

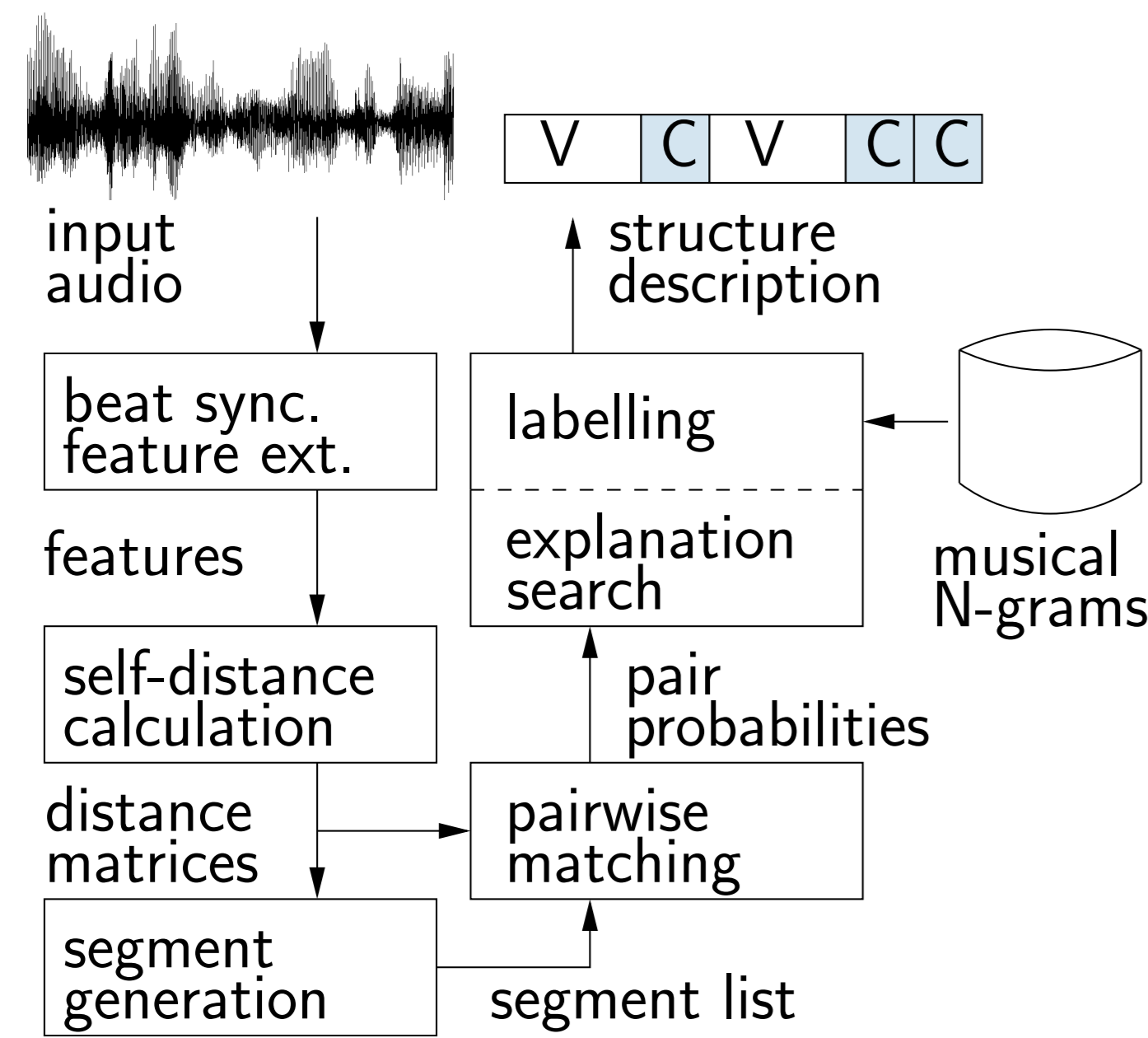
Music Structure Analysis Using a Probabilistic Fitness Measure And an Integrated Musicological Model

Jouni Paulus and Anssi Klapuri

Department of Signal Processing, Tampere University of Technology, Tampere, Finland

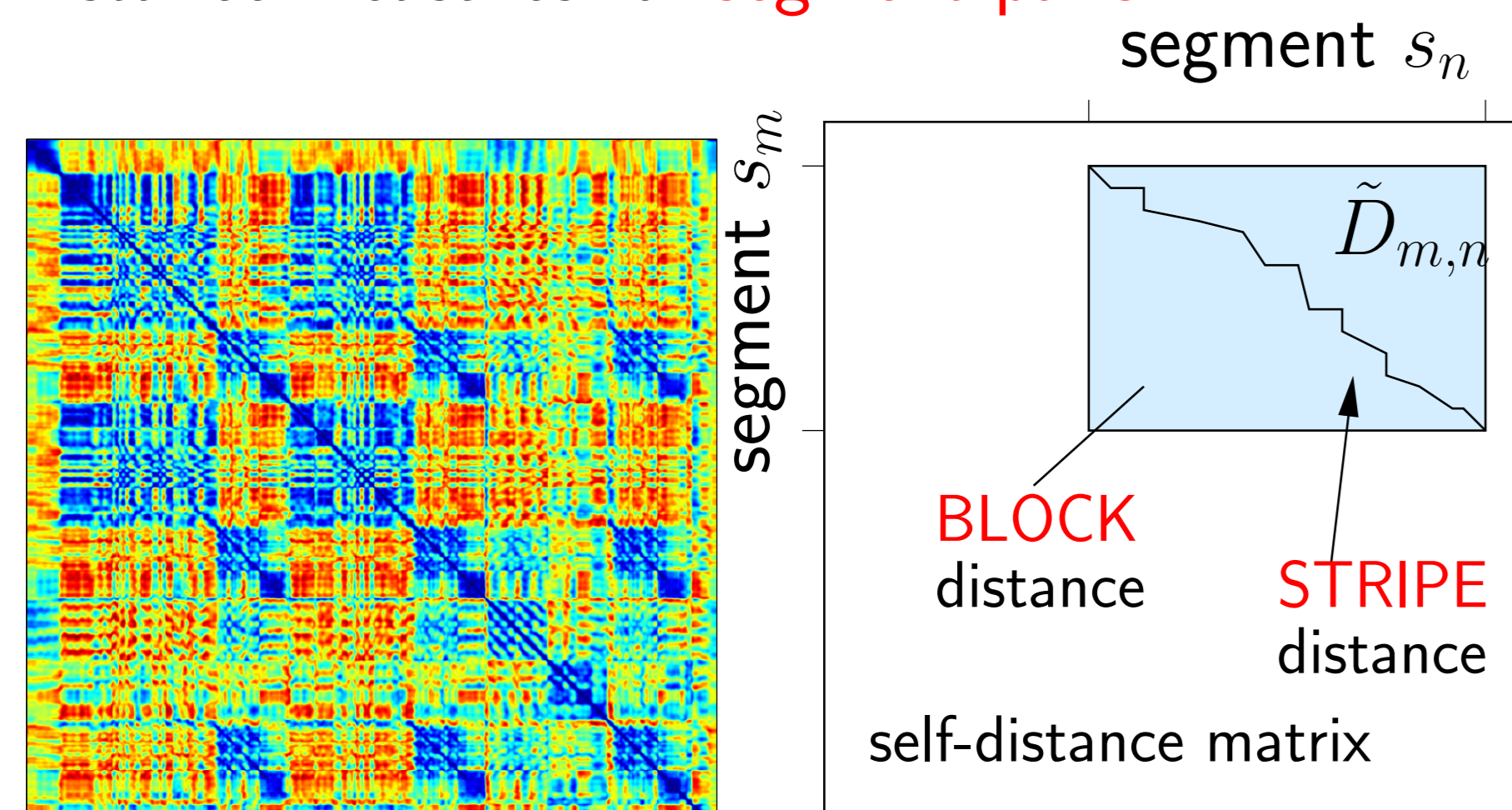
Introduction

- Structure analysis: from audio input
 - find **segmentation** to musical parts (e.g., chorus and verse)
 - **group** segments with similar content, and
 - assign musically meaningful **labels** to groups.

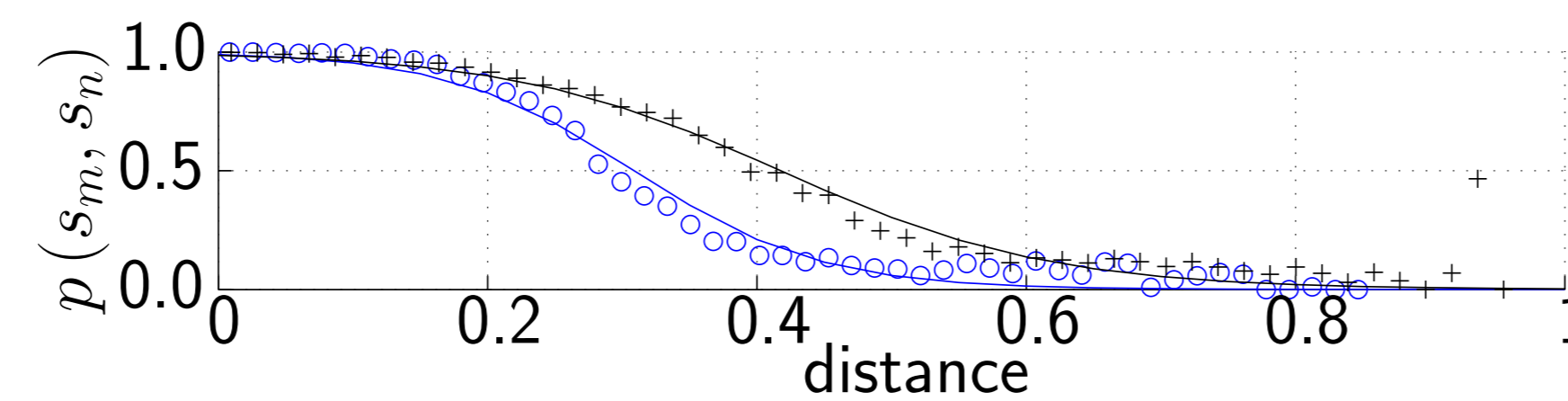


Segment matching

- Three acoustic features for different aspects:
 - general timbre → **MFCCs**,
 - tonal / harmonic content → **chroma**,
 - rhythmic content → **rhythmogram**.
- Self-distance matrix (SDM)
 - Cos-distance between all frame pairs.
- Distance measures for **segment pairs**:



- Map distances to probability that the segments belong to same group. (E.g., blocks, **stripes**.)



Fitness measure

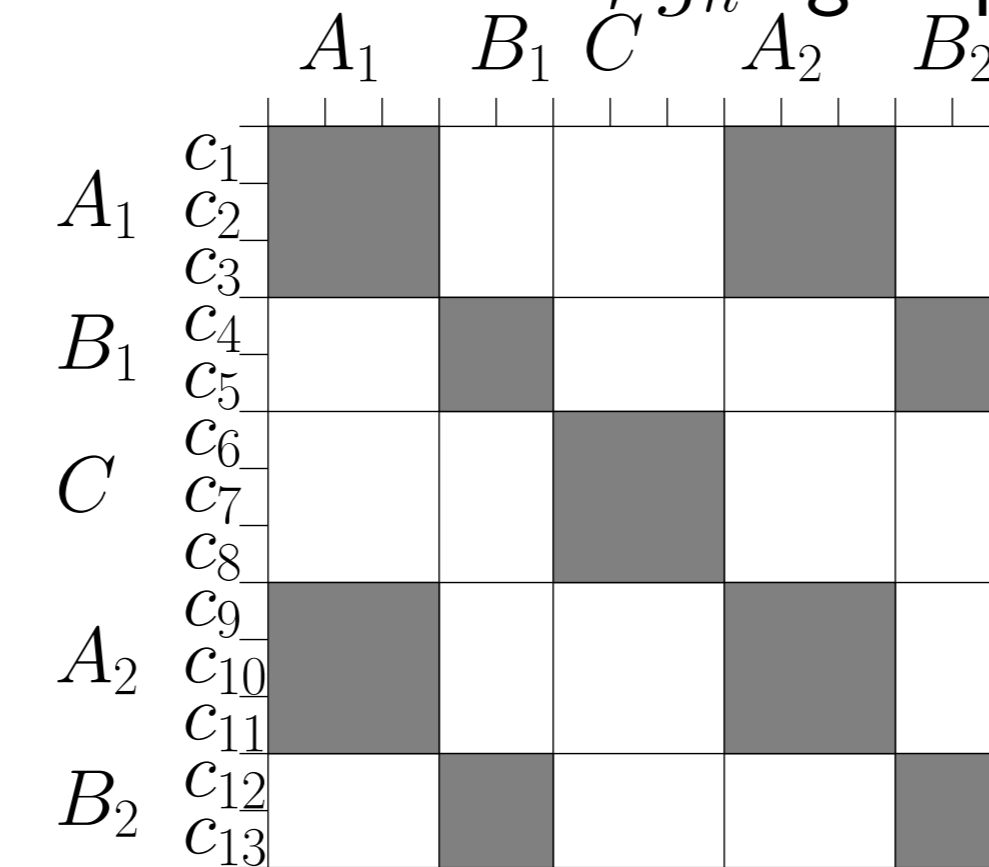
- Find the structural description E maximising

$$P(E) = \sum_{m=1}^M \sum_{n=1}^M A(s_m, s_n) L(s_m, s_n),$$

where

$$L(s_m, s_n) = \begin{cases} \log(\hat{p}(s_m, s_n)), & \text{if } g_m = g_n \\ \log(1 - \hat{p}(s_m, s_n)), & \text{if } g_m \neq g_n \end{cases}$$

$A(s_m, s_n)$: submatrix area, g_n : group of s_n .



Musicological knowledge

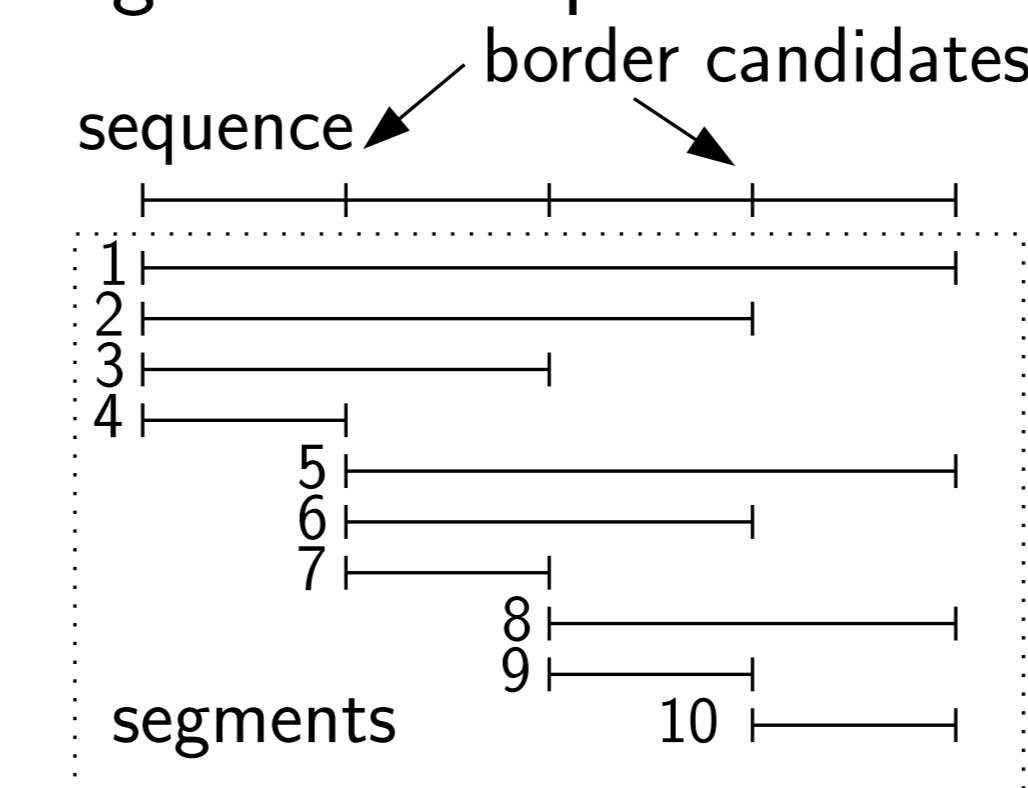
- **N-grams** for musical part sequences proved to contain useful information.
 - Labelling the groups as a post-processing step (presented @ CMMR2008).
- Use information in search by adding a term:

$$P(E) = \sum_{m=1}^M \sum_{n=1}^M A(s_m, s_n) L(s_m, s_n) + \frac{w}{M-1} \sum_{o=1}^M \log(p_N(g_o | g_{1:(o-1)})) \sum_{m=1}^M \sum_{n=1}^M A(s_m, s_n)$$

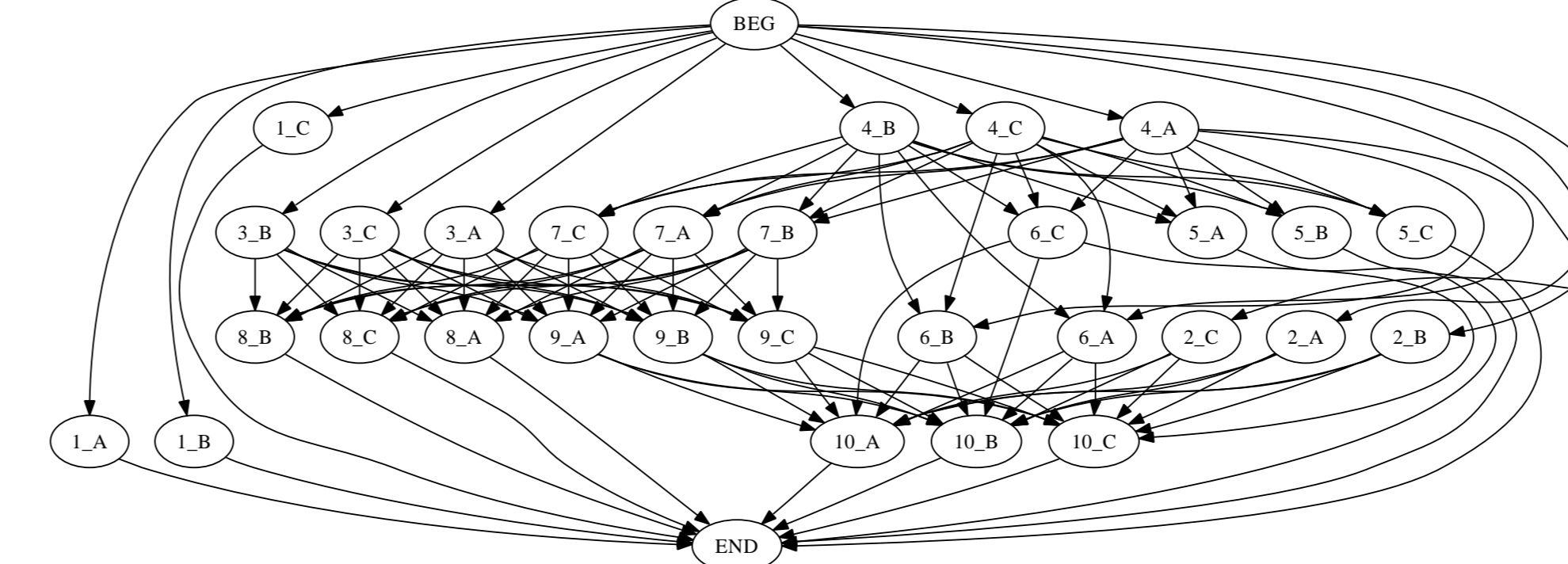
p_N : N-gram probability, w : weighting factor.

Search problem

- **Rapid increase** of search space size as a function of number of segmentation point candidates. E.g.,



Allowing three different groups (A, B, C) produces DAG:



- Find optimal path from BEG to END.
- Problem: state transition costs depend on the **whole** earlier path.

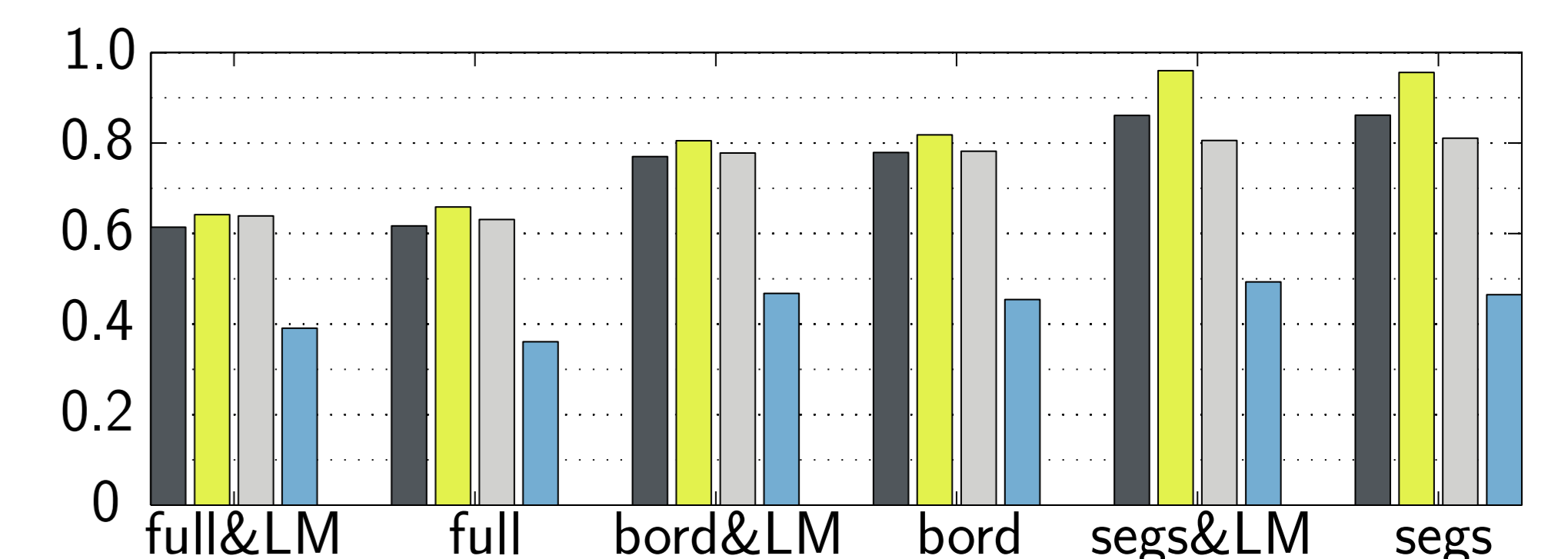
Bubble Token Passing

- Each segment & group combination is a state.
- States contain an ordered buffer of tokens. At each iteration
 - arriving tokens are inserted to the buffer, and
 - the N best tokens are propagated and removed from the buffer.
- Tokens store travelled state sequence.
- Tokens arriving to end state contain found structure descriptions.
- Operation parametrised by number of propagated tokens and maximum number of stored tokens.

- **Controllably greedy**.
- Finds a solution quickly, iterations increase search scope and may produce better solutions.
- Store all tokens and run until all tokens have arrived to end state → exhaustive search.

Experiments

- Evaluations with 557 manually annotated popular music pieces, TUTstructure07.
- Evaluation metrics on frame-by-frame basis:
 - Correct segmentation and grouping of frame pairs, F-measure, **precision rate**, **recall rate**.
 - Correct musical **label** to frame.



Conclusions:

- Probabilistic fitness function for music structure descriptions.
- Fitness function optimisation presented as a graph search.
- A novel greedy search algorithm presented.
- The effect of using musical part N-grams in the fitness function studied.
- Musicological model has very small effect on frame-by-frame grouping result.
- Musicological model improves result when evaluated by correctly labelled frames (compared to post-process labelling).