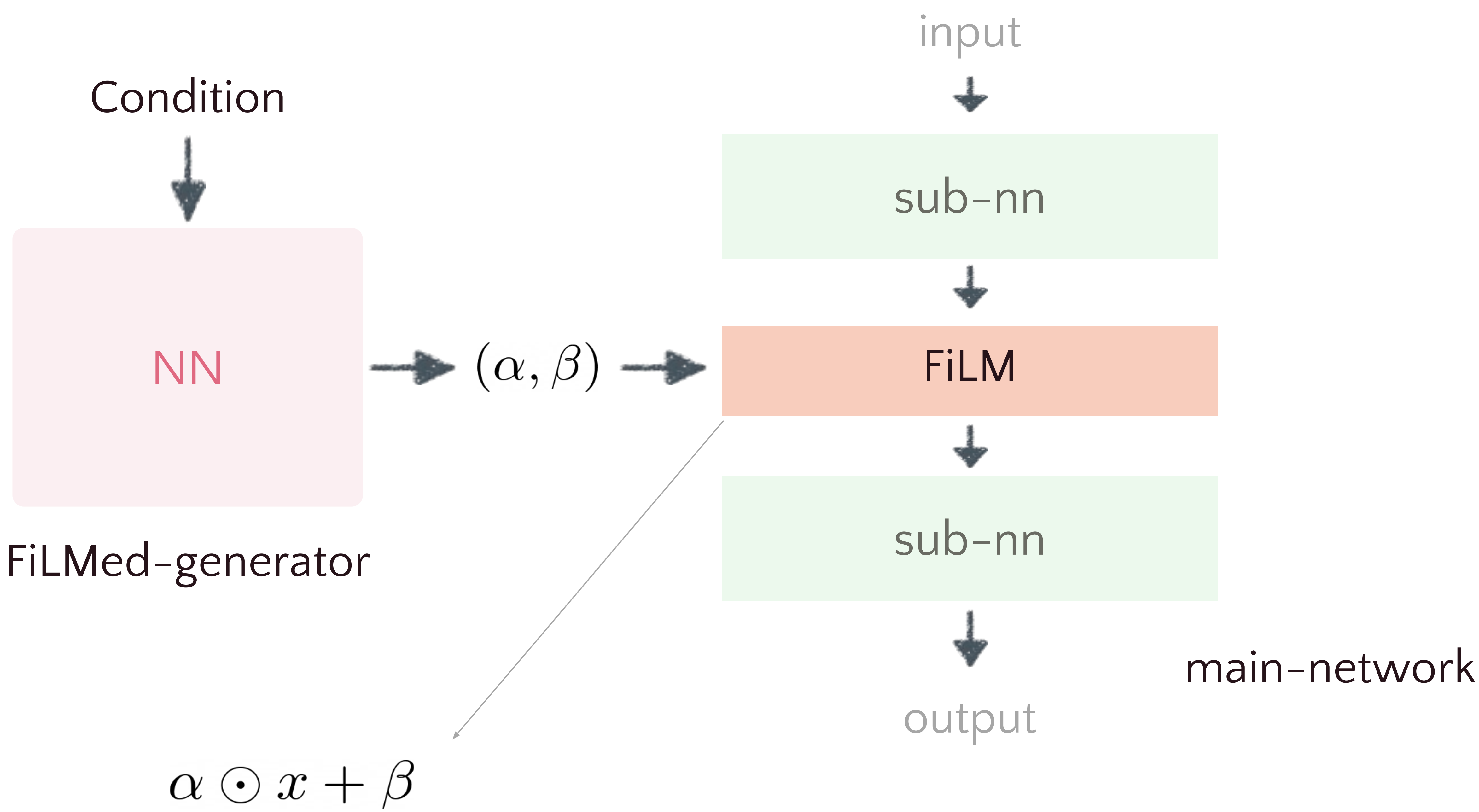
Proposed method

Overview

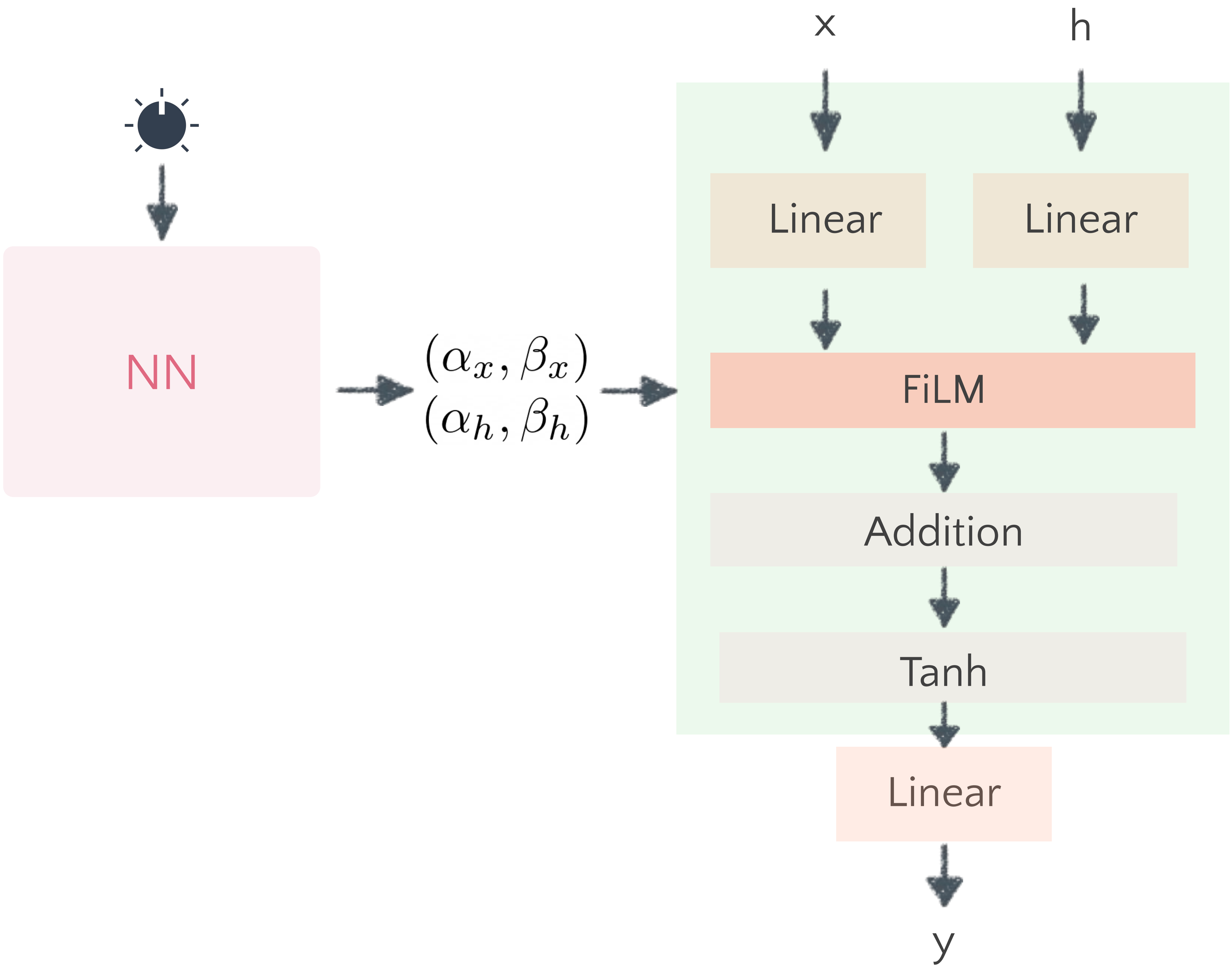
FiLM	StaticHyper	DynamicHyper

FiLM (Feature-wise Linear Modulation)

- Inject conditioning value via element-wise linear modulation



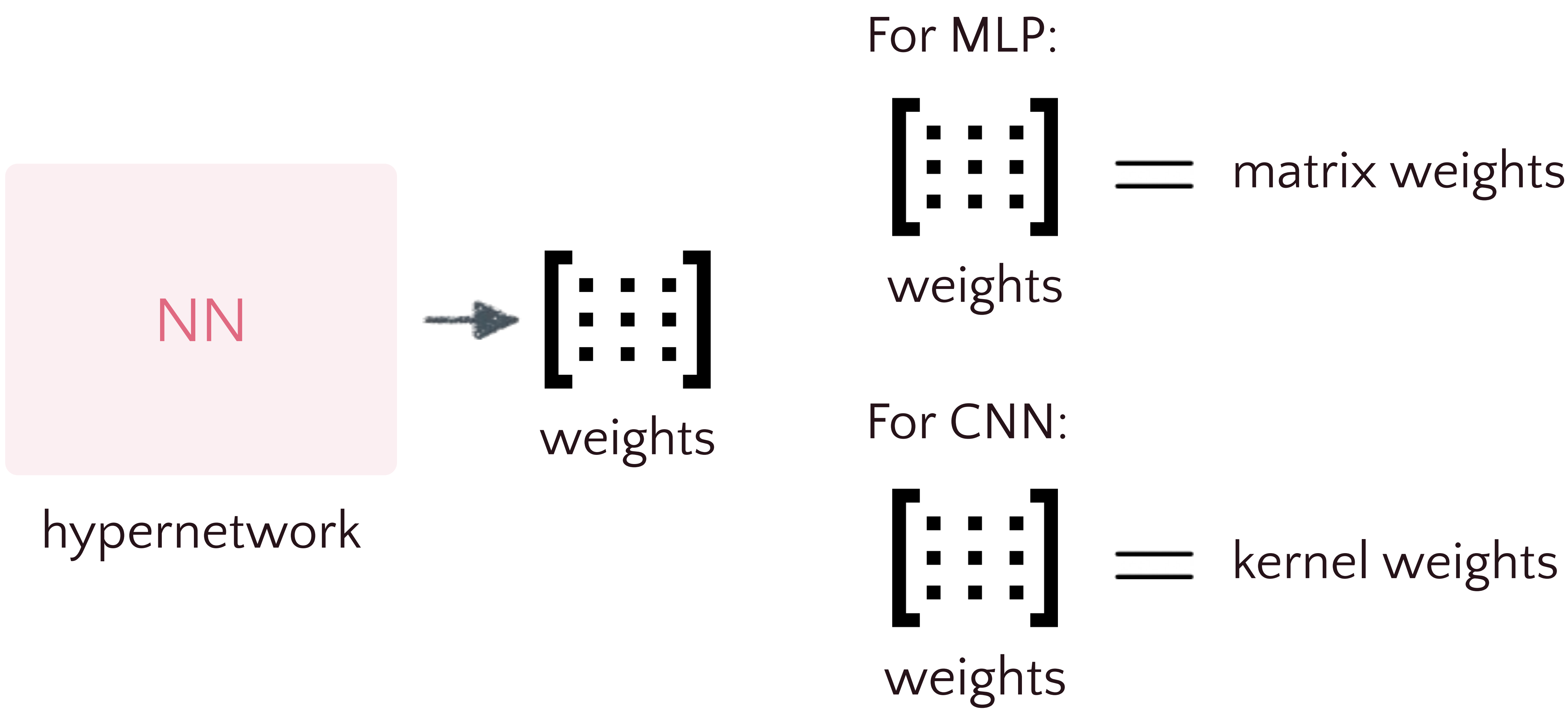
FiLM-RNN



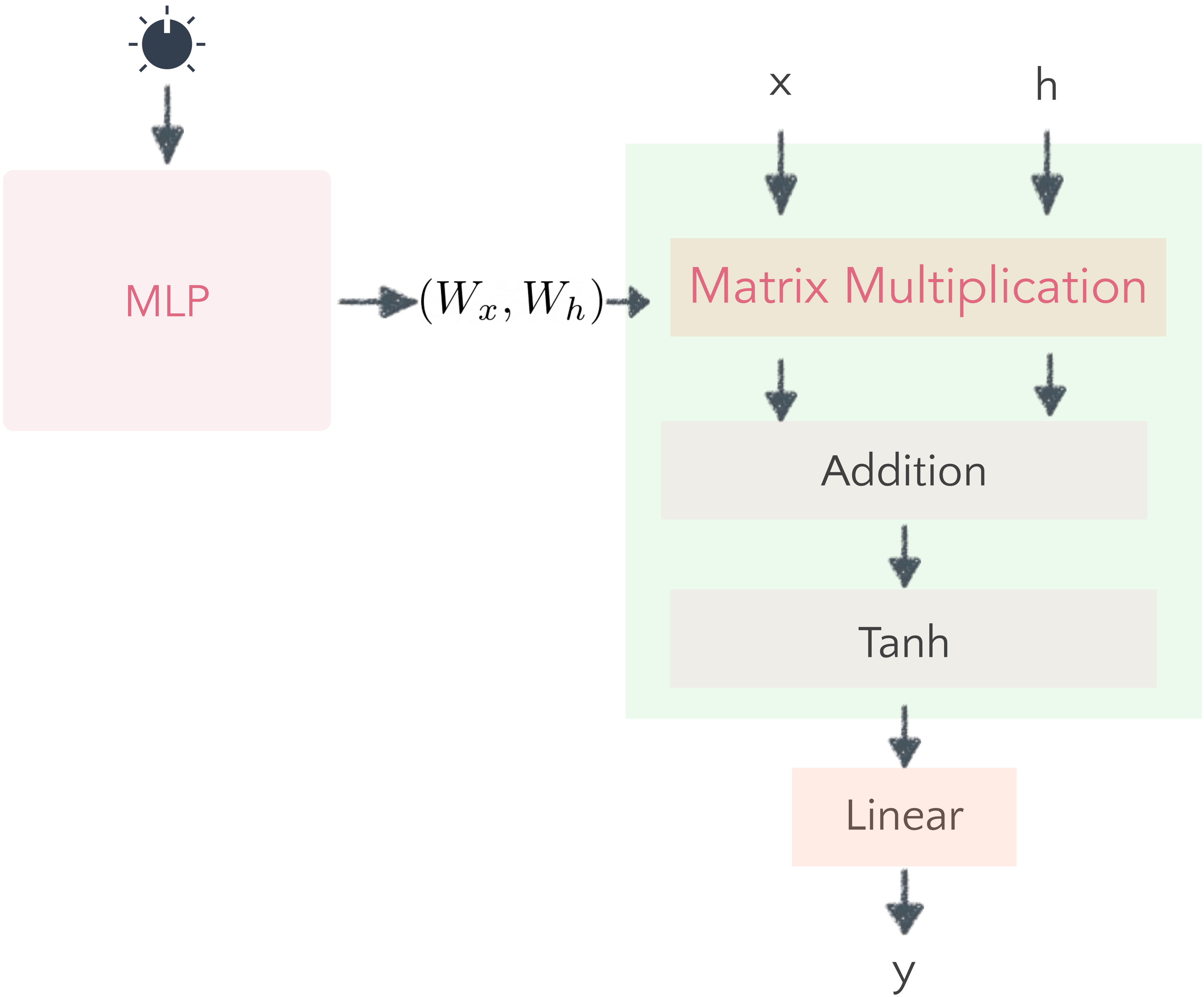
x
h
NN
(α_x, β_x)
(α_h, β_h)
Linear
Linear
FiLM
Addition
Tanh
Linear
y

HyperNetwork

- A network generates weights of another neural network

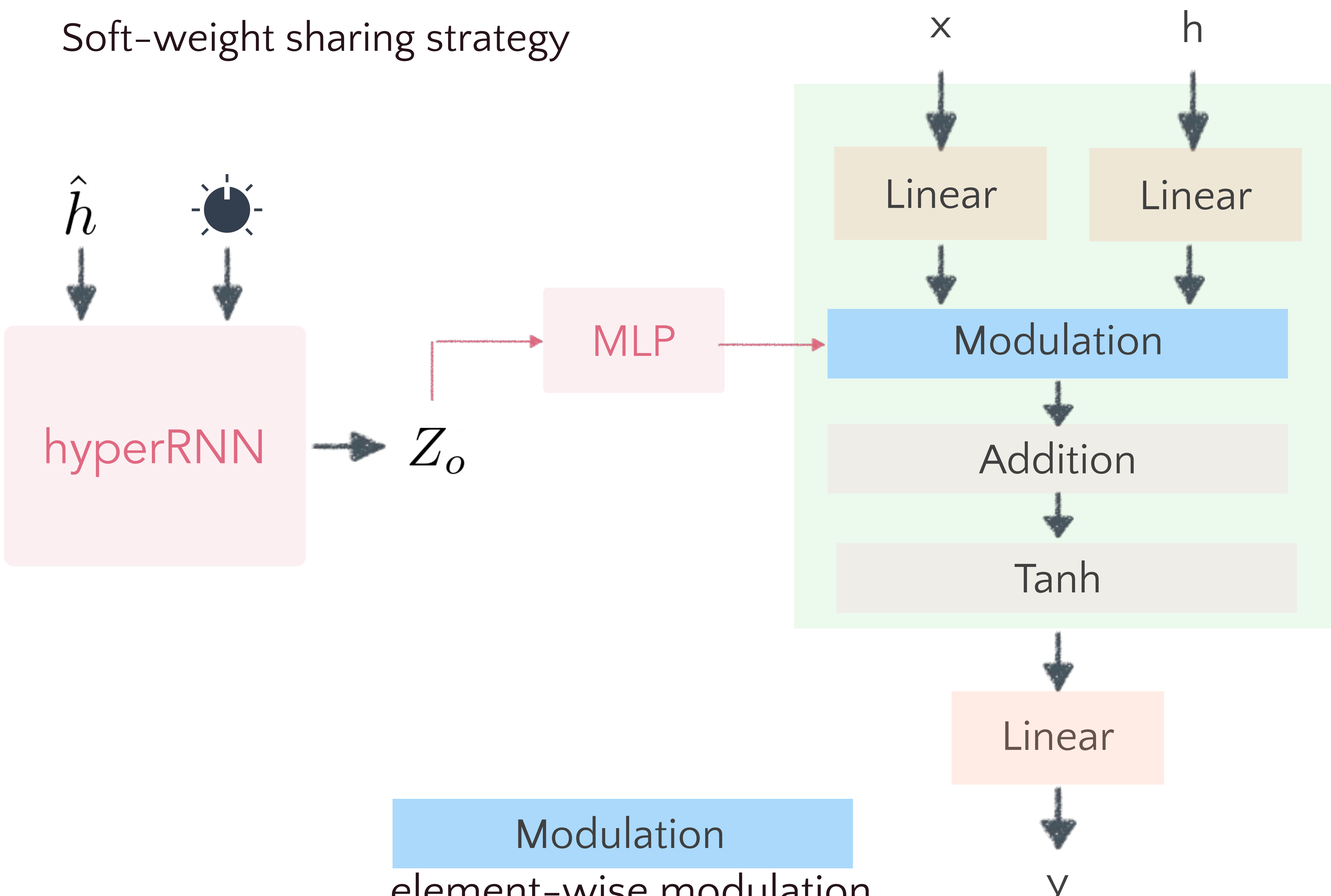


StaticHyper-RNN



DynamicHyper-RNN

Soft-weight sharing strategy



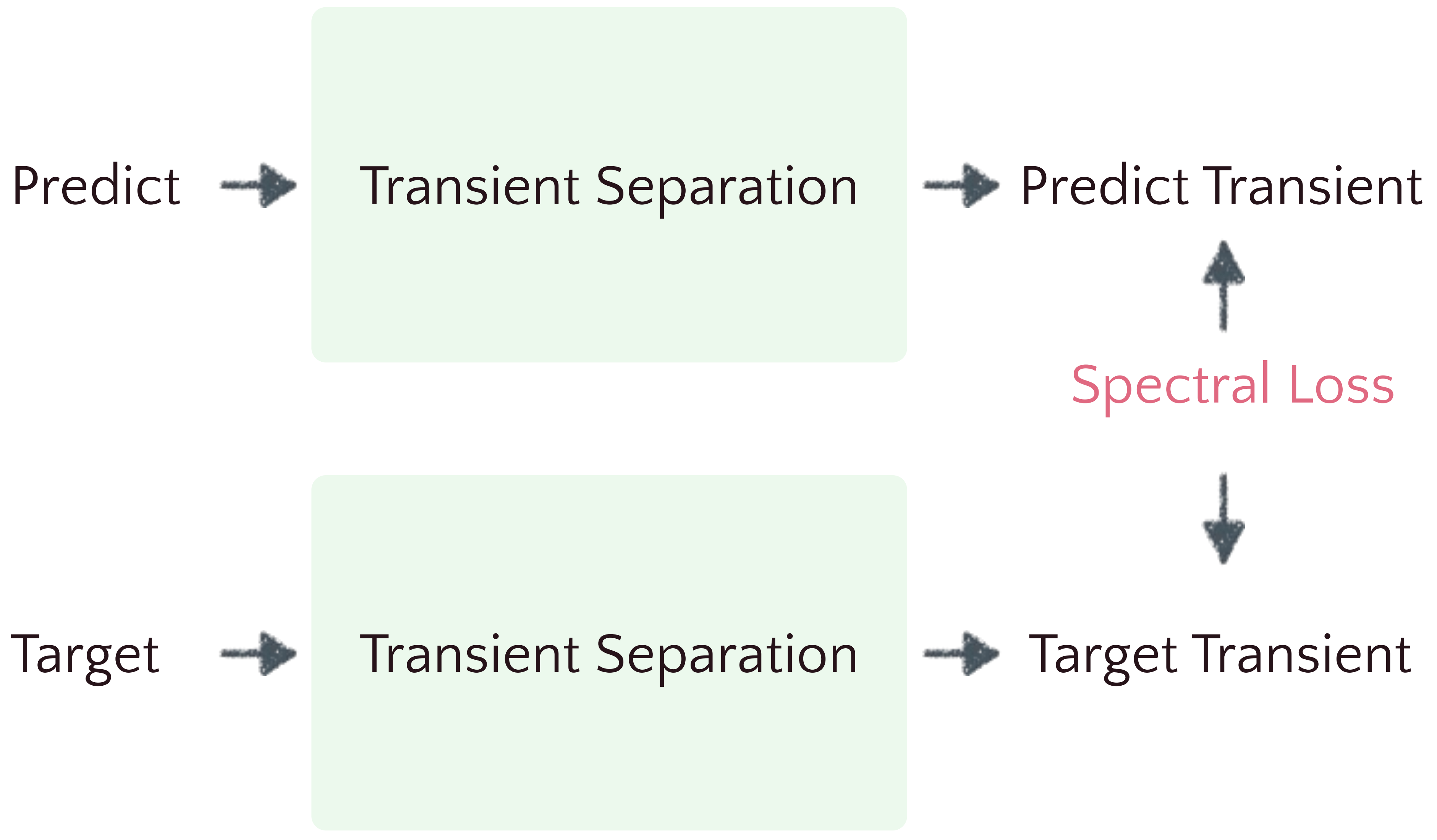
element-wise modulation

Experiments & Results

Implementation details

- Data
  - Boss OD-3 Overdrive pedal (self-collected)
  - LA2A Compressor
- Training
  - Initialize hidden state with zeros
  - BPTT (Backpropagation through time)
    - 2048 samples for Boss OD-3
    - 8192 samples for LA2A Compressor
  - Loss function: MAE Loss + Multiresolution STFT Loss

Transient Metric



Results

Overdrive (Boss OD-3)

Model	Condition	OD-3 (Overdrive)						Params
		L1	STFT	LUFS	CF	RMS	Transient	
GRU	Concat	0.120	1.933	0.455	2.932	1.096	27.338	3585
	FiLM	**0.011***	**0.536**†	0.176	**0.676***	0.401	12.504	17217
	StaticHyper	0.017	0.698	0.165	1.650	**0.318**†	**12.347**†	30369
	DynamicHyper	0.150	**0.428***	**0.075***	0.883	**0.153***	**11.308***	20289

Compressor (LA2A)

Model	Condition	LA2A (Compressor)						Params
		L1	STFT	LUFS	CF	RMS	Transient	
GRU	Concat	0.108	0.507	0.716	2.081	1.640	21.002	3489
	FiLM	**0.011**†	0.597	1.383	**2.006**†	3.081	**15.825***	17185
	StaticHyper	**0.008***	**0.371***	0.543	2.386	1.211	20.437	30361
	DynamicHyper	0.109	**0.377**†	**0.377**†	**1.919***	**0.819**†	**19.826**†	20169

Results-Computation Cost

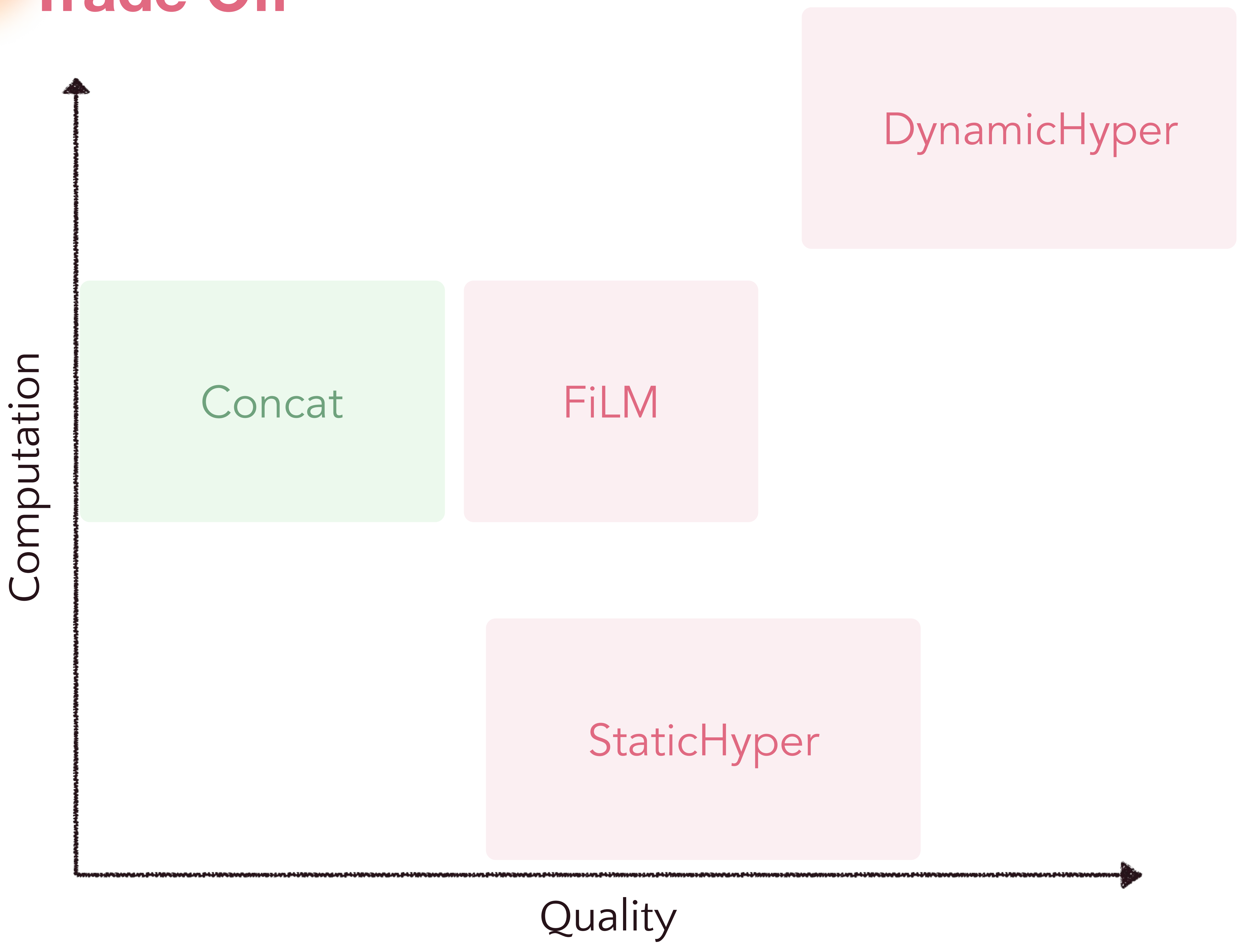
Measuring GFLOPs on one-second audio

Models	GFLOPs
Concat-GRU	0.325
FiLM-GRU	0.307
StaticHyper-GRU	0.003
DynamicHyper-GRU	1.907

When Comparing to Concat:

- FiLM remains similar computation
- StaticHyper save 99% computation
- DynamicHyper cost more computation, but the quality is more better

Trade Off



DynamicHyper
Concat
FiLM
StaticHyper
Computation
Quality

Comprehensive Evaluation

Model	Condition	OD3 (Distortion)						Params
		MAE	STFT	LUFS	CF	RMS	Transient	
LSTM	Concat	0.123	1.901	0.524	2.982	1.259	25.997	4769
	FiLM	0.145	1.057	0.248	1.834	0.552	20.322	22561
								40449
								21857
GRU								3585
								17217
								30369
								20289
TCN	Concat	0.033	0.928	0.305	1.177	0.671	27.634	21769
	FiLM	0.044	0.698	0.338	0.894	0.842	33.678	29849
GCN	Concat	**0.013**†	0.792	0.202	**0.776**†	0.447	19.103	19824
	FiLM	0.149	0.672	**0.141**†	1.200	0.276	24.474	32368

Check out our paper for more results

Conclusion

- Novel condition mechanism for RNN-based architecture
  - Improve emulation quality
  - Save computation cost

More audio samples:
https://yytung.notion.site/HyperRNN