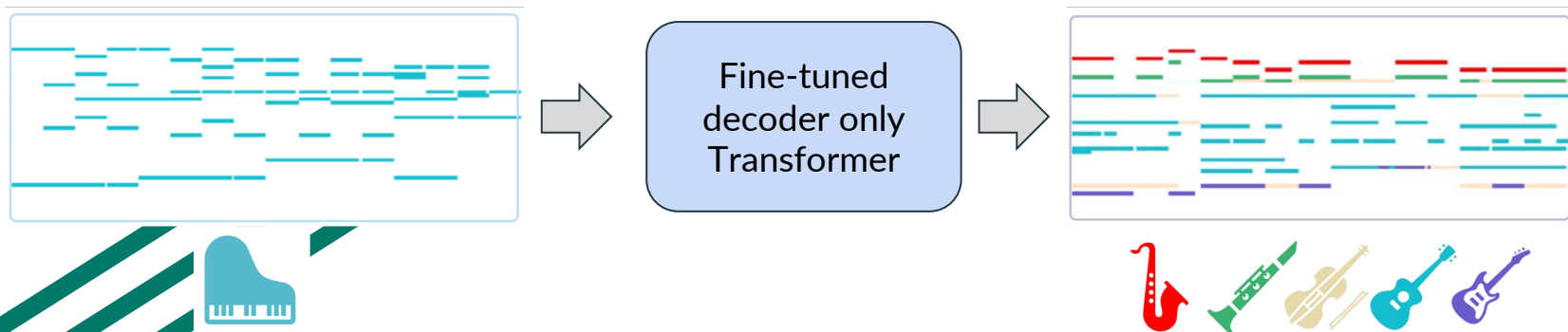


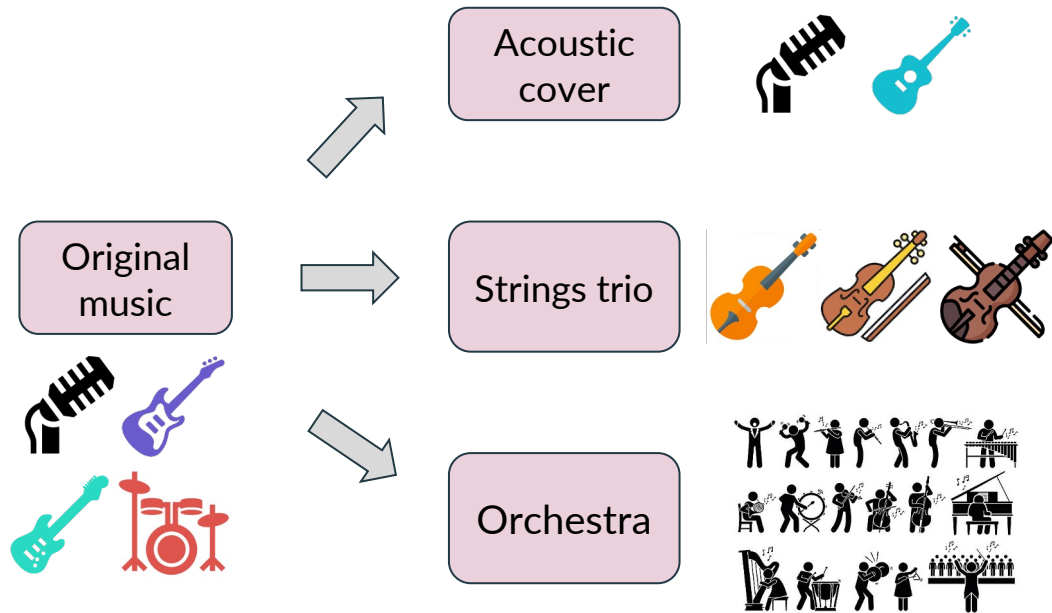
# Unifying Symbolic Music Arrangement: Track-Aware Reconstruction and Structured Tokenization

Longshen Ou, Jingwei Zhao, Ziyu Wang, Gus Xia, Qihao Liang, Torin Hopkins, and Ye Wang

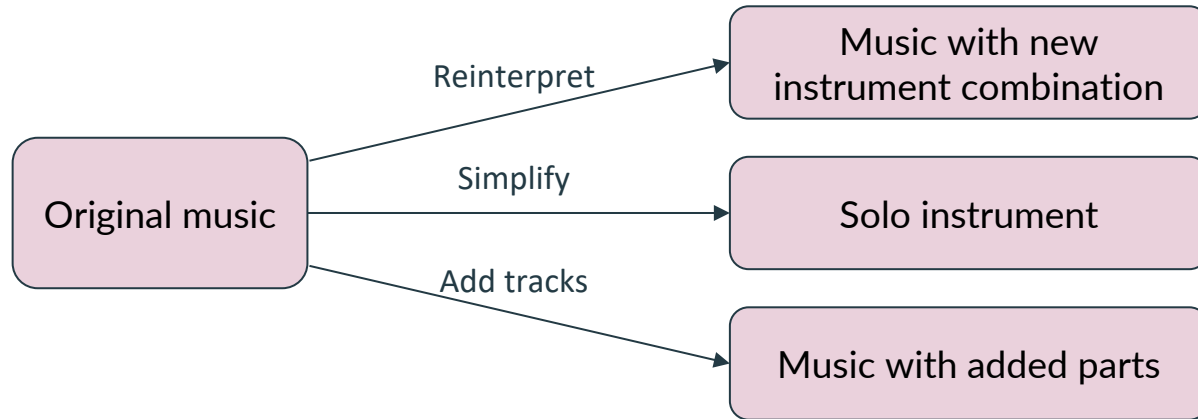


# What is Music Arrangement?

- Transforms the same musical idea into new instrumental forms
- Preserving its essence while changing timbre and interpretation

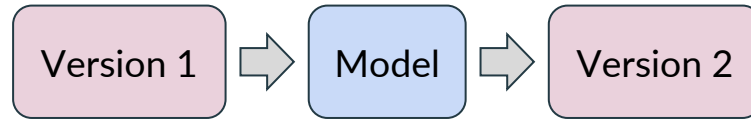


# Problem Definition



- Generate new tracks from existing tracks
- That **reinterpret** the original music or **remain compatible with** original music

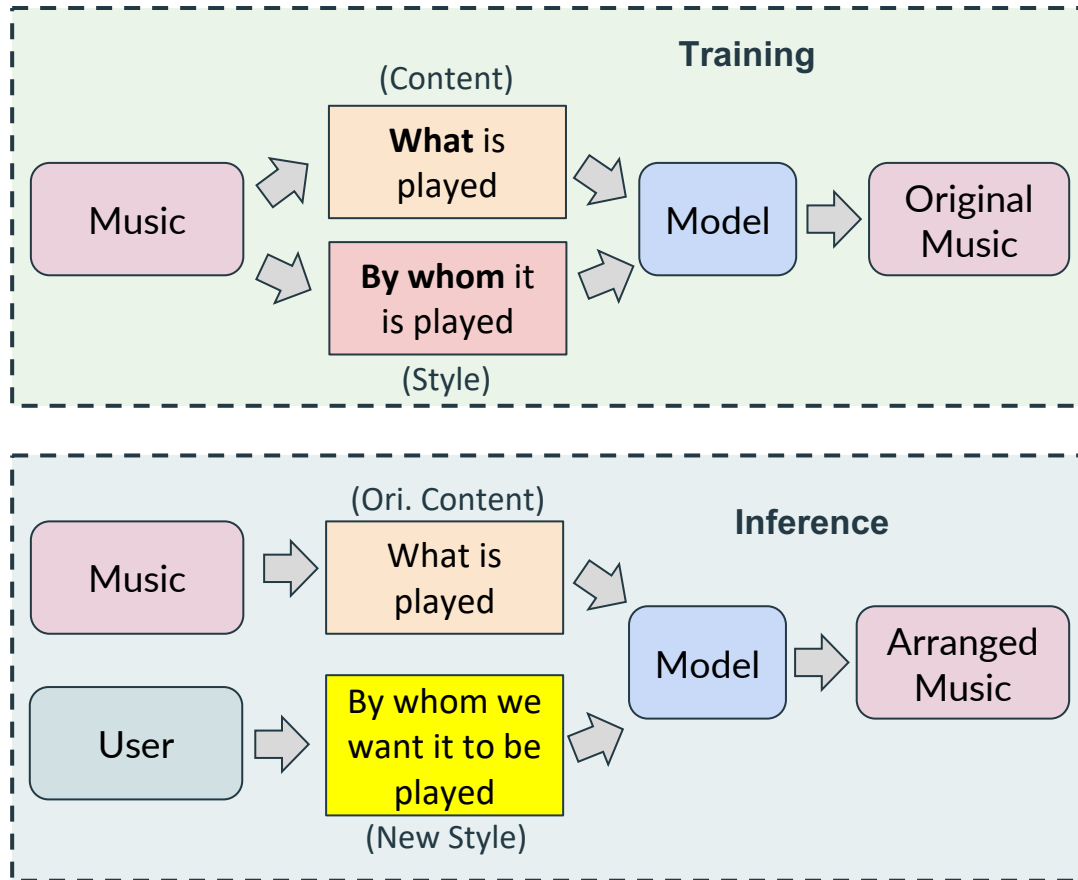
# Methodology



- Paired arrangements data are extremely rare

# Methodology

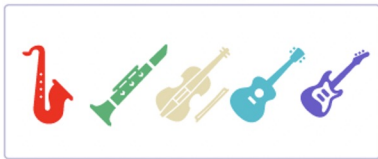
- Learn reconstruction instead



# Unsupervised Reconstruction Objective

*Condition Sequence*

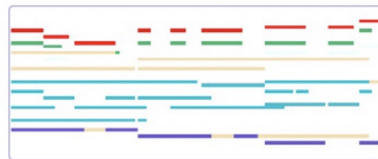
<INSTRUMENT>



<CONTENT>



<HISTORY>



<SEP>



*Target Sequence*

- Condition contain 3 subsequences:  
Instrument, flattened note stream, target-side history
- Fine-tuned from pre-trained decoder only model

# REMI-z: Alleviate Content Fragmentation

**o-X**: onset time, **i-X**: instrument, **p-X**: pitch, **d-X**: duration, **b-1**: end of bar

Contents of same instruments are in same color

REMI+

**o-0** i-26 p-60 d-26 **o-0** i-33 p-36 d-23 **o-0** i-29 p-36 d-10 **o-12** i-29 p-36 d-12  
**o-18** i-80 p-74 d-14 **o-18** i-29 p-48 d-12 **o-24** i-29 p-36 d-8 **o-30** i-29 p-52  
d-11 **o-36** i-80 p-76 d-11 **o-36** i-29 p-36 d-10 **o-42** i-29 p-52 d-7 b-1

Time-ordered tokenization interleave content of different instruments

REMI-z (Ours)

**i-80** o-18 p-74 d-14 o-36 p-76 d-11 **i-26** o-0 p-60 d-26 **i-29** o-0 p-36 d-10  
o-12 p-36 d-12 o-18 p-48 d-12 o-24 p-36 d-8 o-30 p-52 d-11 o-36 p-36 d-10  
o-42 p-52 d-7 **i-33** o-0 p-36 d-23 b-1

Preserve track continuity to easier learning of idiomatic instrument behavior

# Tasks





# Key Findings

- Our model performs consistently and significantly better than task-specific baselines, both in objective metrics and in human evaluations.
- Our tokenization scheme improves every evaluated aspect of arrangement performance.
- Pre-training plays a crucial role.

# Demos

Example: Band → String Trio

Original



Arrangement  
for string trio



# Takeaways

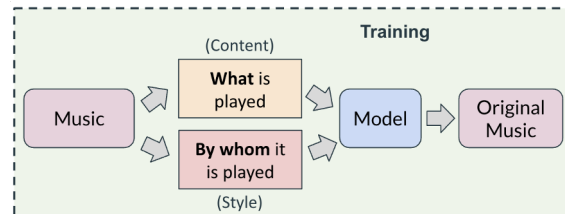
## Unified pipeline

No task-specific model design



## Unsupervised training

Learn interpretation by reconstruction



## Effective tokenization

Track continuity facilitate learning music structure

o-X: onset time, i-X: instrument, p-X: pitch, d-X: duration, b-1: end of bar  
Contents of same instruments are in same color

REMI+  
o-0 i-26 p-60 d-26 o-0 i-33 p-36 d-23 o-0 i-29 p-36 d-10 o-12 i-29 p-36 d-12  
o-18 i-80 p-74 d-14 o-18 i-29 p-48 d-12 o-24 i-29 p-36 d-8 o-30 i-29 p-52  
d-11 o-36 i-80 p-76 d-11 o-36 i-29 p-36 d-10 o-42 i-29 p-52 d-7 b-1

Time-ordered tokenization interleave content of different instruments

REMI-z (Ours)  
i-80 o-18 p-74 d-14 o-36 p-76 d-11 i-26 o-0 p-60 d-26 i-29 o-0 p-36 d-10  
o-12 p-36 d-12 o-18 p-48 d-12 o-24 p-36 d-8 o-30 p-52 d-11 o-36 p-36 d-10  
o-42 p-52 d-7 i-33 o-0 p-36 d-23 b-1

Preserve track continuity to easier learning of idiomatic instrument behavior

# Bonus

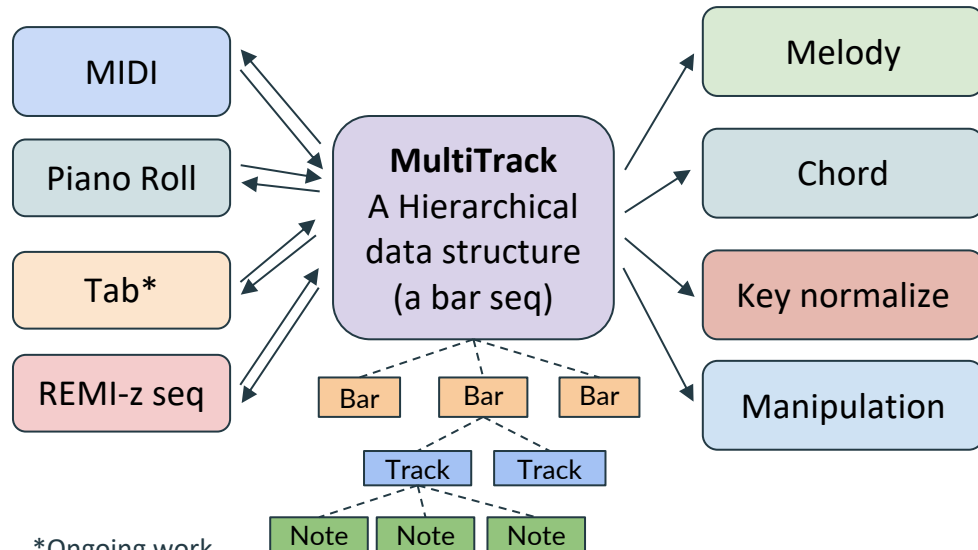
REMI-z simplify note-level modeling  
with lower uncertainty

Arrangement -> General generation tasks

the REMI-z package

```
pip install REMI-z
```

[github.com/Sonata165/REMI-z](https://github.com/Sonata165/REMI-z)



# Thanks

Demo, code, and model  
available online.



[www.oulongshen.xyz/automatic\\_arrangement](http://www.oulongshen.xyz/automatic_arrangement)