

 $\{S_2, S_3\}$ 

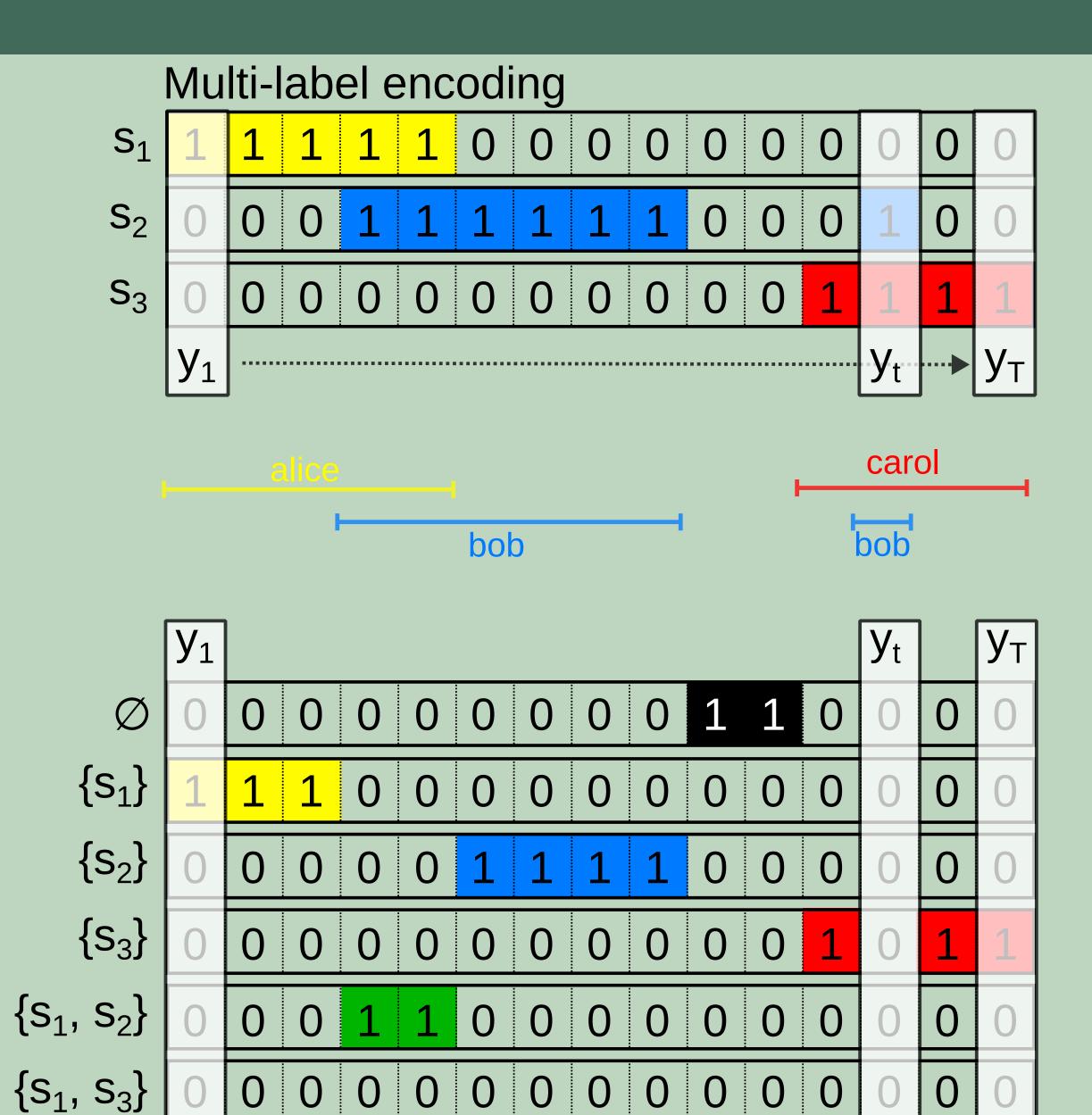
# Powerset multi-class cross entropy loss for neural speaker diarization

Alexis Plaquet<sup>1,2</sup> Hervé Bredin<sup>1,2,3</sup>

<sup>1</sup>Université Toulouse III Paul Sabatier <sup>2</sup>IRIT <sup>3</sup>CNRS



## Summary



Powerset multi-class encoding

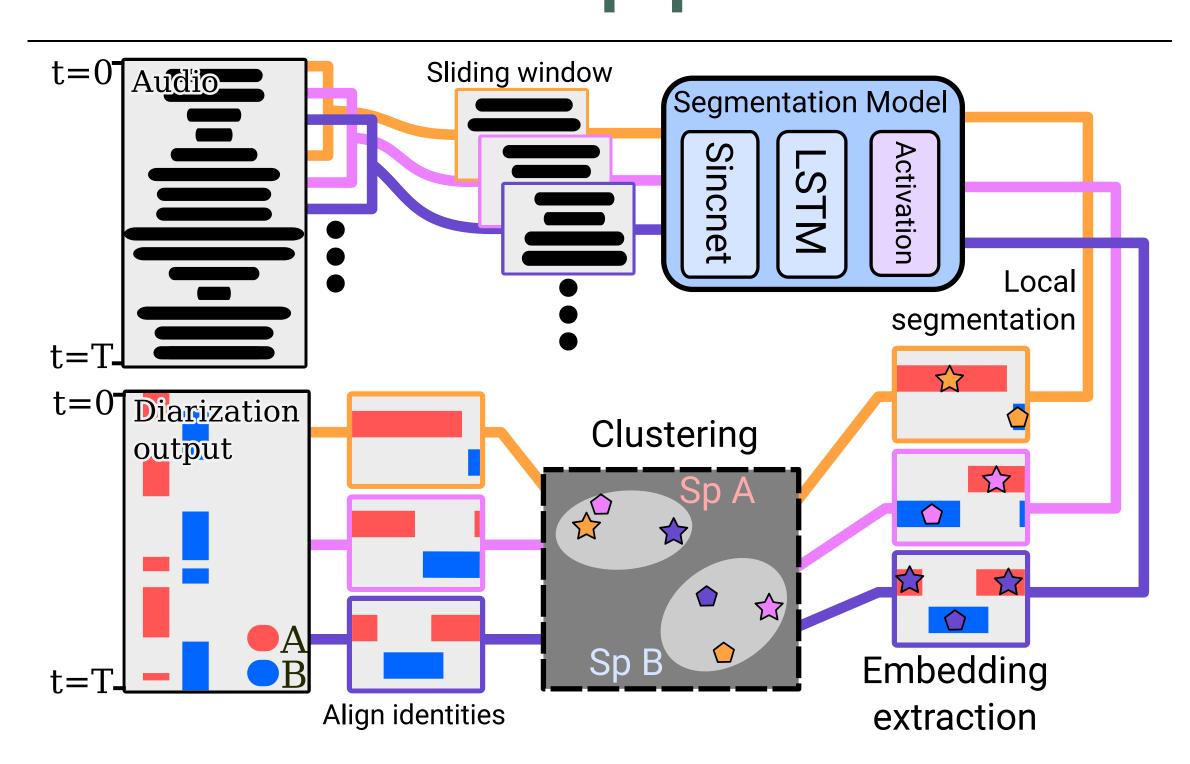
Reframing a multilabel problem as multiclass.

- Powerset speaker diarization provides a consistent boost in Diarization Error Rate (DER), an average
  - 8% in-domain relative improvement
  - 11% out-of-domain relative improvement (on DIHARD)
- Removes the binarization threshold hyperparameter.
- Domain adaptation with powerset provides an additional 8% relative improvement, and has a 6% relative improvement over multilabel domain adaptation.
- Results on 8 in-domain datasets, as well as tests on all 11 DIHARD domains as out-of-domain data.
- No additional computational cost, and straightforward implementation.

## (Powerset) Speaker Diarization

#### "Who speaks when?" Speaker ASpeaker B Output $\dot{t}=0$ $t=T \dot{t}=0$ t=TBinarization Threshold Output Output tensor (multilabel) Segmentation Model Speaker ASpeaker B LSTM 0 Output tensor (multiclass) t=0Segmentation Model ● B ● A+B t=TPowerset spea argmax LSTM diarization

# The full pipeline



### Results

