

MusicGen音乐生成

# 语音与文本

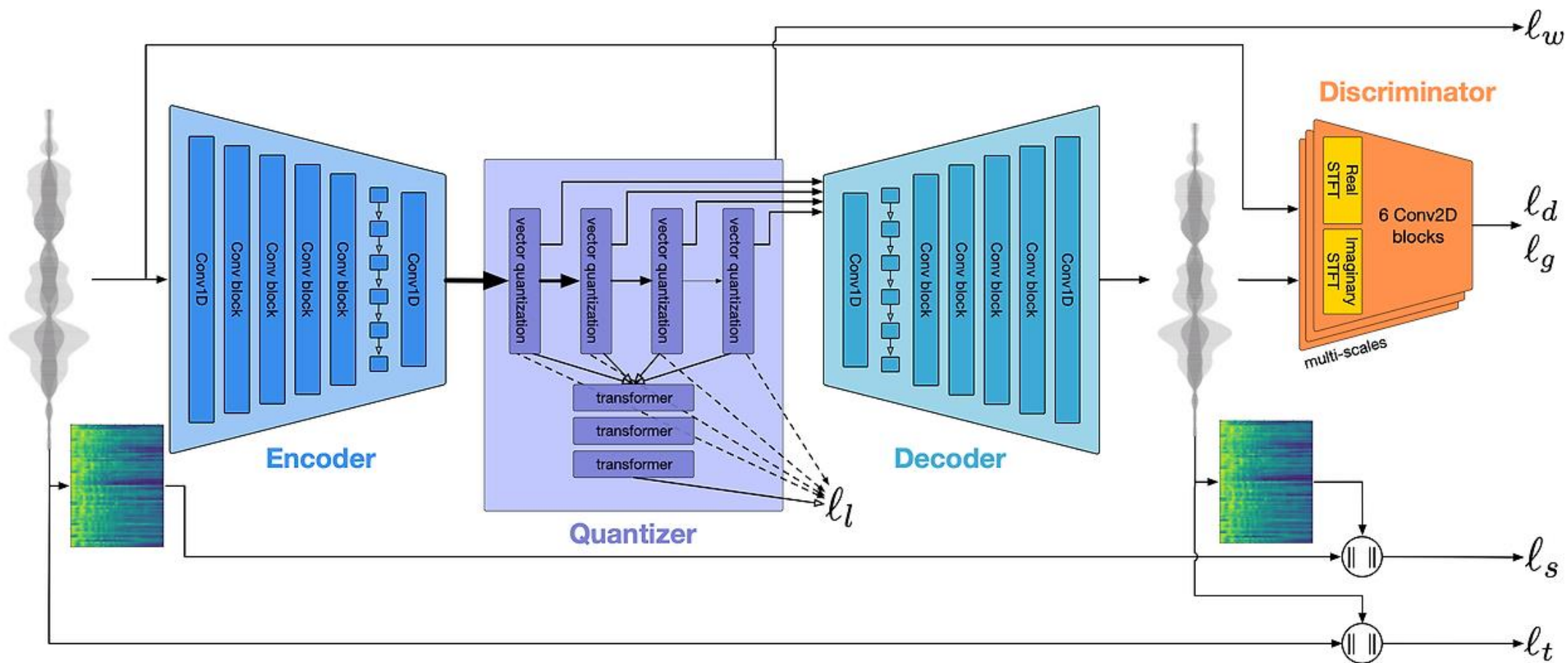
文本模态：

- 表示为离散的token序列；
- 短序列，例如每个句子10-20词；
- 高度抽象，几乎每个词都包含语义信息，因此信息密度较高。

• 语音模态：

- 表示为连续值的序列；
- 极长序列，例如每句话3s，16k采样率，则每句话由48000个连续值的序列表示；
- 信息密度极低，因此有短时不变性，可以从一个片段推测相邻片段的信号。

# Encoder: Encodec



# Vector Quantization

$8 = 2^3$ .  
So at least 3 bits  
needed!

000  
001  
010  
100  
011  
110  
101  
111

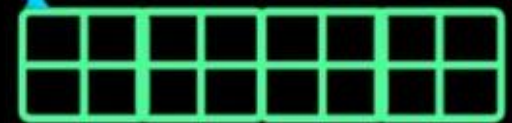
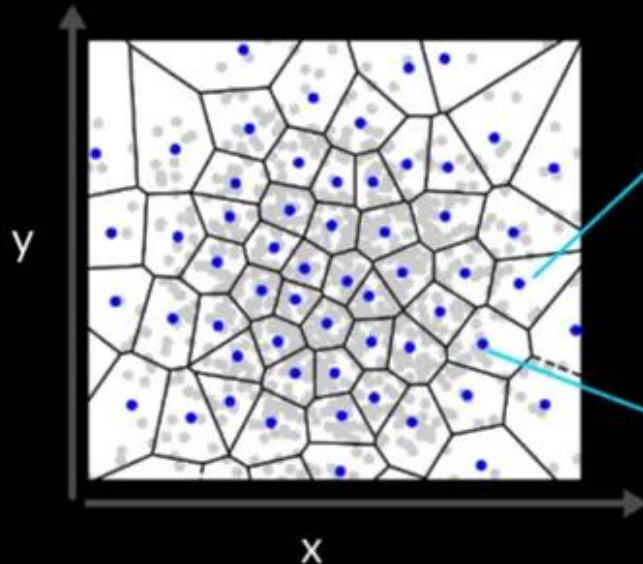
## Vector Quantization

process of converting continuous or discrete data into vectors

### Steps

we cluster the given data and create centroids

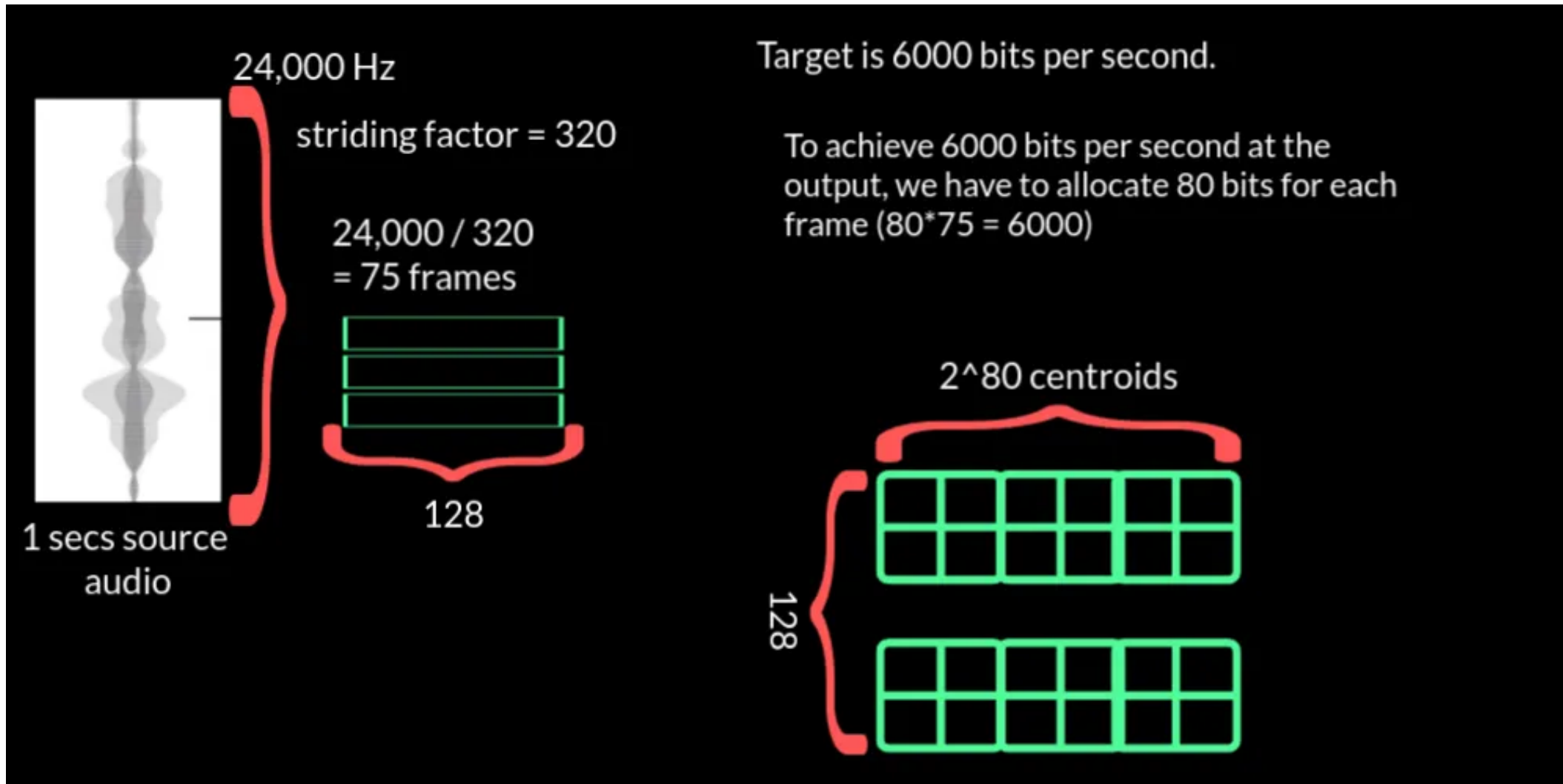
create a codebook with all the centroids put together in a table



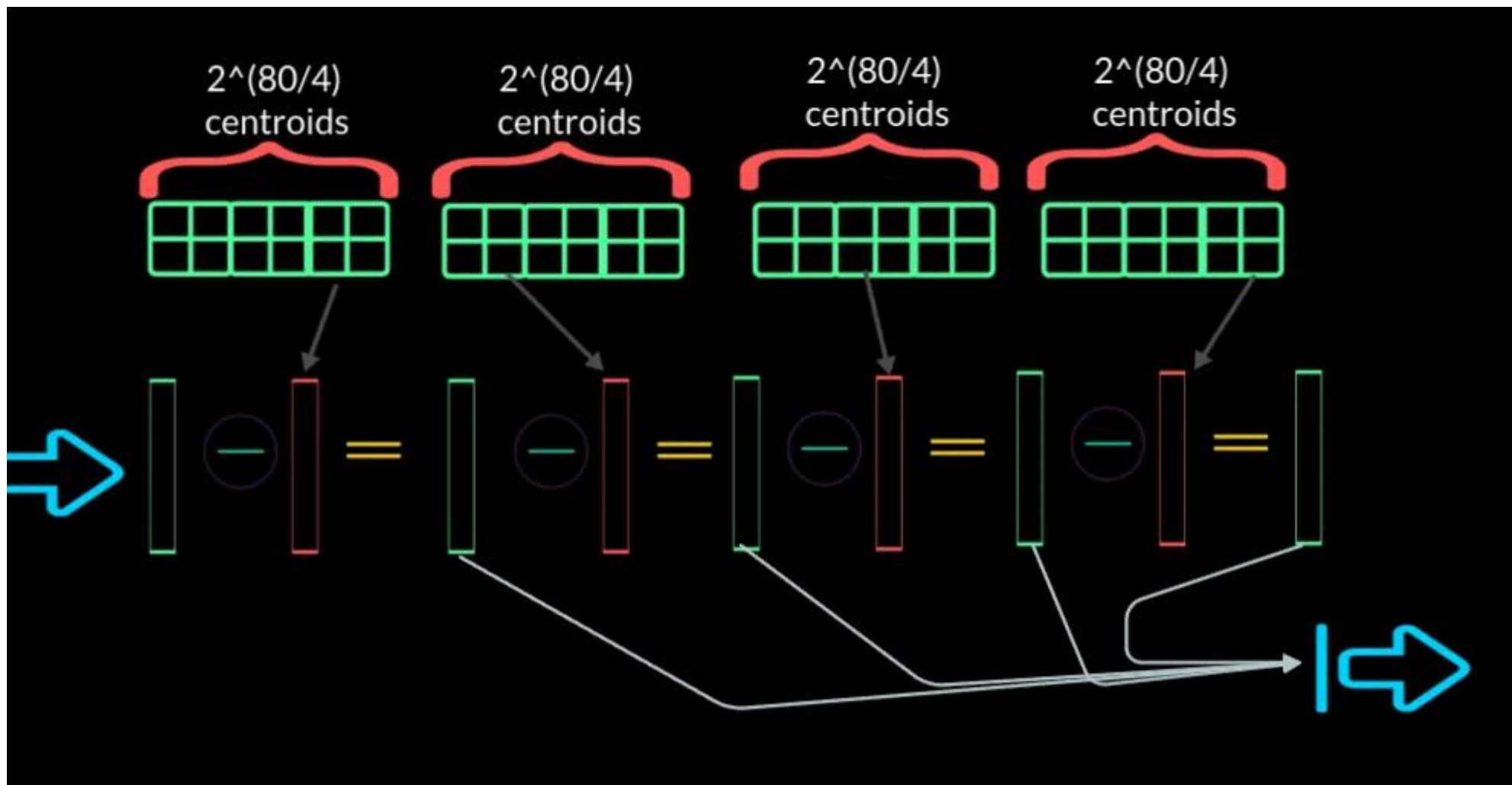
size of codebook  
-> compression  
needed

8 centroids

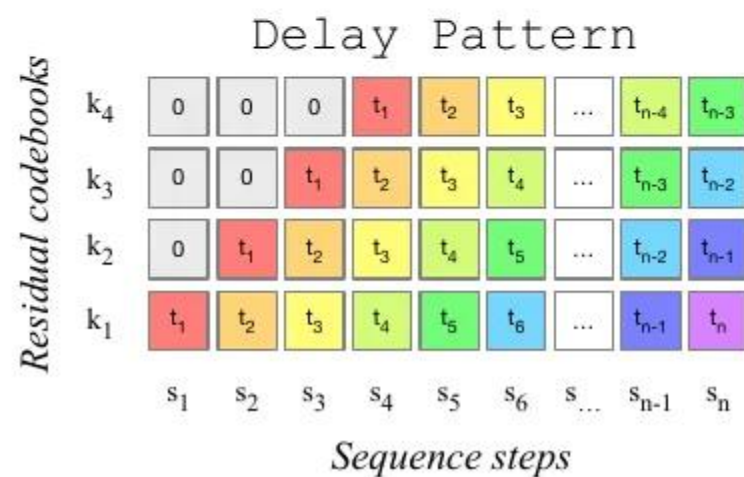
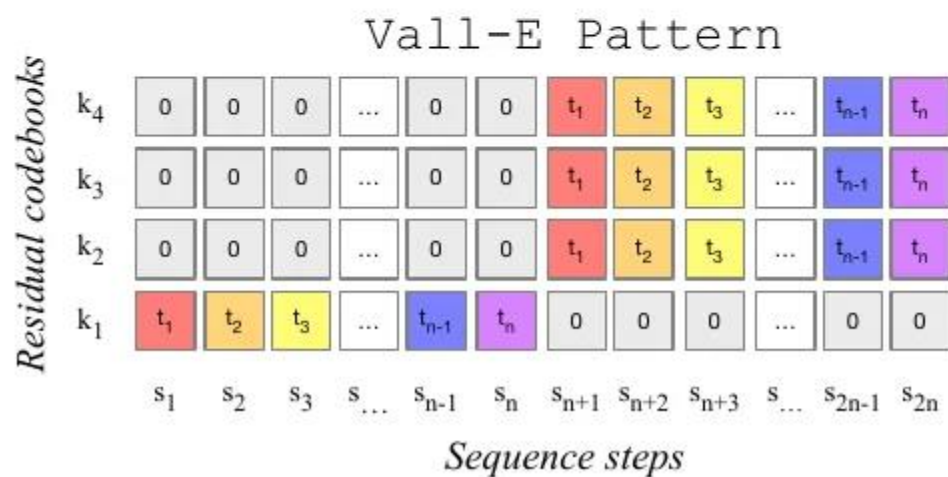
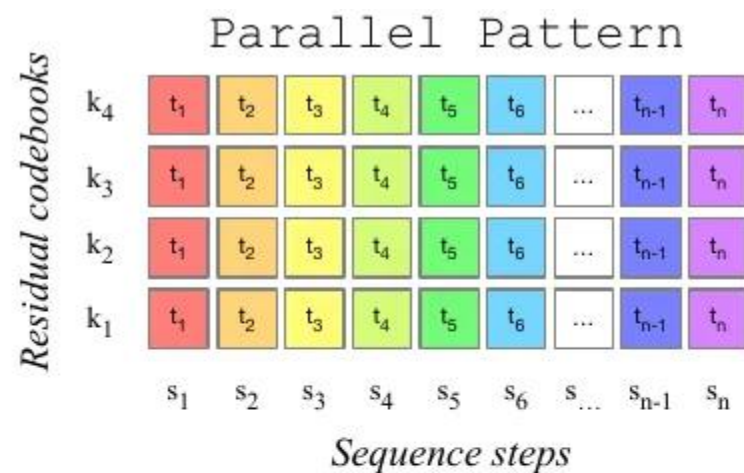
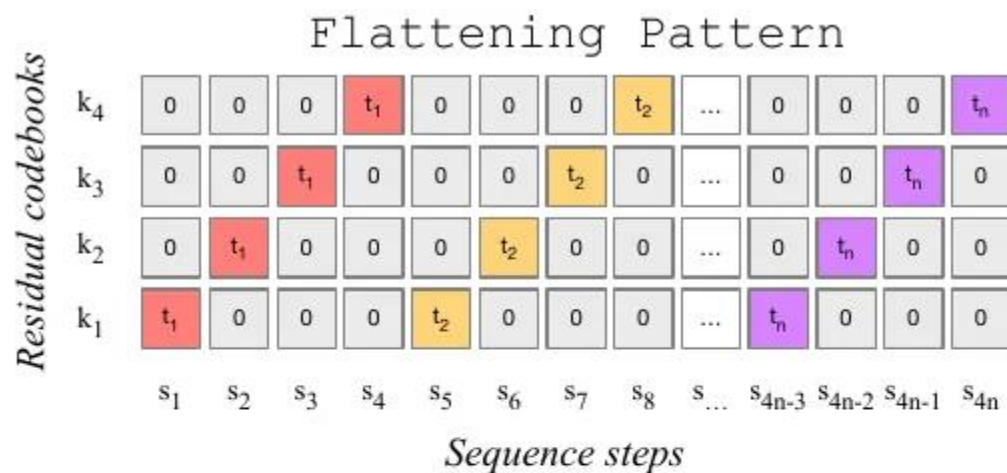
# Limitations of VQ



# Residual Vector Quantization

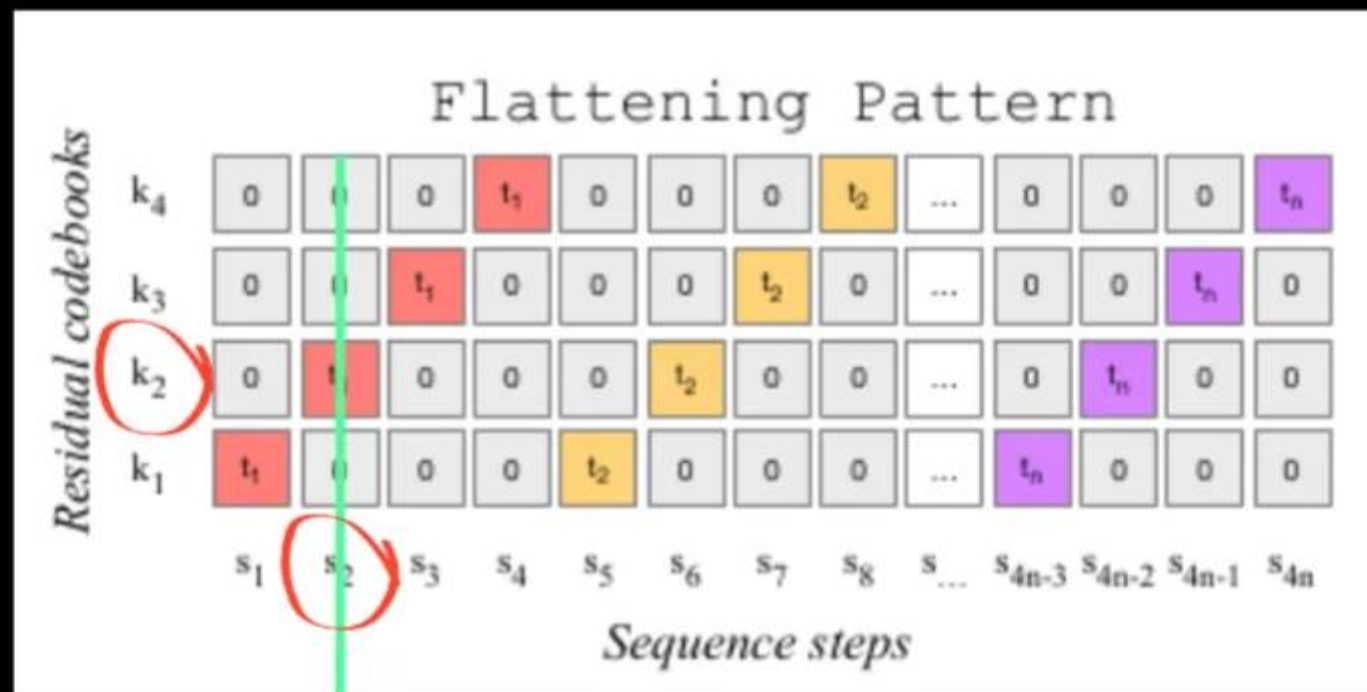


# Interleaving Patterns





# Codebook Projection and Positional Embedding



Sum sinusoidal  
positional embedding

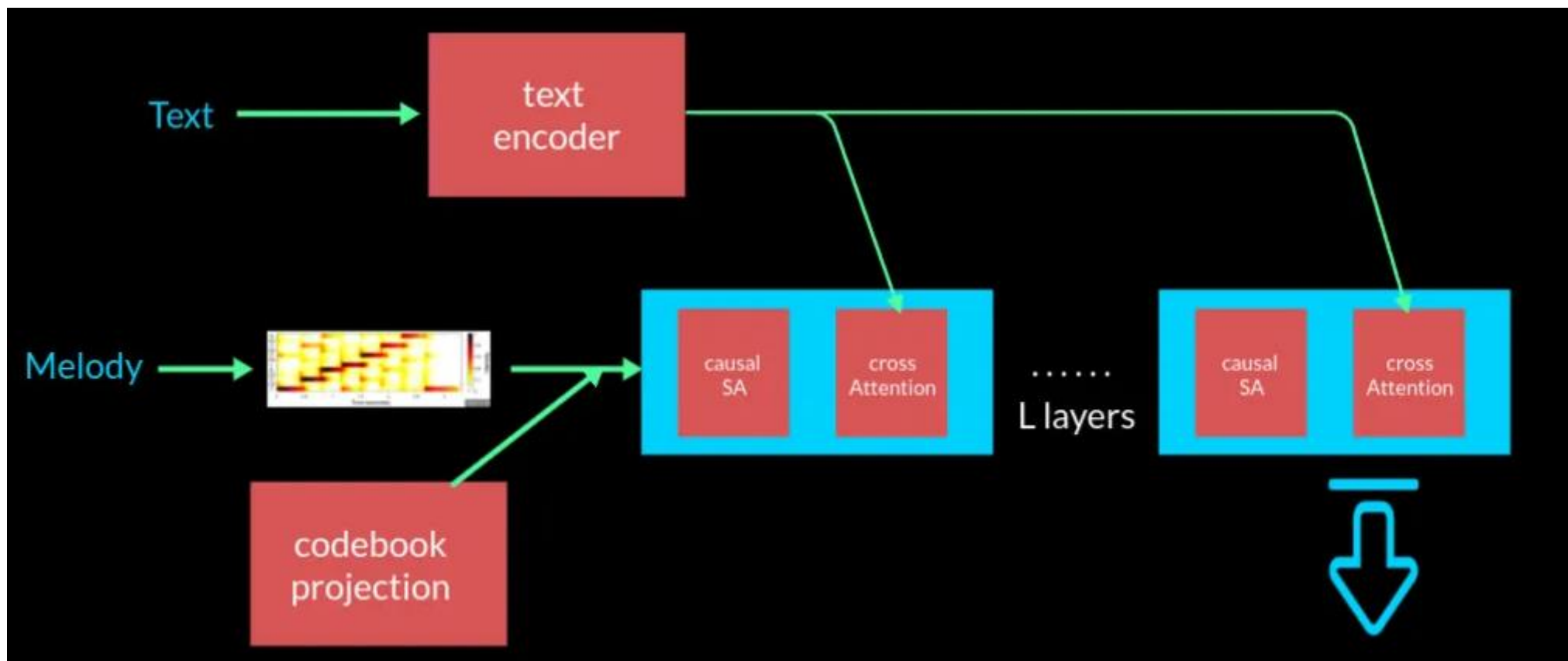


sum codebook values  
corresponding to  $t_1, k_2$

→ **Decoder**



# Decoder



# MusicGen流式音乐生成

基于MindNLP使用MusicGen-Small模型进行流式音乐生成，生成一段音乐片段后立即播放。（生成速度可能比较慢）

提示Prompt

80s pop track with synth and instrumentals

生成音乐长度

15

每个片段长度

时间越短解码步数越大

1.5

随机种子

5

Clear

Submit

🎵 请欣赏生成的音乐

Flag

☰ Examples

提示Prompt	生成音乐长度	每个片段长度	随机种子
An 80s driving pop song with heavy drums and synth pads in the background	30	1.5	5
A cheerful country song with acoustic guitars	30	1.5	5
90s rock song with electric guitar and heavy drums	30	1.5	5
a light and cheerful EDM track, with syncopated drums, airy pads, and strong emotions bpm:			

# 执行命令

```
conda create -n ms python=3.9 -y
```

```
conda activate ms
```

```
wget https://mindspore-demo.obs.cn-north-4.myhuaweicloud.com/mindnlp_install/mindnlp-0.3.1-py3-none-any.whl
```

```
pip install mindnlp-0.3.1-py3-none-any.whl
```

```
git clone https://github.com/mindspore-lab/mindnlp
```

```
cd mindnlp/llm/inference/musicgen
```

```
pip install gradio
```

```
python app_zh.py
```