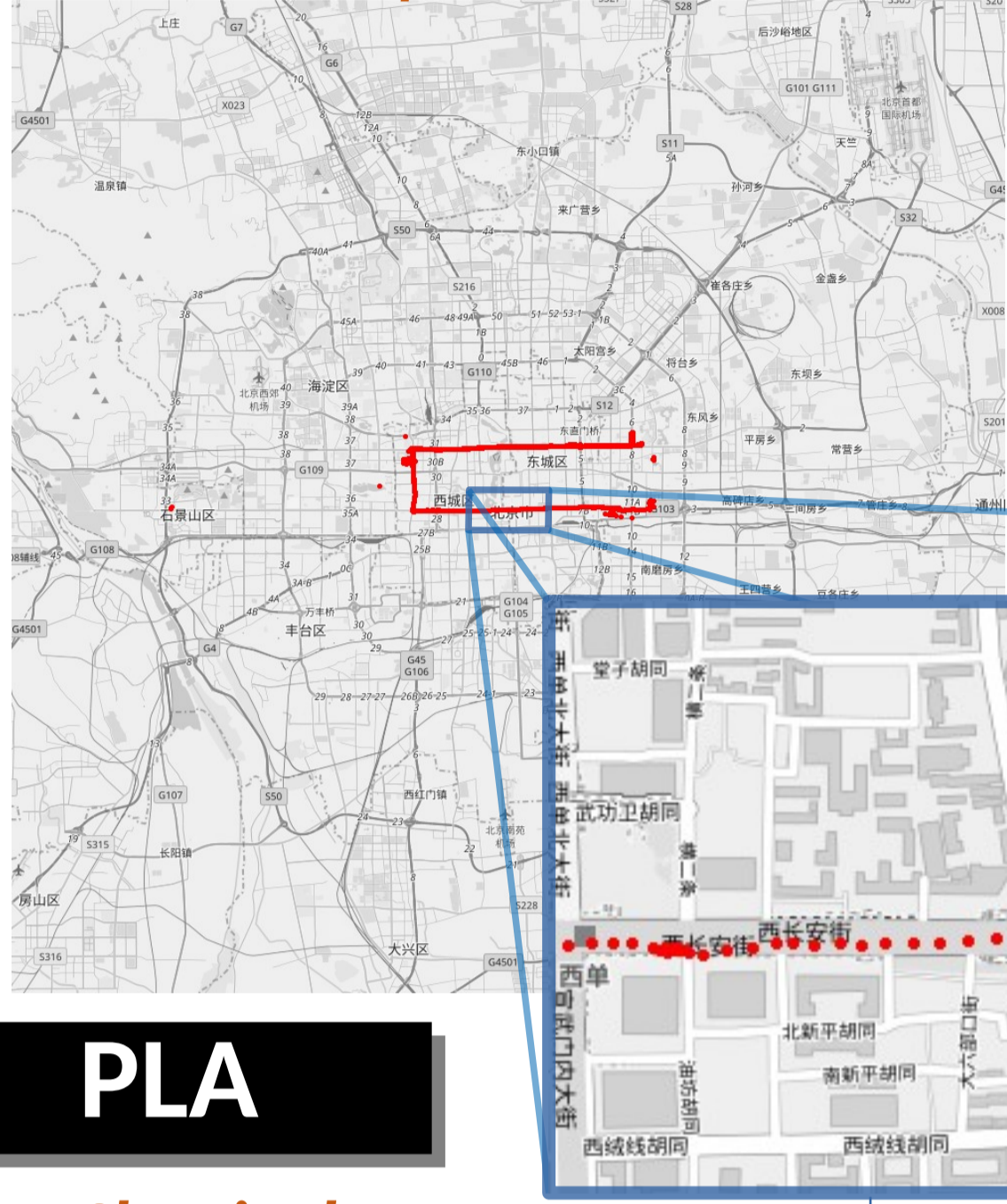


# Streaming Piecewise Linear Approximation for Efficient Data Management in Edge Computing

## INTRODUCTION: PIECEWISE LINEAR APPROXIMATION (PLA) AT THE EDGE

Example: GPS Traces.

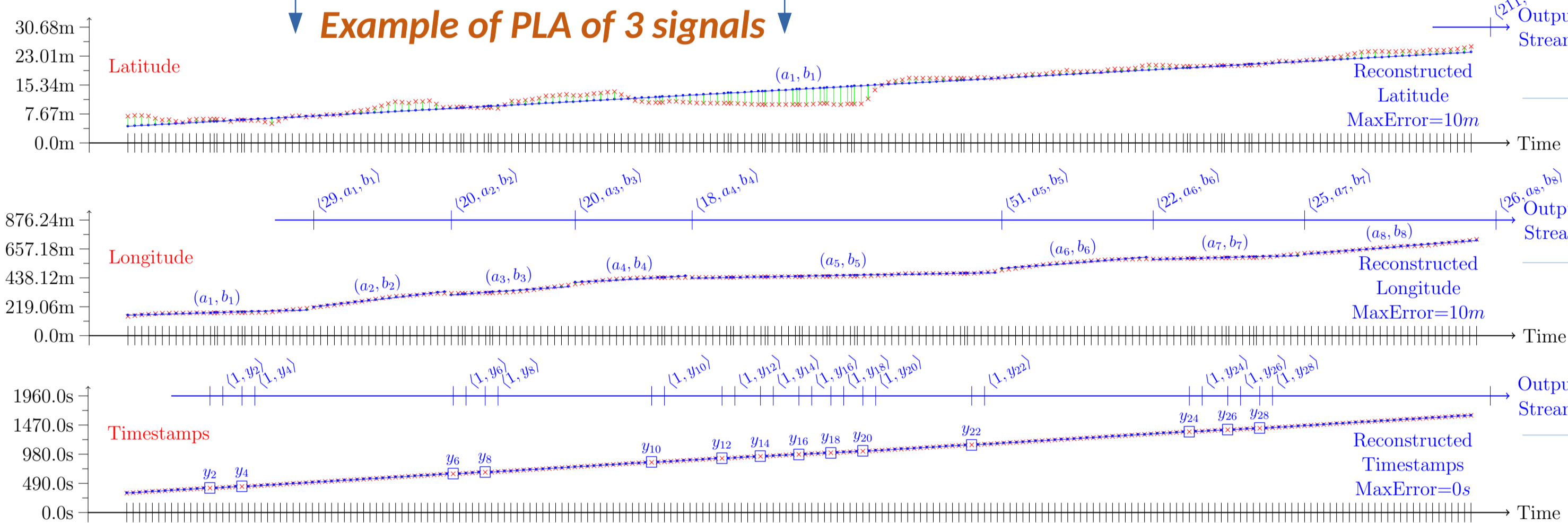


- **Context:** Big Data challenges due to large volume of data produced at high rate by small devices in large Cyber-Physical Systems (as smart grids, homes, cars, ...).
- **Idea:** Use low-overhead compression mechanism (PLA) to facilitate data retrieval and maintenance in large IoT systems; Aim at reducing both **bandwidth** and **storage** consumption at the edge by forwarding/storing only the PLA of the stream.
- **PLA:** Replace consecutive input values by *line segments* approximating them within some predefined *error bound*.

➤ **Contribution:** Streaming PLA improving different trade-offs in our context.

### PLA

- A Classical problem.
- Previous works: focus on the # of segments.
- Greedy approach: continue till it's not possible (optimal).

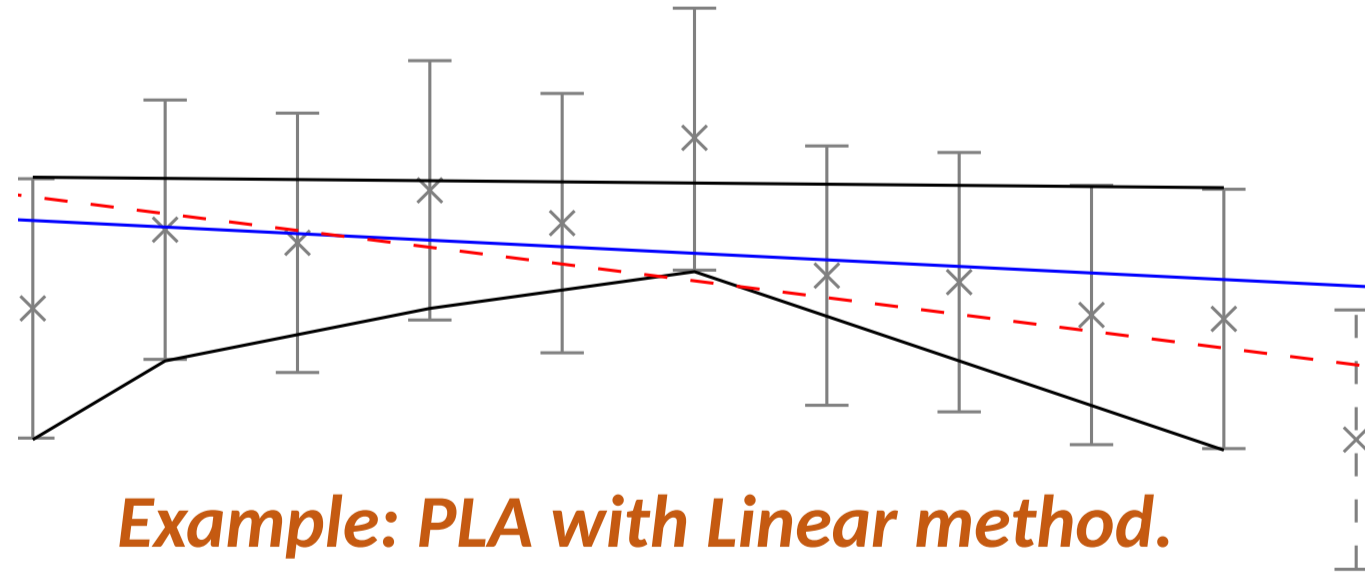


### TRADE-OFFS

Compression (# bytes ratio)	Avg Latency (# of tuples)	Error (average)
1%	106	1m
8%	16	3.4m
23%	11.3	0s

## OUR NEW METHODS

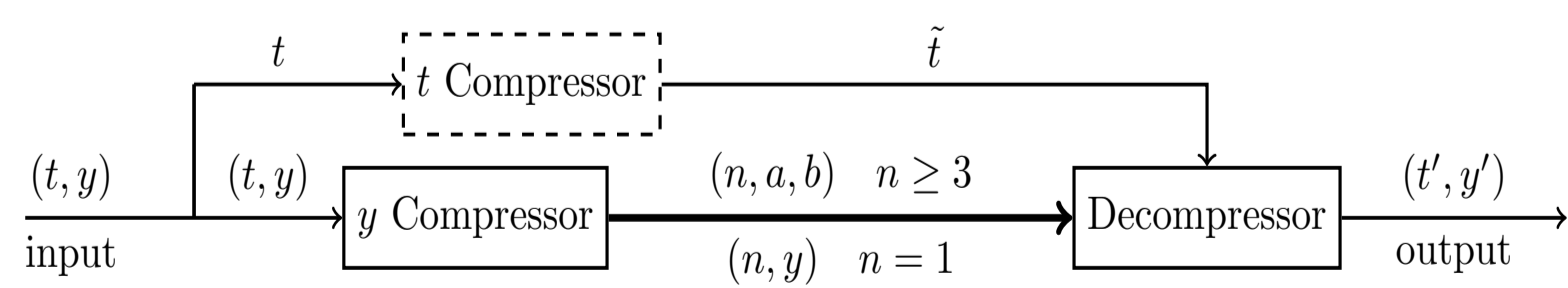
- **Linear PLA method:** combining several known approaches (best-fit line approximation and convex-hull maintenance for error checking).
- *Lightweight, fast, and reduces errors.*



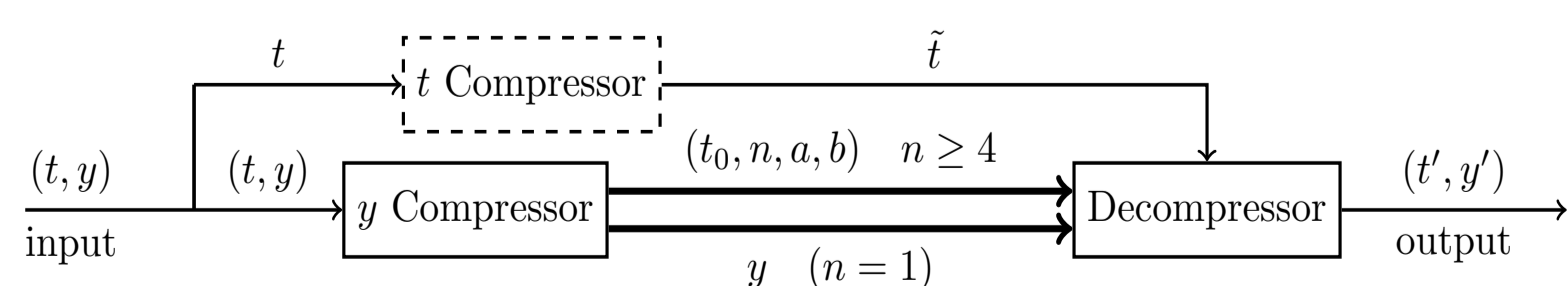
Example: PLA with Linear method.

- **New PLA protocols:** bounds segments' length and introduce singletons (isolated values).

### Our Protocol Flowcharts:



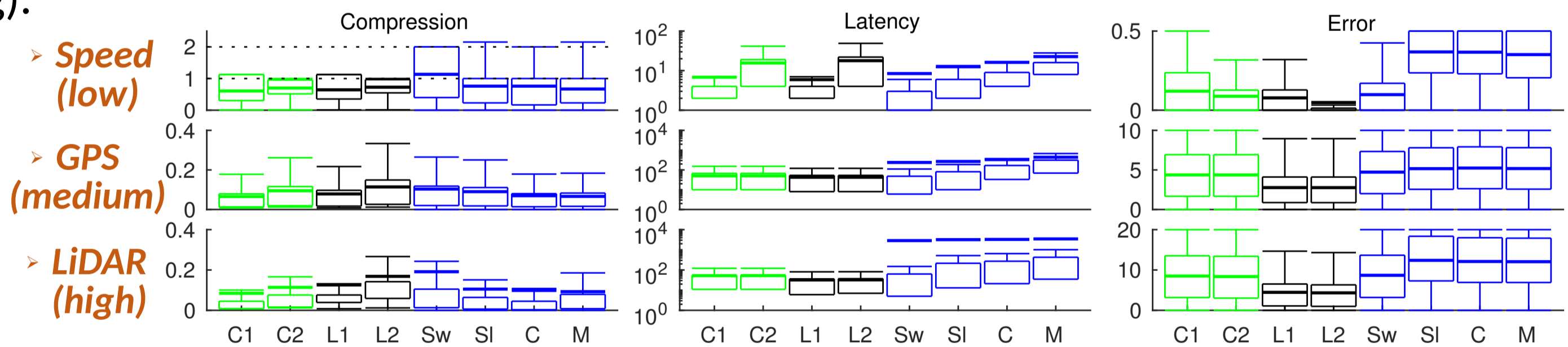
➤ **One-Stream:** best compression & latency.



➤ **Two-Streams:** never inflates input data.

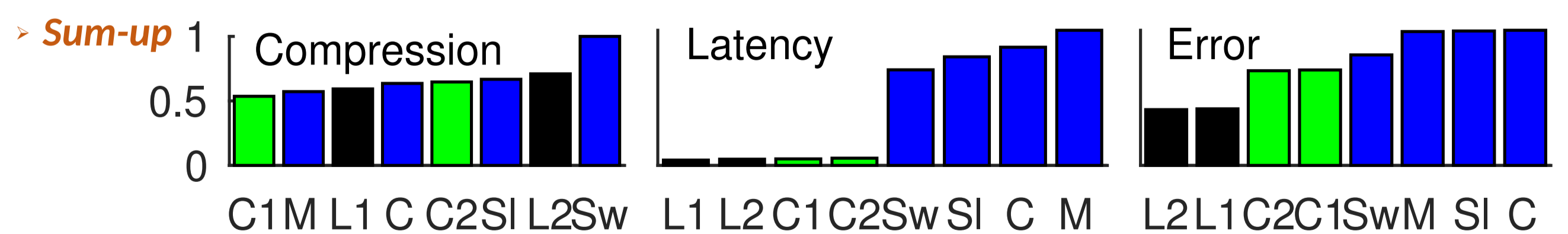
## EVALUATION

- **Extensive evaluation:** 3 large datasets (19k GPS traces 0.5Gb, 145k LiDAR scans 2.8Gb and 7k 76 days of Speed measurements 2.3 Gb).



Streaming statistics over the 3 datasets with following methods & protocols:

- **C1/C2:** known PLA method (Convex Hull which produces longest possible segments) coupled with One-Stream (1) or Two-Stream (2).
- **L1/L2:** Linear PLA method with One/Two-Stream protocol.
- **State-of-the-art:** SwingFilter (**Sw**) a simple greedy heuristic; SlideFilter (**SI**) equivalent to Convex Hull; Continuous (**C**) best continuous solution (segments share endpoints); Mixed-PLA (**M**) mixed-approach optimizing output size.



➤ **Conclusions:**

- **Two-Streams** is the only safe protocol (never inflates data).
- **Our protocols** achieve 1-2 orders of magnitude better latencies.
- **Linear method** achieves far smaller errors for a small trade in compression and produces comparable latencies.

