

State Space Models and RISC-V Co-design for Energy-Efficient Audio Applications

Presented by
Jorge Luis Vega Pinto · PhD Student, SDU Microelectronics
 Principal supervisor: Farshad Moradi, PhD · Co-supervisors: Hooman Farkhani, PhD and Yasser Rezaeiyan, PhD

INTRODUCTION

Always-on audio applications—such as hearing devices, wearables and embedded acoustic sensors—must process long, noisy streams in real time under strict energy and memory constraints. This project investigates selective state-space models (SSMs) as an efficient alternative to attention-based architectures and co-designs their core operations with a low-power RISC-V system. The aim is accurate, causal audio intelligence with lower latency, memory use and energy consumption.

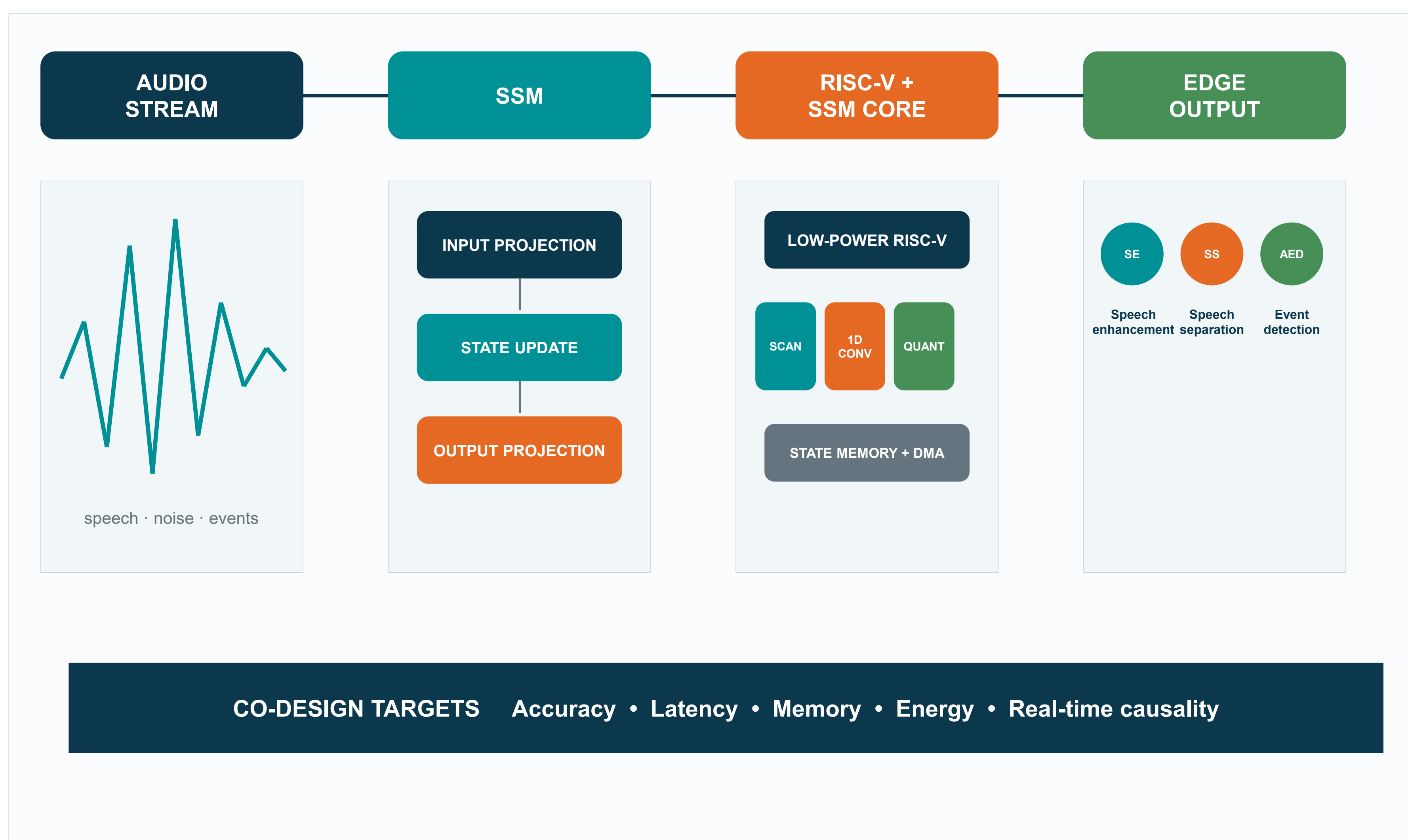


Figure 1. Hardware–algorithm co-design workflow from streaming audio to efficient edge inference.

RESEARCH OBJECTIVES

- Develop causal S4-, Mamba- and neuromorphic-inspired models for edge audio.
- Benchmark accuracy and efficiency against CNN, RNN and Transformer baselines.
- Compress models through quantization, pruning and knowledge distillation.
- Design an SSM accelerator tightly coupled to a low-power RISC-V processor.

CORE HYPOTHESIS

Hardware-aware SSMs can preserve long-range temporal modelling while improving the accuracy–efficiency trade-off for always-on audio.

BACKGROUND

Transformers provide strong sequence modelling but their attention cost grows with sequence length. SSMs replace explicit attention with recurrent state updates that can be executed in streaming mode. Their structure is promising for embedded audio, yet efficient deployment requires joint optimization of model architecture, numerical precision, memory movement and hardware datapaths.

AUDIO TASKS

- Speech enhancement and noise cancellation.
- Speech separation.
- Audio-event detection.

METHODOLOGY

1. MODEL DEVELOPMENT

Train causal S4, diagonal SSM, Mamba and spiking variants with multi-timescale memory.

2. BENCHMARKING

Evaluate PESQ, STOI, SI-SDR, model size, MACs, latency and memory on public audio datasets.

3. HARDWARE CO-DESIGN

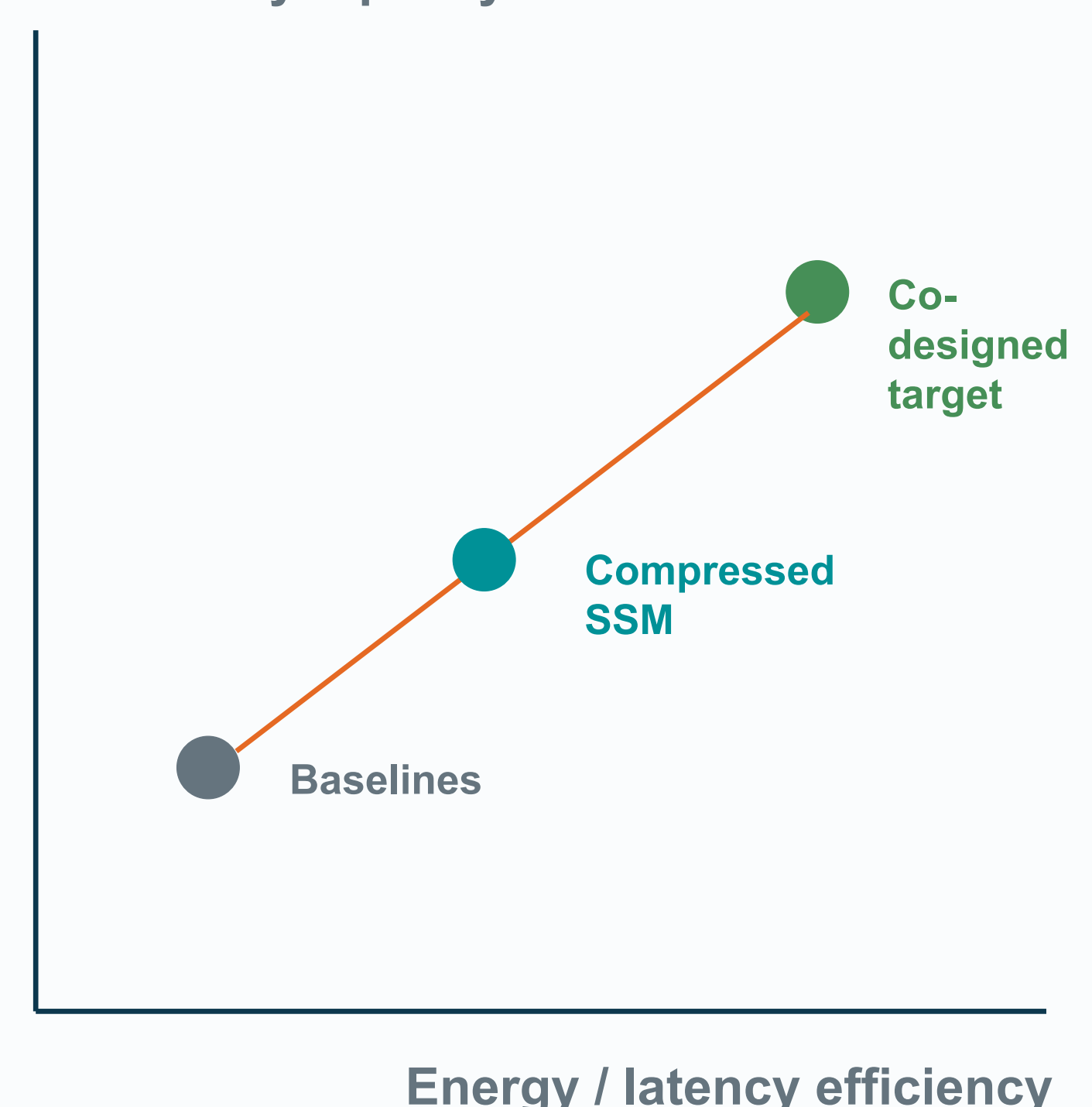
Implement scan, 1D-convolution and state-memory units in RTL; integrate with RISC-V and validate by simulation and FPGA prototyping.

EXPECTED OUTCOMES

A reproducible edge-audio benchmark, compressed causal SSMs, and a reusable RISC-V-coupled SSM-core with measured accuracy–latency–energy trade-offs.

EVALUATION FRAMEWORK

Accuracy / quality



Conceptual target — final positions will be determined experimentally.