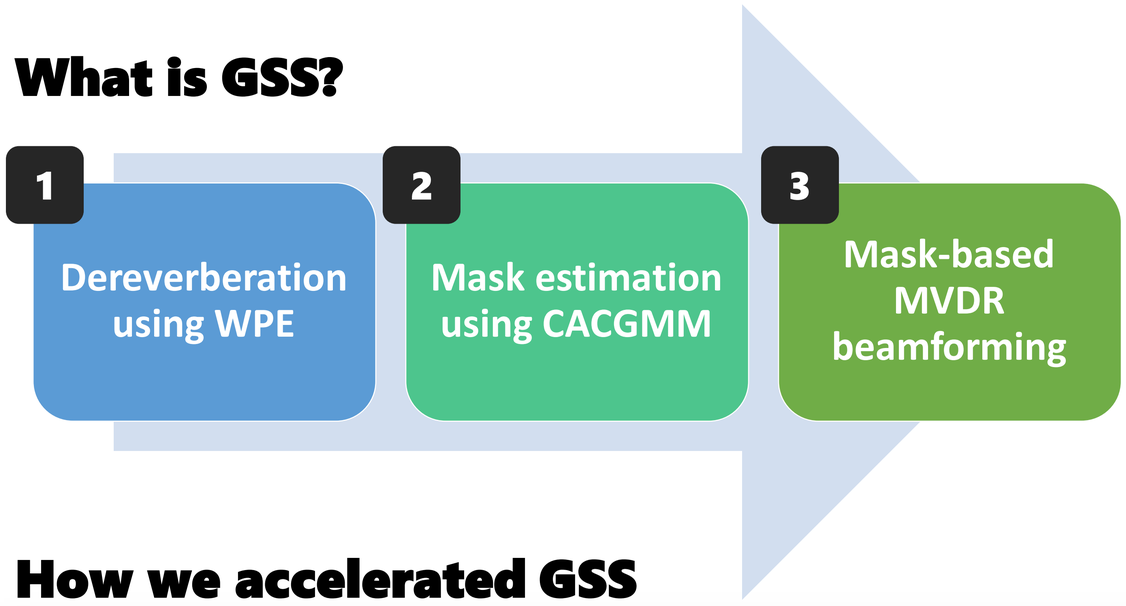
Robust front-end for CHiME-6. Now ~300x faster!

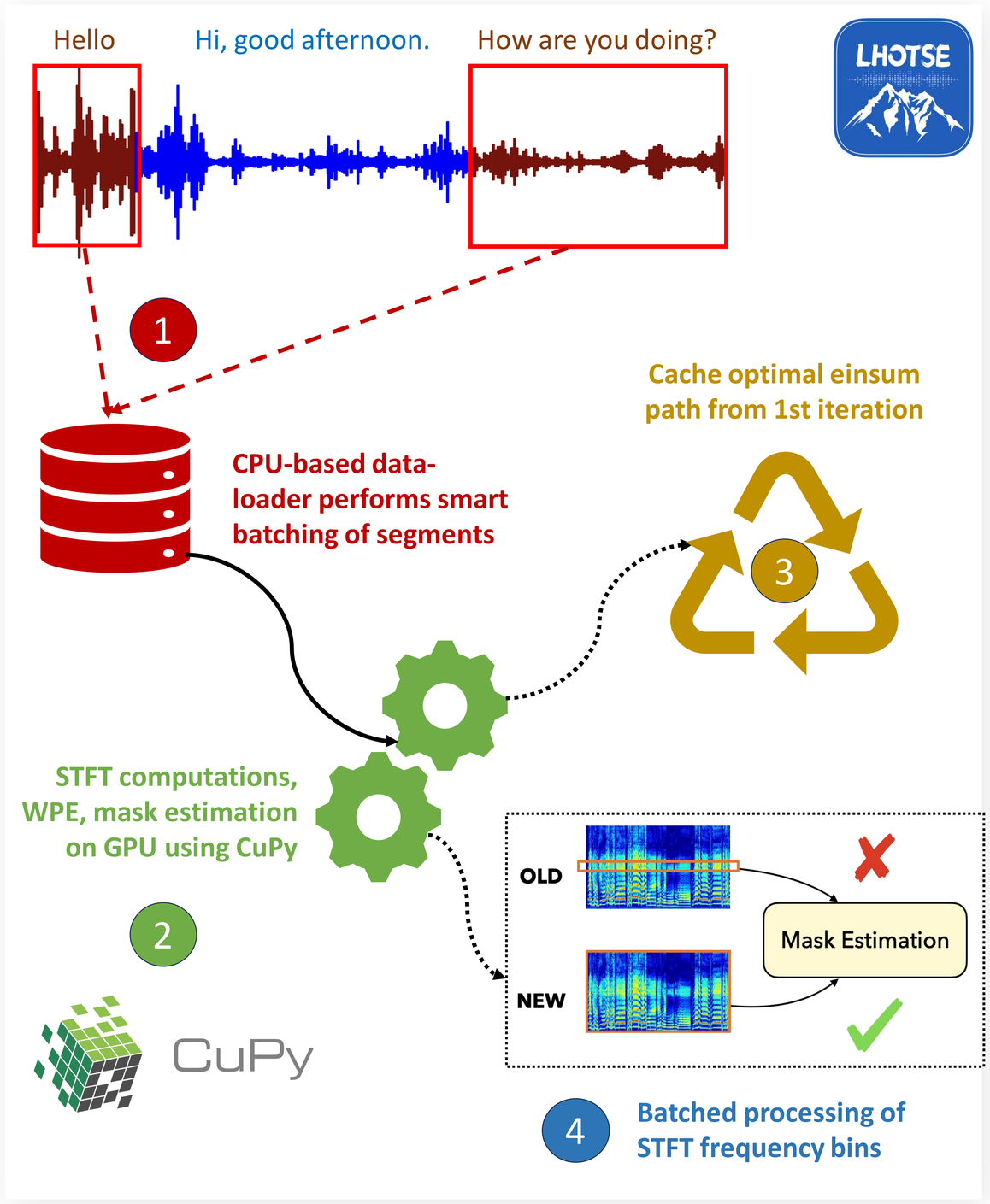
"GPU-accelerated Guided Source Separation for Meeting Transcription"

Background: GSS is an **unsupervised, multi-channel** target-speaker extraction method first proposed for CHiME-5 (Boeddeker et al.) and provided ~10% absolute WER reduction. We present an improved implementation which uses GPU-acceleration and modern data pipelines to make it significantly faster!

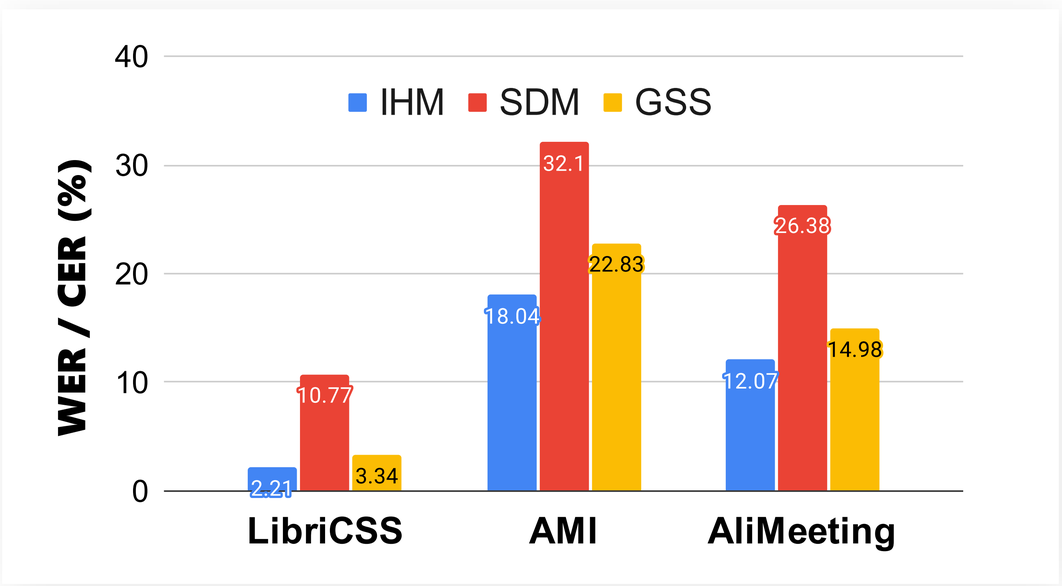
What is GSS?



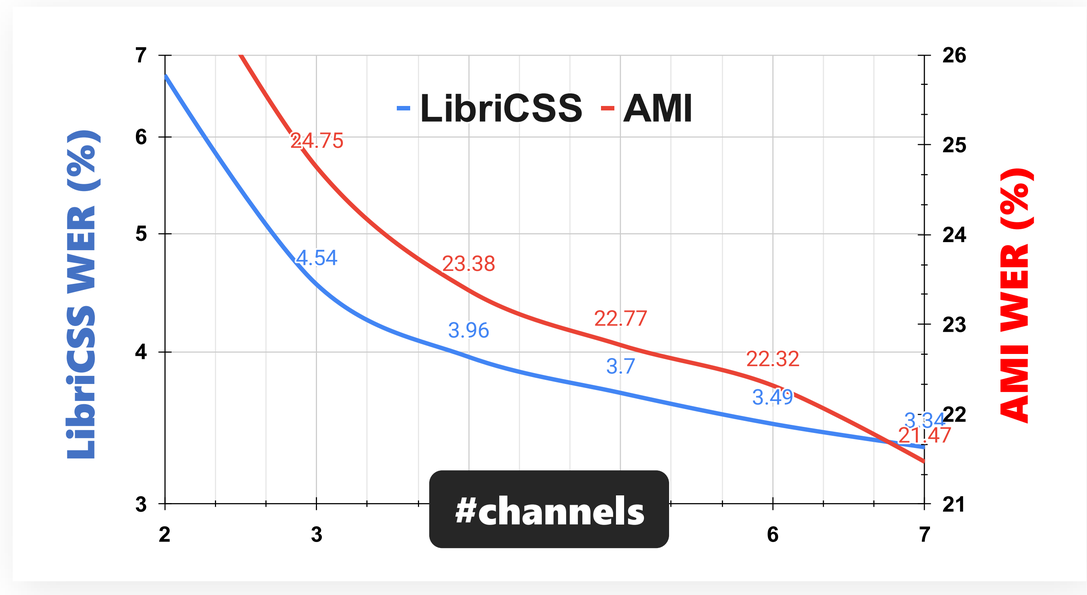
How we accelerated GSS



Take-away #1 GSS can recover up to 80% of the WER gap between close-talk and far-field ASR.



Take-away #2 More input channels give better GSS performance, but with diminishing returns.



Easy-to-use for your favorite datasets!

```
#!/bin/bash

# 1. Create Lhotse manifests for corpus
lhotse prepare libricss --type mdm /export/libricss data/

# 2. Prepare recording-level cuts
lhotse cut simple -r data/recordings.jsonl.gz -s data/supervisions.jsonl.gz exp/cuts.jsonl.gz

# 3. Prepare segment-level cuts
lhotse cut trim-to-supervisions --discard-overlapping exp/cuts.jsonl.gz exp/segments.jsonl.gz

# 4. Perform enhancement
gss enhance cuts --max-batch-duration 50.0 exp/cuts.jsonl.gz exp/segments.jsonl.gz exp/enhanced
```

Further reading: Check out the paper for results using overlap-aware diarization, analysis of GSS hyperparameters, including context size, number of iterations, and noise class, and more.

Boeddeker, Christoph et al. "Front-end processing for the CHiME-5 dinner party scenario." *CHiME 2018*.

Desh Raj, Daniel Povey, Sanjeev Khudanpur

JOHNS HOPKINS UNIVERSITY