Hi!

Benjamin (Beni) Eckhardt

25.1.’23

Benjamin (Beni) Eckhardt

Göttingen. Erasmus @ Tartu

25.1.’23

Benjamin (Beni) Eckhardt

Göttingen. Erasmus @ Tartu

Artificial and Natural Intelligence

25.1.’23

Benjamin (Beni) Eckhardt
Göttingen. Erasmus @ Tartu

Artificial and Natural Intelligence
Literature Study

25.1.'23

Benjamin (Beni) Eckhardt
Göttingen. Erasmus @ Tartu

Artificial and Natural Intelligence
Literature Study

Symbolic Music Representation Learning

25.1.'23

Benjamin (Beni) Eckhardt
Göttingen. Erasmus @ Tartu

Artificial and Natural Intelligence
Literature Study

Symbolic Music Representation Learning

publish.obsidian.md/nai22-project4a

25.1.'23

Benjamin (Beni) Eckhardt
Göttingen. Erasmus @ Tartu

Artificial and Natural Intelligence
Literature Study

Symbolic Music Representation Learning

publish.obsidian.md/nai22-project4a

beijn.github.io/nai22-project4a

25.1.’23

A Field Guide to
Symbolic Music Representation Learning

Title	Creator	Year	Item Type	Publication
A Survey on Artificial Intelligence for Music Generation: Agents, Domains and Perspectives	Hernandez-Olivan et al.	2022	Journal Article	
Attend to Chords: Improving Harmonic Analysis of Symbolic Music Using Transformer-Bas...	Chen and Su	2021	Journal Article	Transactions of the International Soc...
Code Generation Using Machine Learning: A Systematic Review	Dehaerne et al.	2022	Journal Article	IEEE Access
ComMU: Dataset for Combinatorial Music Generation	Hyun et al.	2022	Preprint	
Compound Word Transformer: Learning to Compose Full-Song Music over Dynamic Direct...	Hsiao et al.	2021	Journal Article	Proceedings of the AAAI Conferenc...
Content-Based Musical Similarity Computation using the Hierarchical Dirichlet Process	Hoffman et al.	2008	Conference ...	
Deep One-Class Classification	Ruff et al.	2018	Conference ...	Proceedings of the 35th Internation...
Deep Visual Instruments: Realtime Continuous, Meaningful Human Control over Deep Ne...	Akten, Memo	2021	Journal Article	
Detecting Symmetries of All Cardinalities With Application to Musical 12-Tone Rows	Venkatesh and Sachdev	2022	Conference ...	Proceedings of the First MiniCon Co...
DiffRoll: Diffusion-based Generative Music Transcription with Unsupervised Pretraining Ca...	Cheuk et al.	2022	Journal Article	
Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset	Hawthorne et al.	2022	Conference ...	
Encyclopedia of Machine Learning and Data Mining		2017	Book	
FIGARO: Generating Symbolic Music with Fine-Grained Artistic Control	Rutte et al.	2022	Journal Article	ArXiv
First Steps of an Approach to the ARC Challenge based on Descriptive Grid Models and the...	Ferré	2021	Journal Article	
From Music to Mathematics and Backwards: Introducing Algebra, Topology and Category T...	Andreatta	2018	Book Section	Imagine Math 6: Between Culture a...
Generating Diverse High-Fidelity Images with VQ-VAE-2	Razavi et al.	2019	Conference ...	Advances in Neural Information Pro...
Generation of musical patterns through operads	Giraudo	2021	Preprint	
Harmony Transformer: Incorporating Chord Segmentation into Harmony Recognition	Chen and Su	2019	Presentation	
Hybrid computing using a neural network with dynamic external memory	Graves et al.	2016	Journal Article	Nature
Learning Compositional Rules via Neural Program Synthesis	Nye et al.	2020	Conference ...	Advances in Neural Information Pro...
MidiTok: A python package for MIDI file tokenization	Nathan Fradet, Jean-Pierre Briot	2021	Conference ...	Extended abstracts for the late-brea...
Minimum description length revisited	Grünwald and Roos	2019	Journal Article	International Journal of Mathematics...
Multitrack Music Transformer	Dong et al.	2022	Journal Article	
Music Composition with Deep Learning: A Review	Hernandez-Olivan and Beltrán	2021	Journal Article	ArXiv
Music Translation: Generating Piano Arrangements in Different Playing Levels	Gover and Zewi	2022	Conference ...	Proceedings of the First MiniCon Co...
musicaiz: A Python Library for Symbolic Music Generation, Analysis and Visualization	Hernandez-Olivan and Beltran	2022	Journal Article	
Nested Hierarchical Dirichlet Processes	Paisley et al.	2015	Journal Article	IEEE Transactions on Pattern Analysis...
Neural Abstract Reasoner	Kolev et al.	2020	Journal Article	
Neural Discrete Representation Learning	Oord et al.	2017	Conference ...	Advances in Neural Information Pro...
Omnizart: A General Toolbox for Automatic Music Transcription	Wu et al.	2021	Journal Article	Journal of Open Source Software
Pop Music Transformer: Beat-based Modeling and Generation of Expressive Pop Piano Co...	Huang and Yang	2020	Conference ...	Proceedings of the 28th ACM Intern...
Rethinking Attention with Performers	Choromanski et al.	2020	Journal Article	ArXiv
Sequence-to-Sequence Learning with Latent Neural Grammars	Kim	2021	Conference ...	Advances in Neural Information Pro...

ISMIR 2022
BENGALURU

NEURAL INFORMATION
PROCESSING SYSTEMS

JOSS

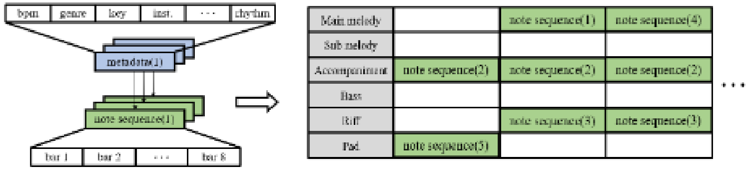
acm
Association for
Computing Machinery

nature

Dataset for Combinatorial Music Generation

automated mainstream drag-and-drop composing from high quality metadata

[1], #NIPS 2022
NIPS workshop page and source



(a) stage1 (b) stage2

Figure 1: **The whole process of *Combinatorial Music Generation*.** In stage1, a note sequence (green box) is generated from a set (blue box) of metadata. Stage2 then combines note sequences generated from several metadata sets to create one completed music. The number of bars for a note sequence and the number of note sequences for a complete piece of music can be flexible (*ComMU* mostly has a note sequence of 4, 8, and 16 bars). In this work, we focus on solving stage1.

Motivation

- Industrializable artistic control over #generative models (in preformatted mainstream domains)

Tasks

- Combinatorial Music Generation (#conditional) (training is #supervised)

Ideas

- high quality #human-expert data with powerful domain specific constraints ("12 metadata")
- #autoregressive #sequence-completion (with Transformer-XL[2]) as a baseline to show potential
  - #teacher-forcing : #chord symbols are injected to intended positions during generation

Inductive Biases

- music = repeated overlay of unrelated, short, monophonic sequences. Relation only through common metadata in prompt.
- The usual flatness concern in latent codes applies (see Hierarchical Latents)

Context

- see also ComMU Dataset

On FIGARO

- similar in Description to Sequence, but in FIGARO:
- description is automatically generated from a bigger (unsorted) dataset
- using fewer assumptions
  - and thus fewer description components
  - #per-bar , thus less overall structure, but more flexible

On MMM[3]

- is the most similar in previous literature, but
- not feasable for Combinatorial Music Generation > Homophonic Composition, because of no track role- nor chord-conditioning
- specialized for #resampling or Inpainting

Data

- see ComMU Dataset
- #data-augmentation : with transposition and time dilation, yields ×60 data amount

Evaluation

Controllability

- whether metadata information is faithfully adhered to
- determinable via simple algorithms, that heuristically extract information like note density, chords, tempo

Diversity

- Sum of pairwise #cosine-similarity between Chroma Vectors (#harmony) and Groove[4] (#rhythm) across multiple generations from same metadata

Fidelity

- #winrate between generated and real sample (#human-survey)

Results

- Sampling > Temperature controls a between trade-off Controllability and Diversity.
- Sampling > Top-K ([5], [6]) K-value has neglegible effect - unlike with text
  - due to smaller vocab size

Generation from humans

- composers tasked with extracting composing guidlines for specific industrial genres that use Combinatorial Music Generation > Homophonic Composition.
- Composers tasked with composing short MIDI sequences that match this guidlines.
- Composed MIDI sequence and guidline (metadata) converted to REMI-like tokens.

Metadata

- precise tokens: *bpm, key, instrument, time signature, pitch range, number of measure, min/max velocity, and rhythm*
  - tempo constant accross sequence
- ambiguous tokens: *genre, track role* and *chord progression.*
  Comparison to other datasets regarding metadata given below.

On other Datsets

Table 1: Comparison of ComMU to recent MIDI datasets with various metadata. We compare ComMU to other MIDI dataset on 4 types of metadata: genre, instrument, track-role, and chord progression.

Dataset	Genre	Instrument	Track-role	Chord progression
Lakh MIDI [10]	✓	✓	-	-
MAESTRO [12]	✓	(✓)[1]	-	-
MSMD [13]	-	(✓)[1]	-	✓
ADL Piano MIDI [14]	✓	(✓)[1]	-	-
GiantMIDI-Piano [15]	-	(✓)[1]	-	✓
EMOPIA [23]	-	(✓)[1]	-	✓
ComMU(Ours)	✓	✓	✓	✓

[1] include only one instrument.

See Table 1 of ComMU for an overview and Section 2 for a discussion.

- "Lakh MIDI Dataset [10], MAESTRO [12], and ADL Piano MIDI [14] have various meta-information such as genre, instrument, key, and time signature, but there is *no chord information* for the harmony of track-level composition."
- "MSMD [13], GiantMIDI-Piano [15], and EMOPIA [23] have chord information that can infuse harmony, but *lack metadata* such as genre, and rhythm that can reflect the intention of composition."
- separates instruments from track role, other than "[10,28,29]"
  - in Symbolic Music Loop Generation both are the same as well

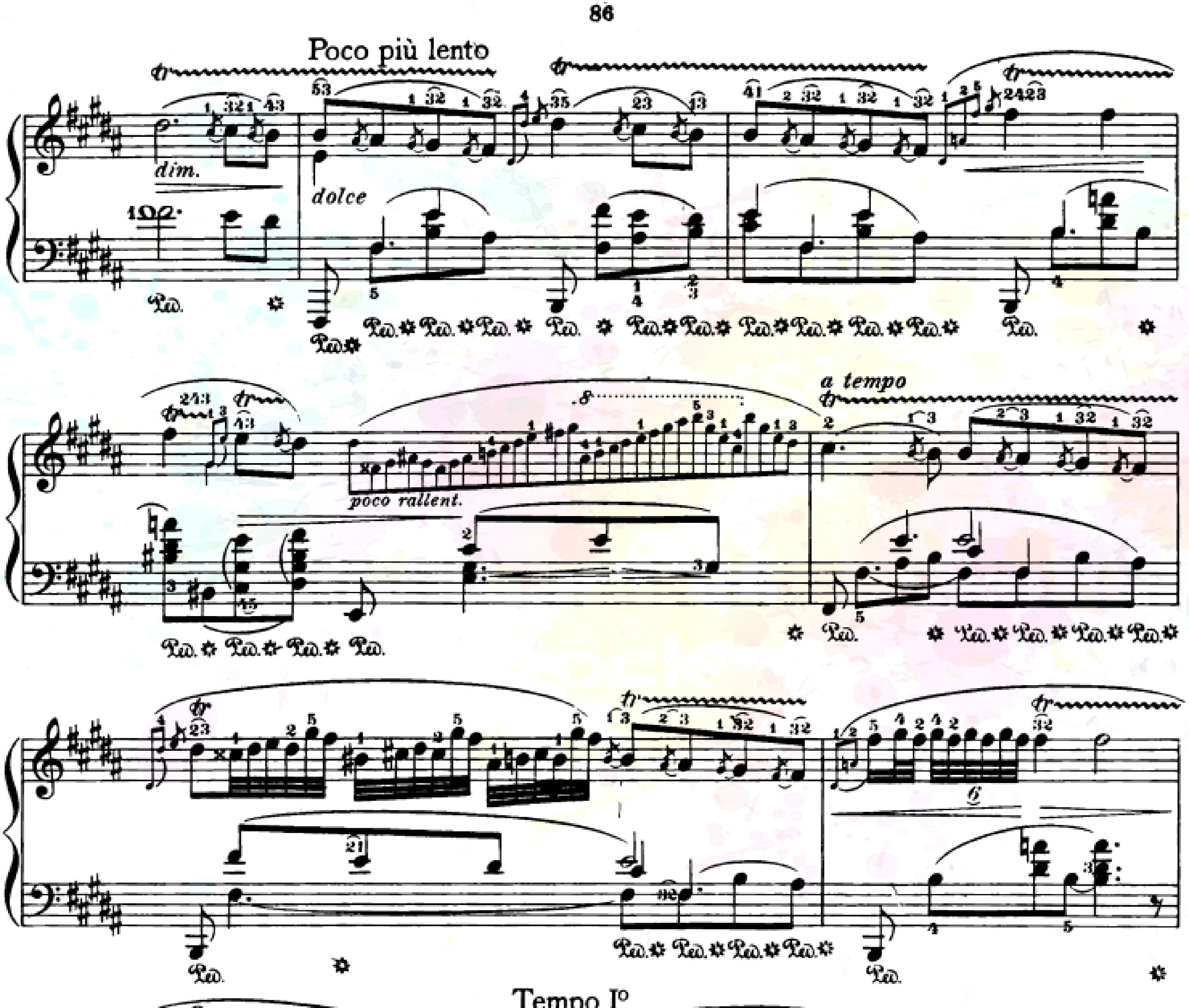
1. Hyun, Lee, Taehyun Kim, Hyolim Kang, Minjoo Ki, Hyeonchan Hwang, Kwanho Park, Sharang Han, and Seon Joo Kim. 2022. "ComMU: Dataset for Combinatorial Music Generation." arXiv. https://doi.org/10.48550/arXiv.2211.09385. ↩ ↩

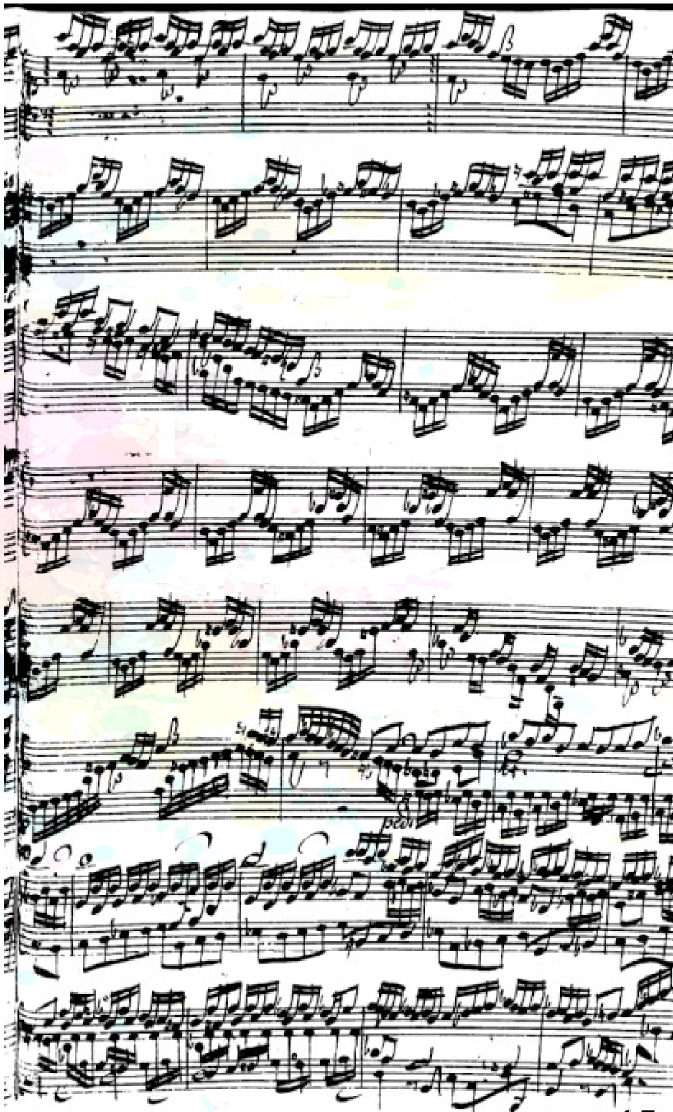
Future Work

- to develop specific architectures for the structure of music rather than text

1. Hyun, Lee, Taehyun Kim, Hyolim Kang, Minjoo Ki, Hyeonchan Hwang, Kwanho Park, Sharang Han, and Seon Joo Kim. 2022. "ComMU: Dataset for Combinatorial Music Generation." arXiv. https://doi.org/10.48550/arXiv.2211.09385. ↩
2. Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G Carbonell, Quoc Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pages 2978–2988, 2019. ↩
3. Jeff Ens and Philippe Pasquier. Mmm: Exploring conditional multi-track music generation with the transformer. arXiv preprint arXiv:2008.06048, 2020. ↩
4. Simon Dixon, Fabien Gouyon, Gerhard Widmer, et al. Towards characterisation of music via rhythmic patterns. In International Society for Music Information Retrieval (ISMIR), 2004. ↩
5. Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), pages 889–898, 2018. ↩
6. Ari Holtzman, Jan Buys, Maxwell Forbes, Antoine Bosselut, David Golub, and Yejin Choi. Learning to write with cooperative discriminators. In Annual Meeting of the Association for Computational Linguistics (ACL), 2018. ↩

A Field Guide to Symbolic ***Music*** Representation Learning

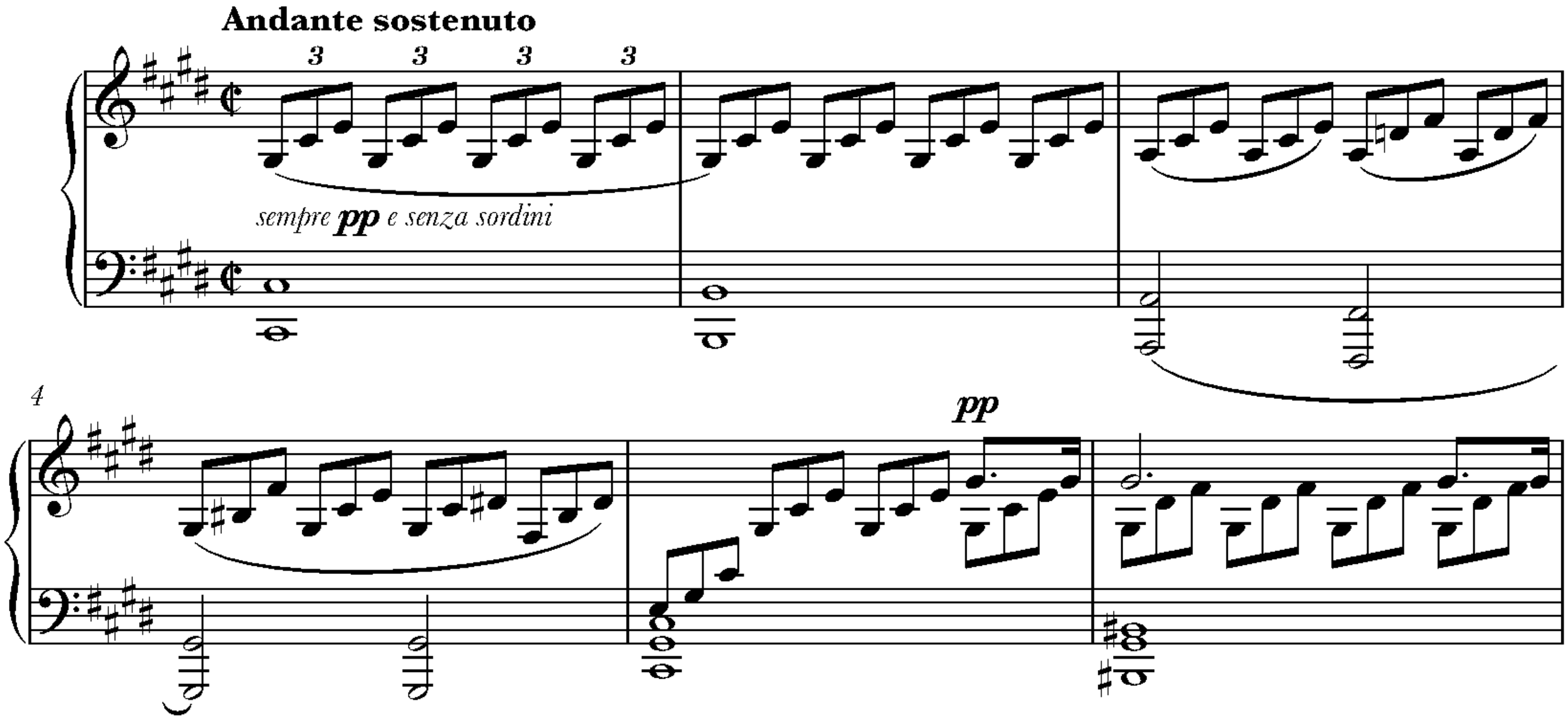

86
Poco più lento
dim.
dolce
poco rallent.
a tempo
Tempo I°

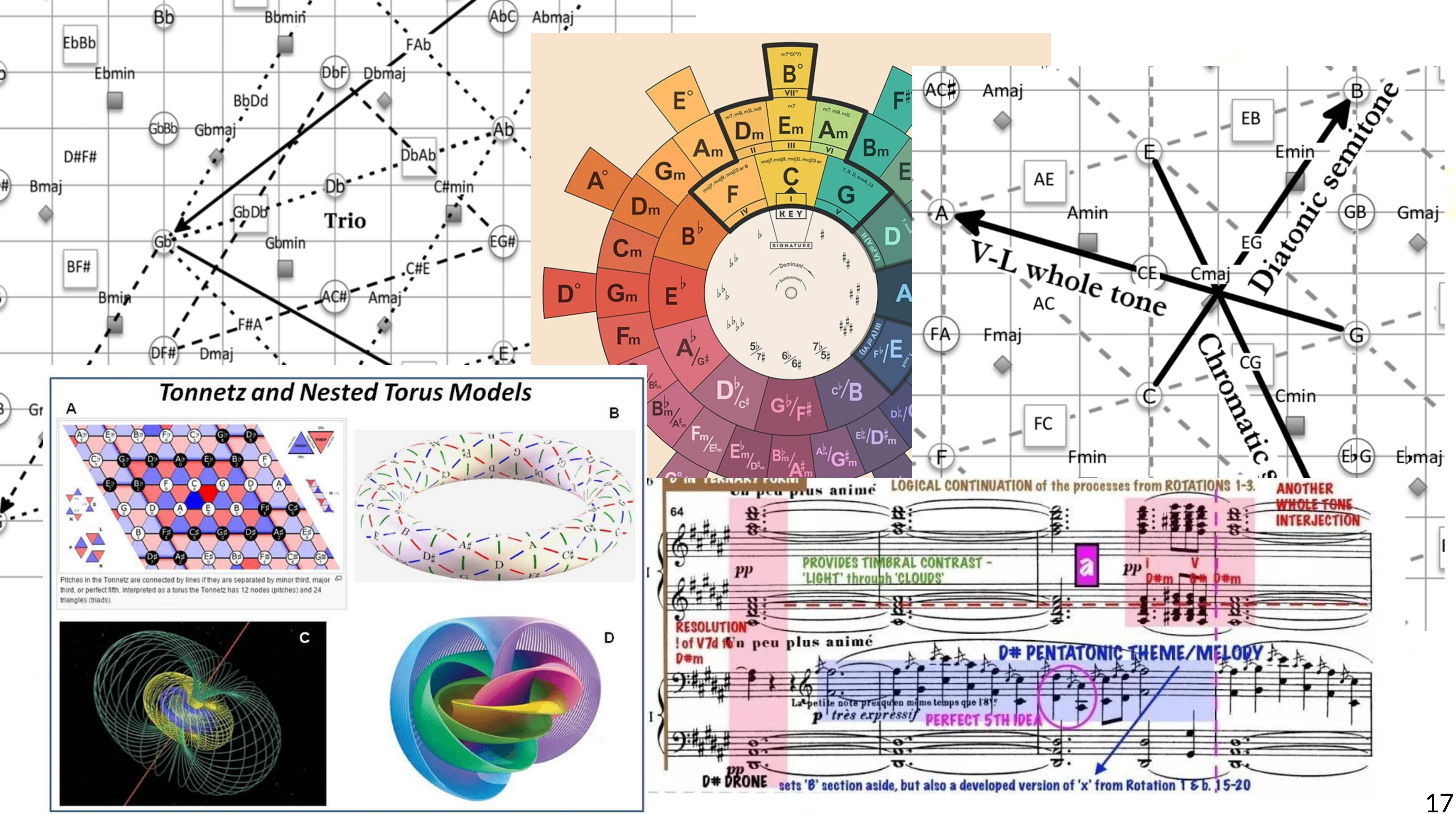


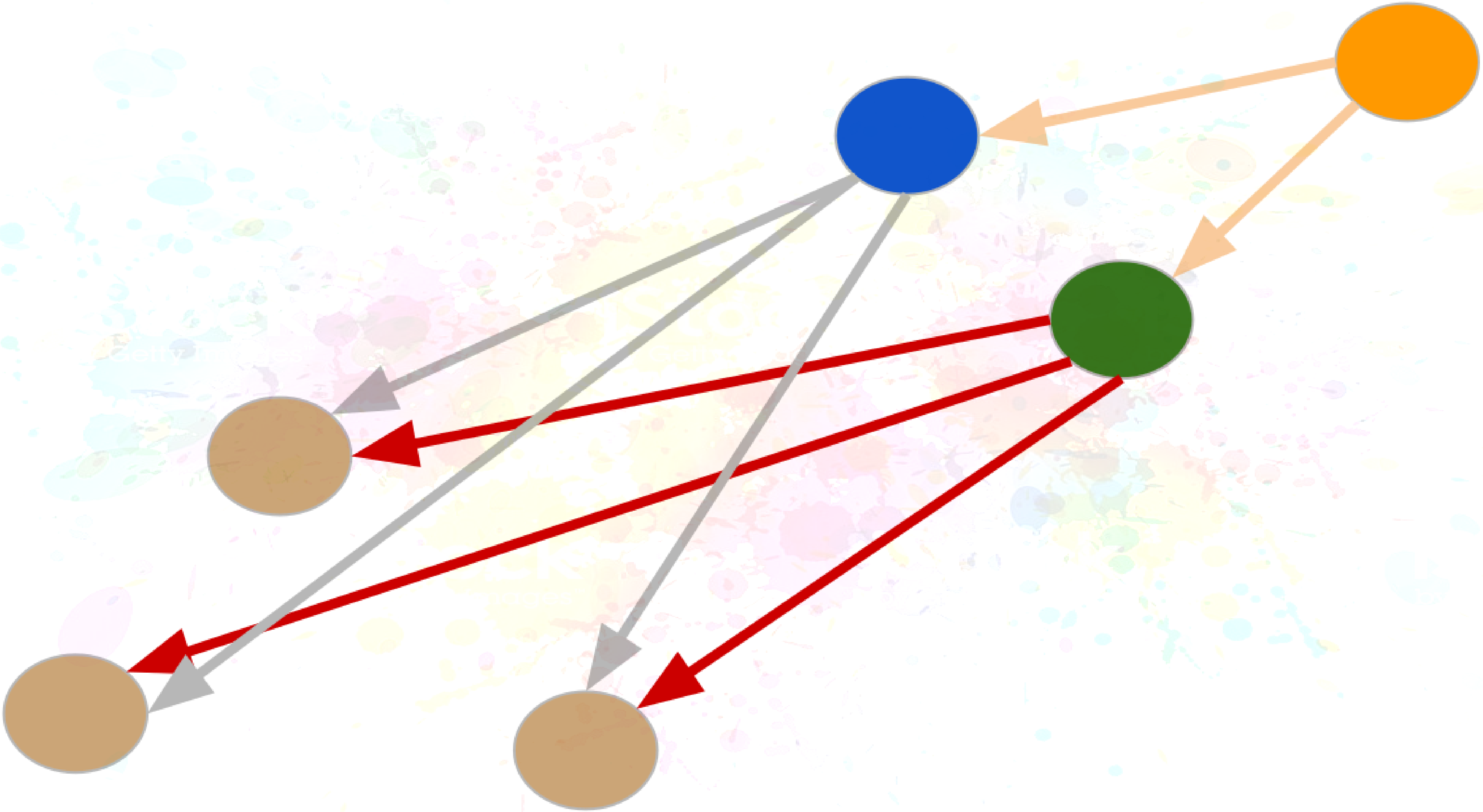
Sonate

Sonata quasi una Fantasia ('The Moonlight')

L.van Beethoven Op 27 No2

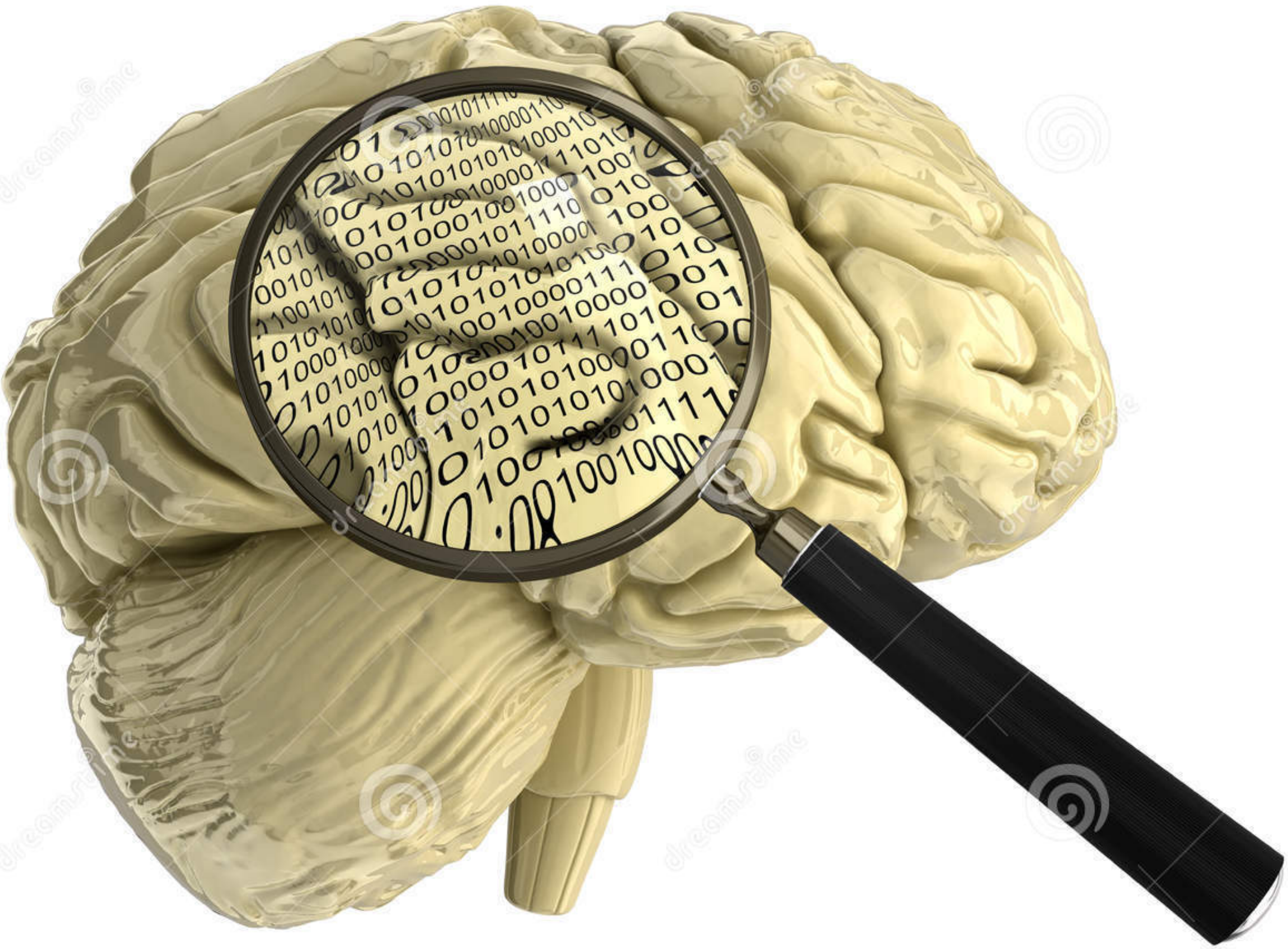


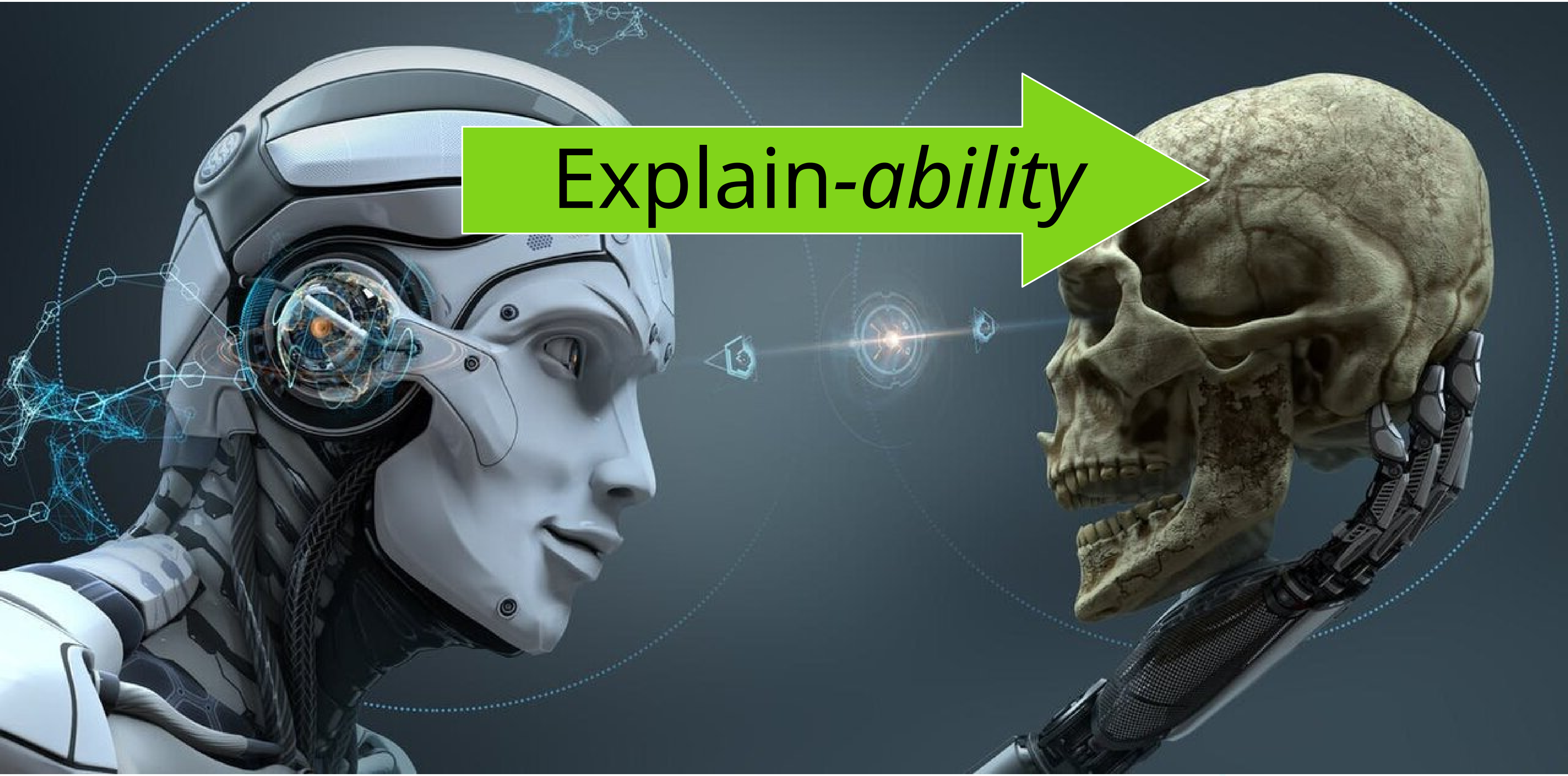

Trio
Tonnetz and Nested Torus Models
A
B
C
D
Pitches in the Tonnetz are connected by lines if they are separated by minor third, major third, or perfect fifth. Interpreted as a torus the Tonnetz has 12 nodes (pitches) and 24 triangles (triads).
KEY
SIGNATURE
V-L whole tone
Diatonic semitone
Chromatic
Amaj
Emin
Amin
Gmaj
Cmaj
Fmaj
Cmin
Fmin
LOGICAL CONTINUATION of the processes from ROTATIONS 1-3.
ANOTHER WHOLE TONE INTERJECTION
Un peu plus animé
PROVIDES TIMBRAL CONTRAST - 'LIGHT' through 'CLOUDS'
RESOLUTION
D#m
D# PENTATONIC THEME/MELODY
très expressif
PERFECT 5TH IDEA
D# DRONE
sets 'B' section aside, but also a developed version of 'x' from Rotation 1 & b. 15-20



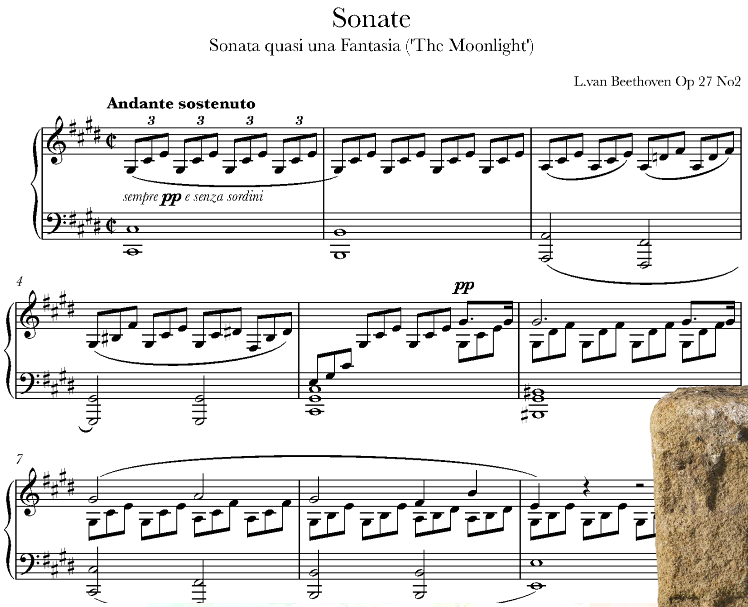
Source: NAI lecture 8, p54

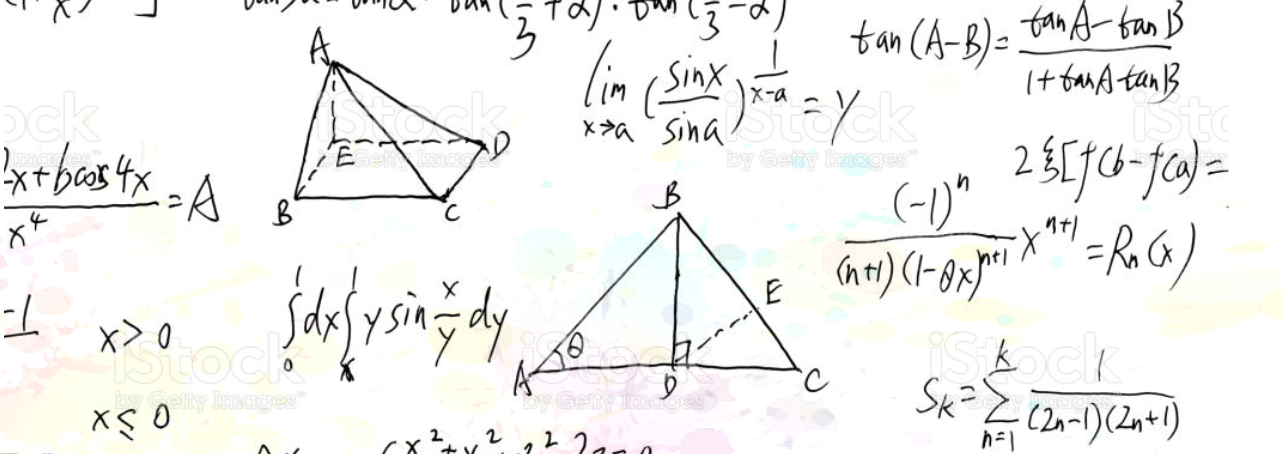
A Field Guide to
Symbolic Music ***Representation*** Learning





Credit: William Bradberry/Shutterstock



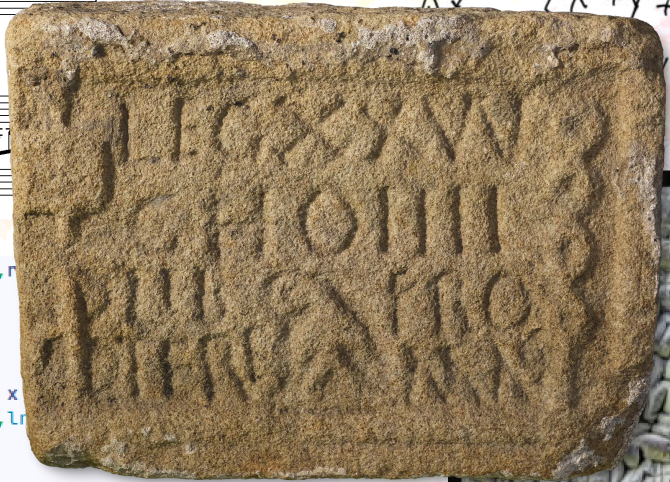



```
bifold :: Foldable t => (x -> (l,r) -> (r,l)) -> (l,r
bifold f lr xs = foldr (\x g -> \(l,r) -> let
  (r'',l'') = g (l',r); (r' ,l' ) = f x (l,r'')
  in (r', l'')) swap xs lr

bifold' :: Foldable t => (x -> rl -> rl) -> rl -> t x
bifold' f lr = bifold (\x -> swap . both (f x)) (lr,lr


foldI :: Foldable t => a -> (Int -> e -> a -> a) -> t e -> a
foldI a f s = snd $ fold (len s, a) (\e (i,a) -> (i-1, f i e a)) s

foldIs :: Foldable t => [Int] -> a -> (Bool -> Int -> e -> a -> a) -> t e -> a
foldIs (sort > reverse > (++ cycle [-1]) -> is) a f = snd . foldI
  (is, a)  \ j e (i:is, a) -> (i==j & is ? (i:is), f (i==j) j e a)
```



Cuneiform tablet image credit: Lecture “Keilschrift über drei Jahrtausende” by Klaus Wagensonner, 20/21 Göttingen

A Field Guide to

Symbolic Music Representation Learning

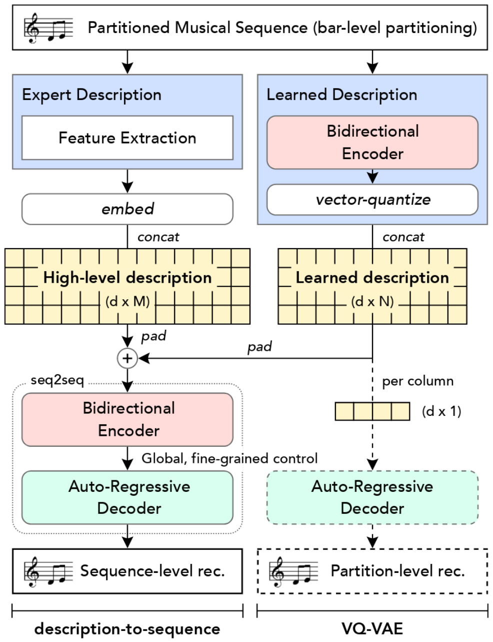
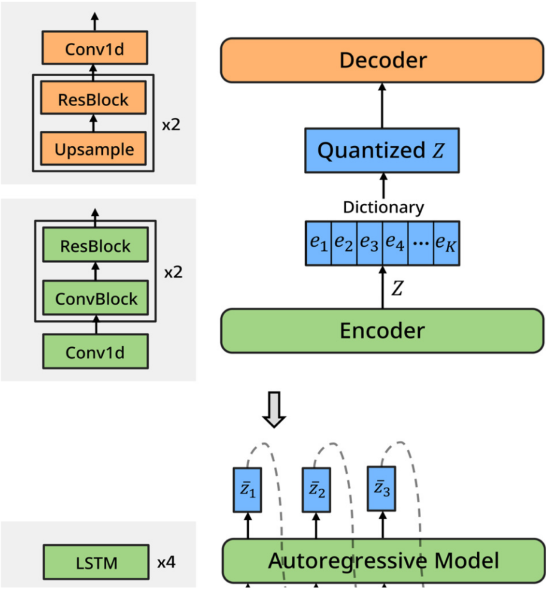


Figure 1. Overview of FIGARO. Dashed lines indicate components that are only used during training. d denotes the hidden dimension of the model, N the number of bars and M the length of the expert description ($M > N$ in our case).



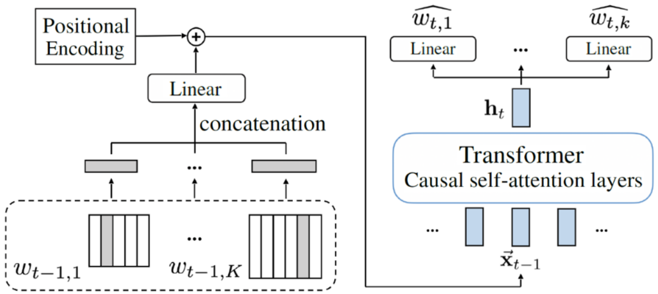


Figure 1: Illustration of the main ideas of the proposed compound word Transformer: (left) *compound word modeling* that combines the embeddings (colored gray) of multiple tokens $\{w_{t-1,k}\}_{k=1}^{K}$, one for each token type k, at each time step $t-1$ to form the input $\vec{\mathbf{x}}_{t-1}$ to the self-attention layers, and (right) *toke type-specific feed-forward heads* that predict the list of tokens for the next time step t at once at the output.

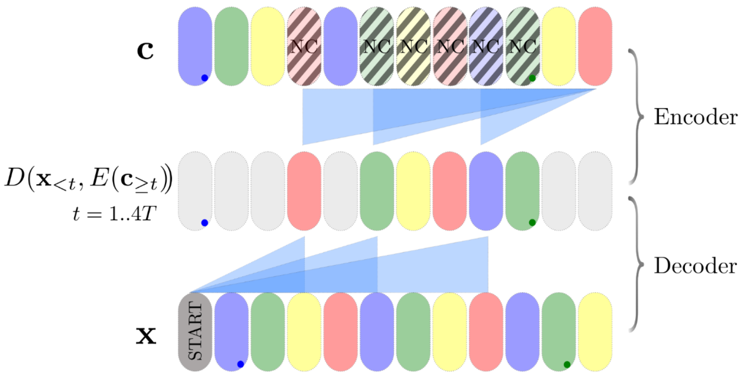


Figure 3: Schematization of our proposed masking scheme. Output tokens (middle row) can only attend to $\mathbf{x}$ and $\mathbf{c}$ tokens delimited by blue triangles; $\mathbf{c}$ is a masked version of $\mathbf{x}$. For ease of implementation, a $\mathbf{START}$ token is appended before $\mathbf{x}$. Our loss function only involves predictions over Non Constrained tokens.

$$\mathcal{B}^{\mathrm{arp}}_{\mathbf{p},\mathbf{d}} := \left(\mathsf{P}_m, \mathfrak{C}, \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}, \mathsf{b}_1\right) \tag{77}$$

$$\mathbf{c}_1 := \left(\mathsf{b}_1, [\mathbf{p}]_m, \mathsf{b}_2^m\right), \quad \mathbf{c}_2 := \left(\mathsf{b}_2, [\mathbf{p}]_m, \mathsf{b}_2^m\right), \tag{78a}$$

$$\mathbf{c}_3 := \left(\mathsf{b}_2, \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_m \end{bmatrix}, \mathsf{b}_3\right). \tag{78b}$$

$$\mathbf{p} := 2102\square1\square0\square$$
$$\mathbf{d} := 05\bar{7}$$
$$k := 3$$

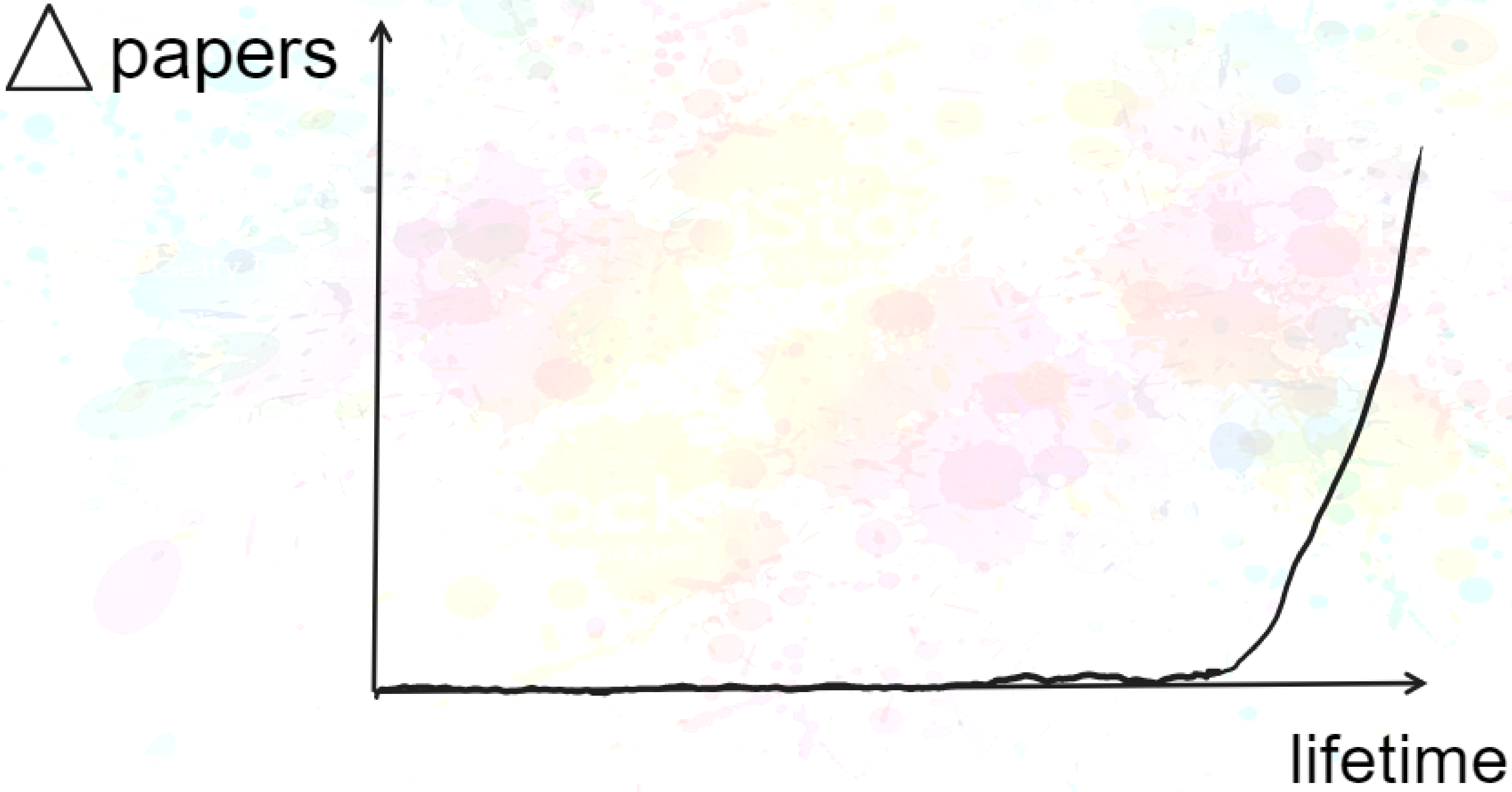
$$\begin{bmatrix} 2 & 1 & 0 & 2 & \square & 1 & \square & 0 & \square \\ 2 & 6 & 5 & 2 & \square & 1 & \square & 0 & \square \\ 2 & \bar{6} & \bar{7} & 2 & \square & 1 & \square & 0 & \square \end{bmatrix}.$$

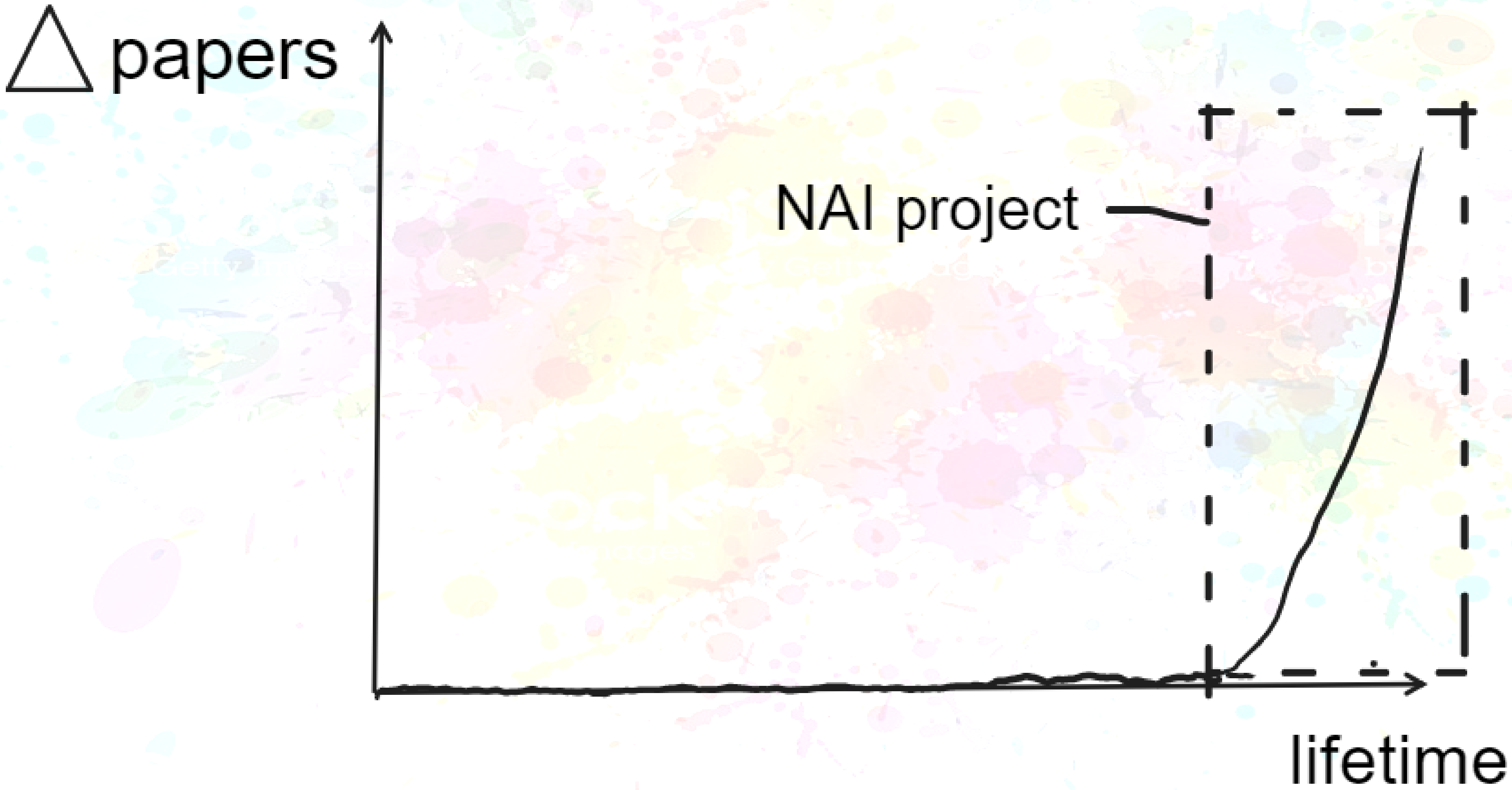
128 $(\boldsymbol{\lambda}, 9_3)$

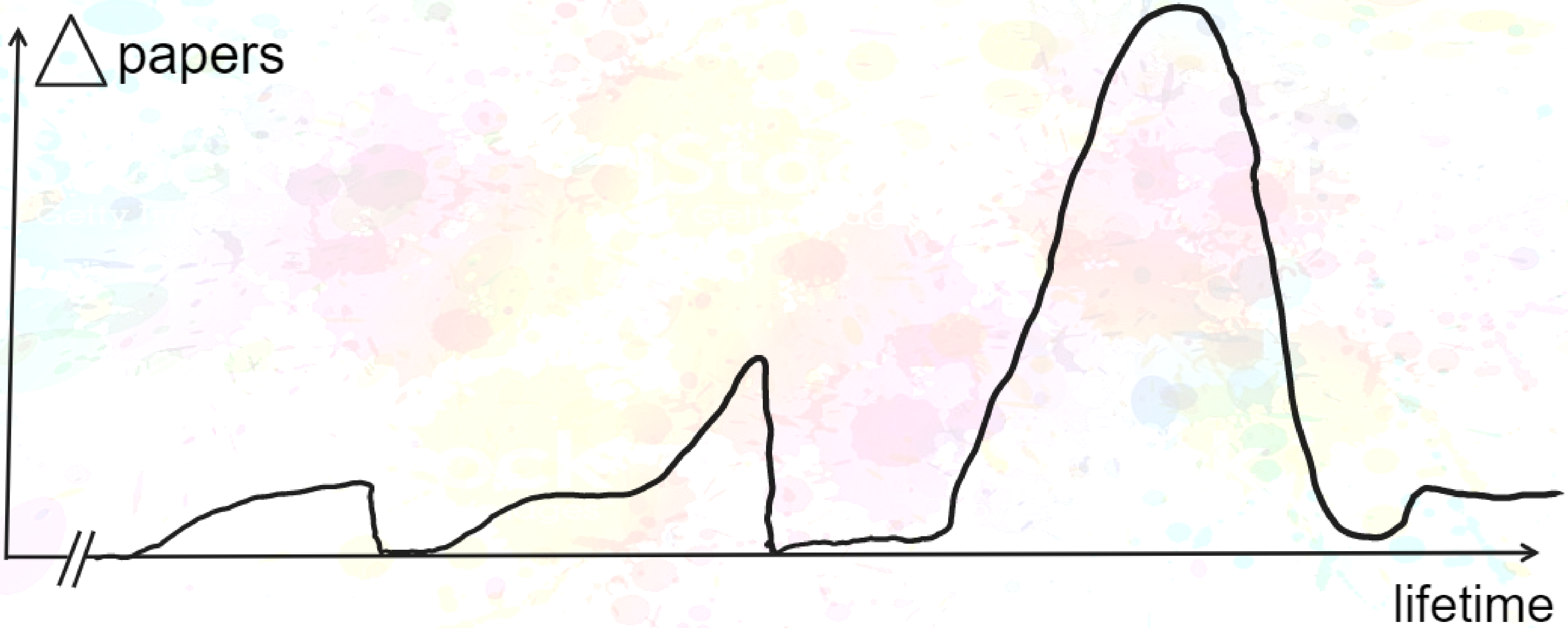
Source: [1] S. Giraudo, “Generation of musical patterns through operads.” arXiv, Apr. 26, 2021. doi: 10.48550/arXiv.2104.12432.

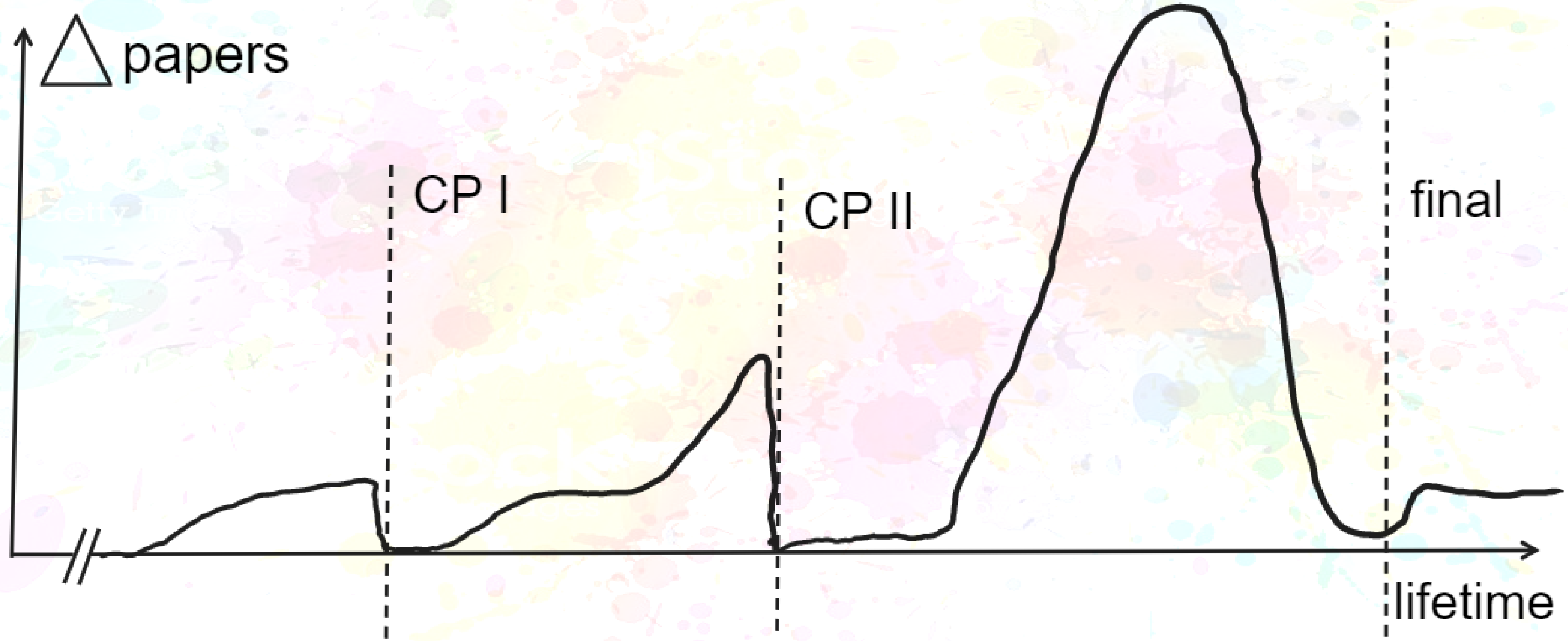
A Field Guide to Symbolic Music Representation ***Learning***

Title	Creator	Year	Item Type	Publication
A Survey on Artificial Intelligence for Music Generation: Agents, Domains and Perspectives	Hernandez-Olivan et al.	2022	Journal Article	
Attend to Chords: Improving Harmonic Analysis of Symbolic Music Using Transformer-Bas...	Chen and Su	2021	Journal Article	Transactions of the International Soc...
Code Generation Using Machine Learning: A Systematic Review	Dehaerne et al.	2022	Journal Article	IEEE Access
ComMU: Dataset for Combinatorial Music Generation	Hyun et al.	2022	Preprint	
Compound Word Transformer: Learning to Compose Full-Song Music over Dynamic Direct...	Hsiao et al.	2021	Journal Article	Proceedings of the AAAI Conferenc...
Content-Based Musical Similarity Computation using the Hierarchical Dirichlet Process	Hoffman et al.	2008	Conference ...	
Deep One-Class Classification	Ruff et al.	2018	Conference ...	Proceedings of the 35th Internation...
Deep Visual Instruments: Realtime Continuous, Meaningful Human Control over Deep Ne...	Akten, Memo	2021	Journal Article	
Detecting Symmetries of All Cardinalities With Application to Musical 12-Tone Rows	Venkatesh and Sachdev	2022	Conference ...	Proceedings of the First MiniCon Co...
DiffRoll: Diffusion-based Generative Music Transcription with Unsupervised Pretraining Ca...	Cheuk et al.	2022	Journal Article	
Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset	Hawthorne et al.	2022	Conference ...	
Encyclopedia of Machine Learning and Data Mining		2017	Book	
FIGARO: Generating Symbolic Music with Fine-Grained Artistic Control	Rutte et al.	2022	Journal Article	ArXiv
First Steps of an Approach to the ARC Challenge based on Descriptive Grid Models and the...	Ferré	2021	Journal Article	
From Music to Mathematics and Backwards: Introducing Algebra, Topology and Category T...	Andreatta	2018	Book Section	Imagine Math 6: Between Culture a...
Generating Diverse High-Fidelity Images with VQ-VAE-2	Razavi et al.	2019	Conference ...	Advances in Neural Information Pro...
Generation of musical patterns through operads	Giraudo	2021	Preprint	
Harmony Transformer: Incorporating Chord Segmentation into Harmony Recognition	Chen and Su	2019	Presentation	
Hybrid computing using a neural network with dynamic external memory	Graves et al.	2016	Journal Article	Nature
Learning Compositional Rules via Neural Program Synthesis	Nye et al.	2020	Conference ...	Advances in Neural Information Pro...
MidiTok: A python package for MIDI file tokenization	Nathan Fradet, Jean-Pierre Briot	2021	Conference ...	Extended abstracts for the late-brea...
Minimum description length revisited	Grünwald and Roos	2019	Journal Article	International Journal of Mathematics...
Multitrack Music Transformer	Dong et al.	2022	Journal Article	
Music Composition with Deep Learning: A Review	Hernandez-Olivan and Beltrán	2021	Journal Article	ArXiv
Music Translation: Generating Piano Arrangements in Different Playing Levels	Gover and Zewi	2022	Conference ...	Proceedings of the First MiniCon Co...
musicaiz: A Python Library for Symbolic Music Generation, Analysis and Visualization	Hernandez-Olivan and Beltran	2022	Journal Article	
Nested Hierarchical Dirichlet Processes	Paisley et al.	2015	Journal Article	IEEE Transactions on Pattern Analysis...
Neural Abstract Reasoner	Kolev et al.	2020	Journal Article	
Neural Discrete Representation Learning	Oord et al.	2017	Conference ...	Advances in Neural Information Pro...
Omnizart: A General Toolbox for Automatic Music Transcription	Wu et al.	2021	Journal Article	Journal of Open Source Software
Pop Music Transformer: Beat-based Modeling and Generation of Expressive Pop Piano Co...	Huang and Yang	2020	Conference ...	Proceedings of the 28th ACM Intern...
Rethinking Attention with Performers	Choromanski et al.	2020	Journal Article	ArXiv
Sequence-to-Sequence Learning with Latent Neural Grammars	Kim	2021	Conference ...	Advances in Neural Information Pro...

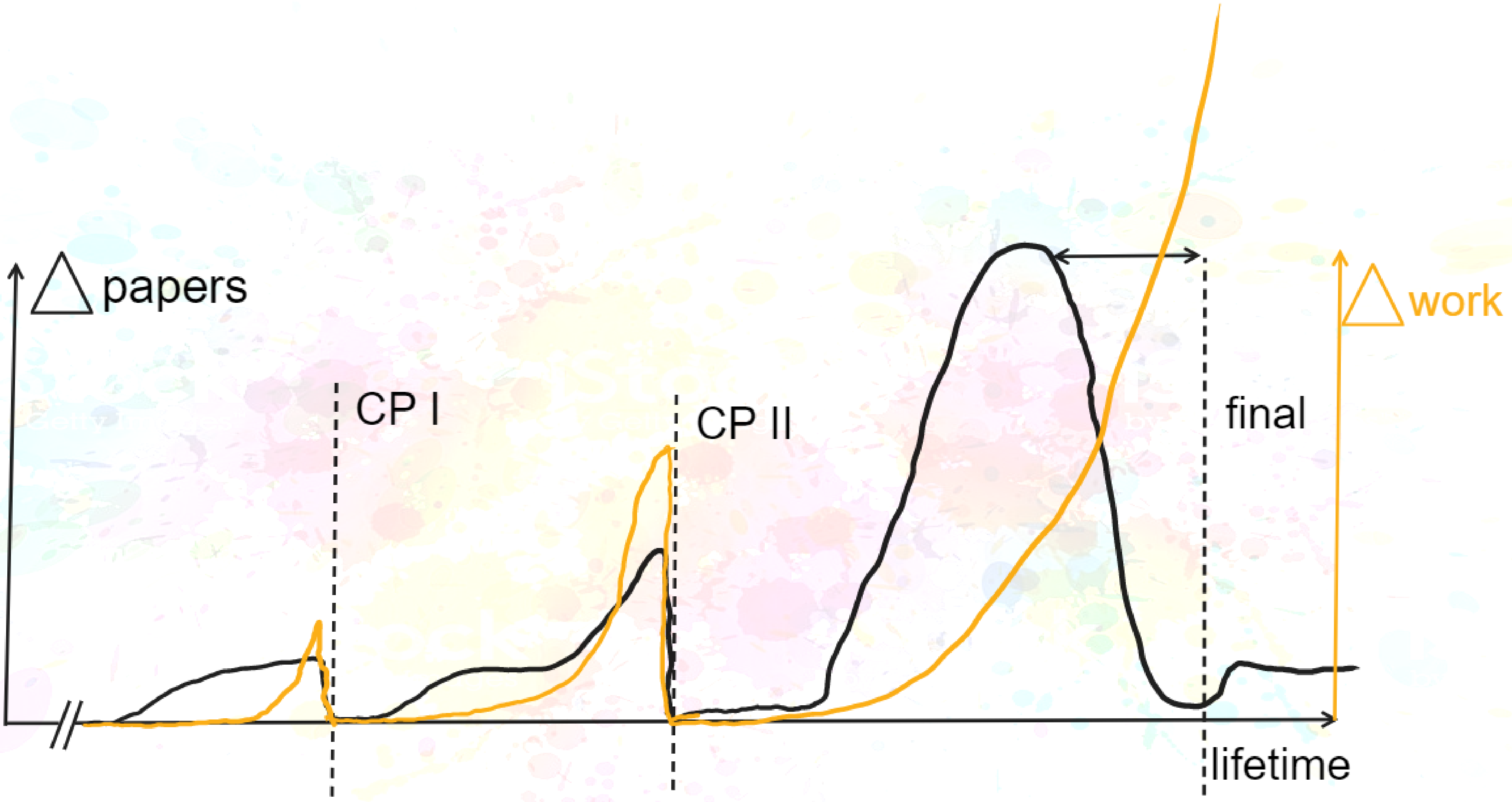

△ papers
lifetime


△ papers
NAI project
lifetime


△ papers
lifetime



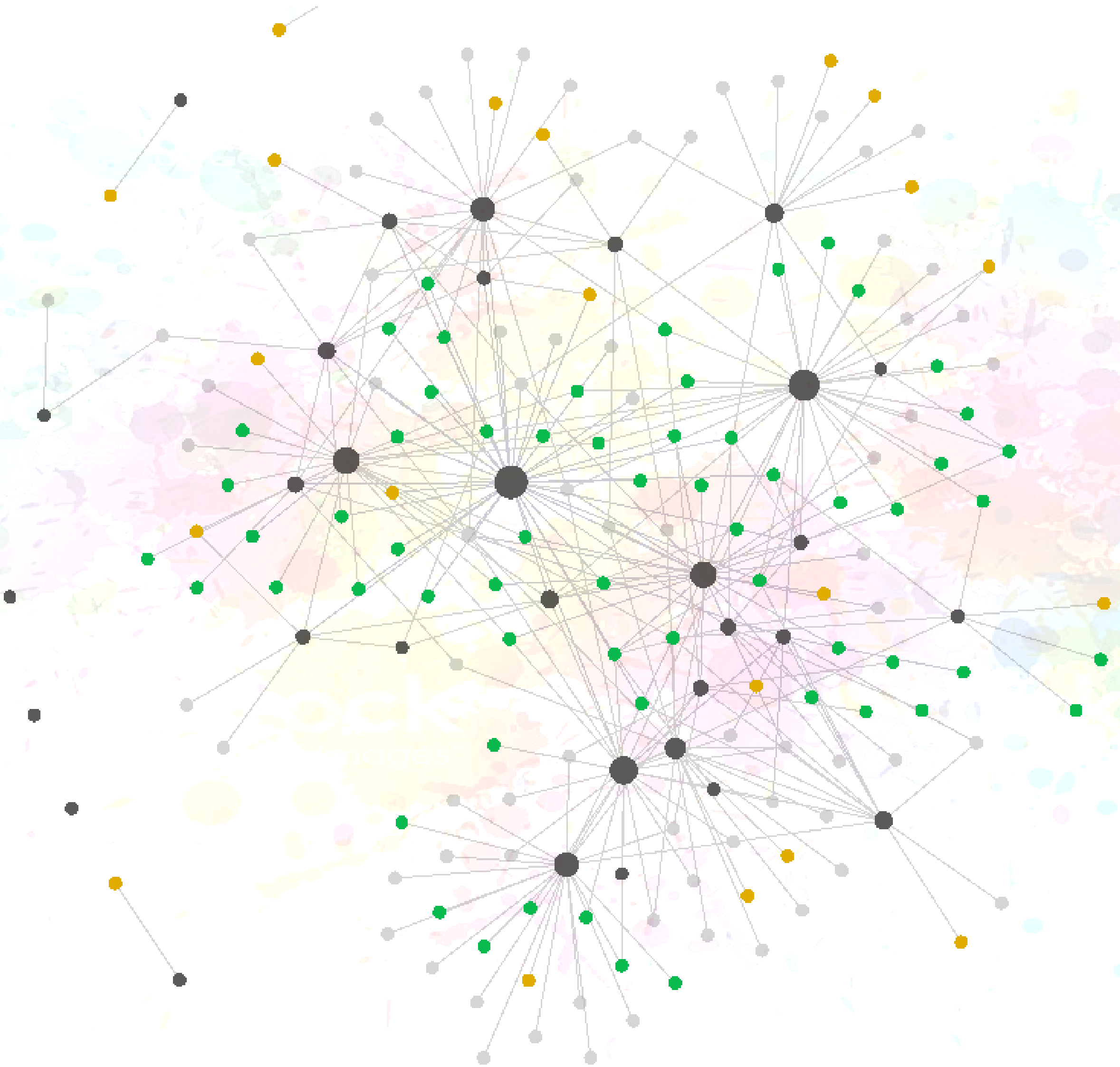
△ papers
CP I
CP II
final
lifetime


△ papers
CP I
CP II
final
△ work
lifetime

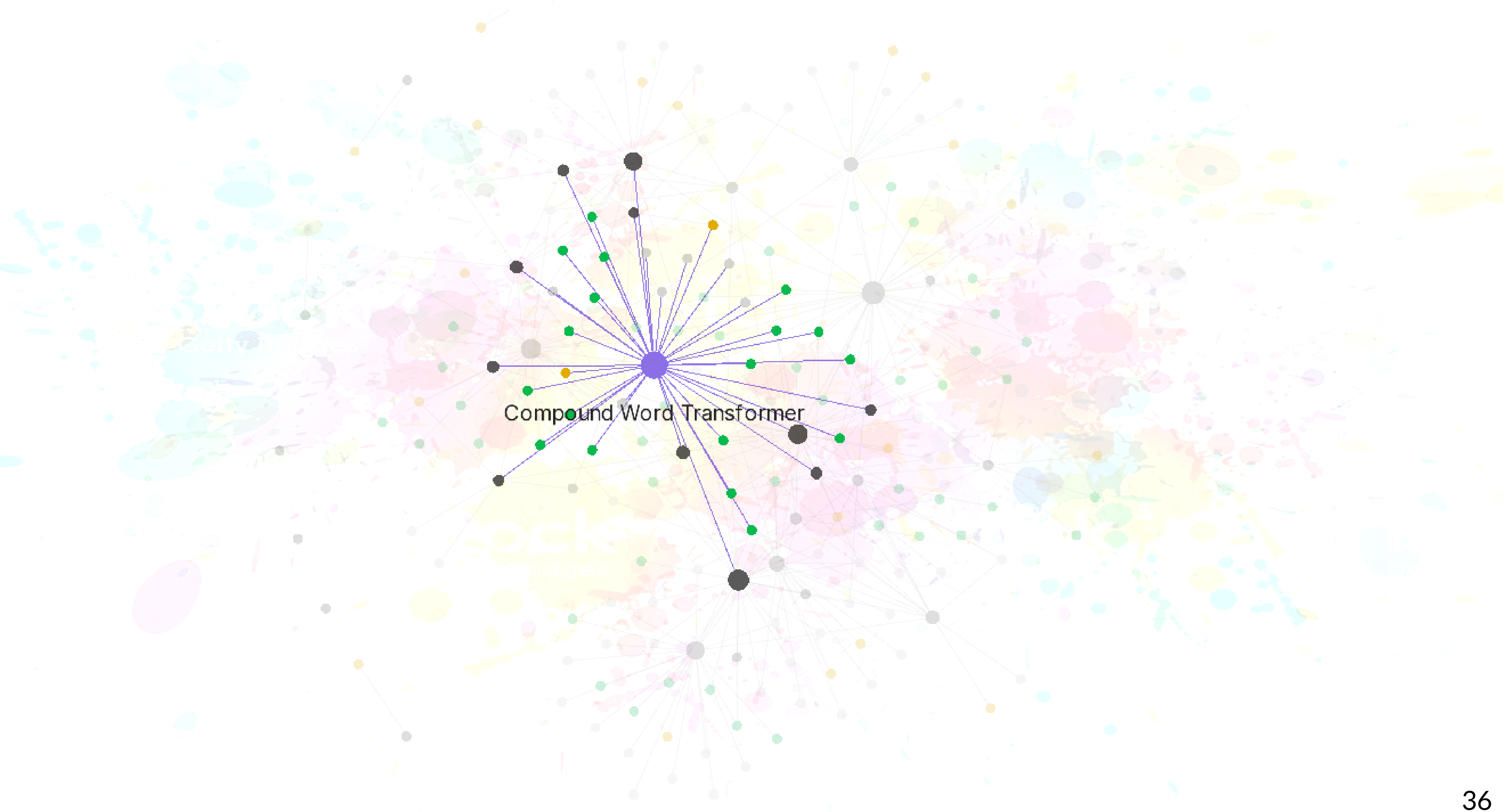
A Field Guide to

Symbolic Music Representation Learning

Graph view



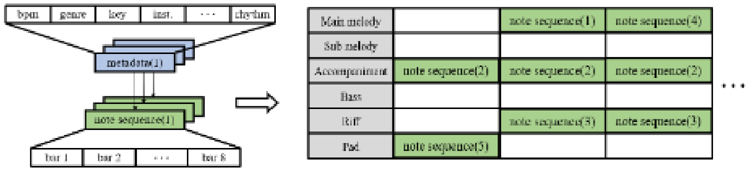
Graph view



Dataset for Combinatorial Music Generation

automated mainstream drag-and-drop composing from high quality metadata

[1], #NIPS 2022
NIPS workshop page and source



(a) stage1 (b) stage2

Figure 1: **The whole process of *Combinatorial Music Generation*.** In stage1, a note sequence (green box) is generated from a set (blue box) of metadata. Stage2 then combines note sequences generated from several metadata sets to create one completed music. The number of bars for a note sequence and the number of note sequences for a complete piece of music can be flexible (*ComMU* mostly has a note sequence of 4, 8, and 16 bars). In this work, we focus on solving stage1.

Motivation

- Industrializable artistic control over #generative models (in preformatted mainstream domains)

Tasks

- Combinatorial Music Generation (#conditional) (training is #supervised)

Ideas

- high quality #human-expert data with powerful domain specific constraints ("12 metadata")
- #autoregressive #sequence-completion (with Transformer-XL[2]) as a baseline to show potential
  - #teacher-forcing : #chord symbols are injected to intended positions during generation

Inductive Biases

- music = repeated overlay of unrelated, short, monophonic sequences. Relation only through common metadata in prompt.
- The usual flatness concern in latent codes applies (see Hierarchical Latents)

Context

- see also ComMU Dataset

On FIGARO

- similar in Description to Sequence, but in FIGARO:
- description is automatically generated from a bigger (unsorted) dataset
- using fewer assumptions
  - and thus fewer description components
  - #per-bar , thus less overall structure, but more flexible

On MMM[3]

- is the most similar in previous literature, but
- not feasable for Combinatorial Music Generation > Homophonic Composition, because of no track role- nor chord-conditioning
- specialized for #resampling or Inpainting

Data

- see ComMU Dataset
- #data-augmentation : with transposition and time dilation, yields ×60 data amount

Evaluation

Controllability

- whether metadata information is faithfully adhered to
- determinable via simple algorithms, that heuristically extract information like note density, chords, tempo

Diversity

- Sum of pairwise #cosine-similarity between Chroma Vectors (#harmony) and Groove[4] (#rhythm) across multiple generations from same metadata

Fidelity

- #winrate between generated and real sample (#human-survey)

Results

- Sampling > Temperature controls a between trade-off Controllability and Diversity.
- Sampling > Top-K ([5], [6]) K-value has neglegible effect - unlike with text
  - due to smaller vocab size

Generation from humans

- composers tasked with extracting composing guidlines for specific industrial genres that use Combinatorial Music Generation > Homophonic Composition.
- Composers tasked with composing short MIDI sequences that match this guidlines.
- Composed MIDI sequence and guidline (metadata) converted to REMI-like tokens.

Metadata

- precise tokens: *bpm, key, instrument, time signature, pitch range, number of measure, min/max velocity, and rhythm*
  - tempo constant accross sequence
- ambiguous tokens: *genre, track role* and *chord progression*.
  Comparison to other datasets regarding metadata given below.

On other Datsets

Table 1: Comparison of ComMU to recent MIDI datasets with various metadata. We compare ComMU to other MIDI dataset on 4 types of metadata: genre, instrument, track-role, and chord progression.

Dataset	Genre	Instrument	Track-role	Chord progression
Lakh MIDI [10]	✓	✓	-	-
MAESTRO [12]	✓	(✓)[1]	-	-
MSMD [13]	-	(✓)[1]	-	✓
ADL Piano MIDI [14]	✓	(✓)[1]	-	-
GiantMIDI-Piano [15]	-	(✓)[1]	-	✓
EMOPIA [23]	-	(✓)[1]	-	✓
ComMU(Ours)	✓	✓	✓	✓

[1] include only one instrument.

See Table 1 of ComMU for an overview and Section 2 for a discussion.

- "Lakh MIDI Dataset [10], MAESTRO [12], and ADL Piano MIDI [14] have various meta-information such as genre, instrument, key, and time signature, but there is *no chord information* for the harmony of track-level composition."
- "MSMD [13], GiantMIDI-Piano [15], and EMOPIA [23] have chord information that can infuse harmony, but *lack metadata* such as genre, and rhythm that can reflect the intention of composition."
- separates instruments from track role, other than "[10,28,29]"
  - in Symbolic Music Loop Generation both are the same as well

1. Hyun, Lee, Taehyun Kim, Hyolim Kang, Minjoo Ki, Hyeonchan Hwang, Kwanho Park, Sharang Han, and Seon Joo Kim. 2022. "ComMU: Dataset for Combinatorial Music Generation." arXiv. https://doi.org/10.48550/arXiv.2211.09385. ↩ ↩

Future Work

- to develop specific architectures for the structure of music rather than text

1. Hyun, Lee, Taehyun Kim, Hyolim Kang, Minjoo Ki, Hyeonchan Hwang, Kwanho Park, Sharang Han, and Seon Joo Kim. 2022. "ComMU: Dataset for Combinatorial Music Generation." arXiv. https://doi.org/10.48550/arXiv.2211.09385. ↩
2. Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G Carbonell, Quoc Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pages 2978–2988, 2019. ↩
3. Jeff Ens and Philippe Pasquier. Mmm: Exploring conditional multi-track music generation with the transformer. arXiv preprint arXiv:2008.06048, 2020. ↩
4. Simon Dixon, Fabien Gouyon, Gerhard Widmer, et al. Towards characterisation of music via rhythmic patterns. In International Society for Music Information Retrieval (ISMIR), 2004. ↩
5. Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), pages 889–898, 2018. ↩
6. Ari Holtzman, Jan Buys, Maxwell Forbes, Antoine Bosselut, David Golub, and Yejin Choi. Learning to write with cooperative discriminators. In Annual Meeting of the Association for Computational Linguistics (ACL), 2018. ↩

A Field Guide to
Symbolic Music Representation Learning

* best before next year

A Field Guide to

Symbolic Music Representation Learning

/nai22-project4a

publish.obsidian.md

beijn.github.io