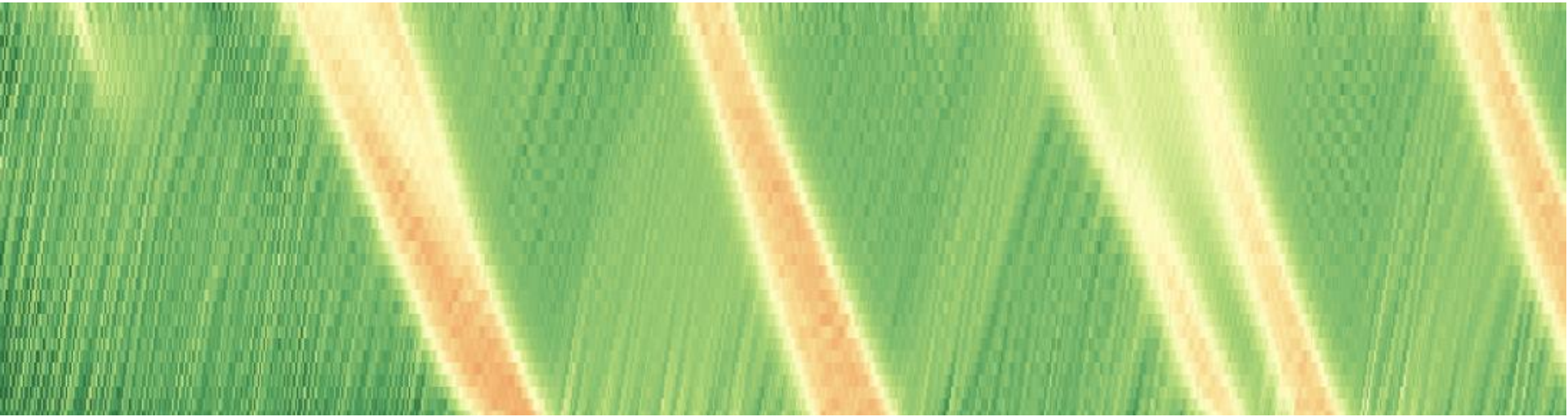
Supervised vs. unsupervised learning for hyperbolic PDEs: application to traffic forecast



Professor Alexandre Bayen, EECS
Associate Provost – Berkeley Space Center
Director, Center for Information Technology in the Interest of Society

Advanced Modeling & Simulation (AMS) Seminar Series
NASA Ames Research Center, September 25th, 2025

Problem statement

We are interested in the hyperbolic PDE [Ole63]:

$$(P): \begin{cases} \partial_t u + \partial_x \phi(t,x,u) = 0, \quad (t,x) \in \mathcal{U}_T \overset{\text{def}}{=} [0,T] \times \mathbb{R} \\ u(t=0,\cdot) = u_0(\cdot) \end{cases}$$

- We are concerned with the solution $u(t,\cdot) \in \mathbf{BV} \cap L^1(\mathbb{R})$, for $t \in [0,T]$. Here, $u \in \mathbf{BV}(\mathbb{R})$ (of bounded variation) if

$$T.V.(u(t,\cdot)) \overset{\text{def}}{=} \sup_{\mathcal{P}} \sum \{|u(t,x_{n+1}) - u(t,x_n)| : x_n \in \mathcal{P}\} < \infty,$$

  where the sup is taken over all possible partitions $\mathcal{P}$.
- $L^\infty(\mathbb{R})$ solutions, measure solutions [DiP85], and semigroup solutions [BC95]

$$\begin{cases} S_t : L^1(\mathbb{R}) \to L^1(\mathbb{R}) \\ u_\circ \mapsto u(t,\cdot) \overset{\text{def}}{=} S_t(u_\circ)(\cdot) \end{cases}$$

  are also studied in the literature under different assumptions for ϕ.
- The generalisation to $x \in \mathbb{R}^n$ with a source term is considered in [Kru70].

OA Oleinik. "Discontinuous solutions of nonlinear differential equations". In: *Amer. Math. Soc. Transl* 26.2 (1963), pp. 95–172.
Ronald J DiPerna. "Measure-valued solutions to conservation laws". In: *Archive for Rational Mechanics and Analysis* 88 (1985), pp. 223–270.
Alberto Bressan and Rinaldo M Colombo. "The semigroup generated by 2× 2 conservation laws". In: *Archive for rational mechanics and analysis* 133 (1995), pp. 1–75.
Stanislav N Krukov. "First order quasilinear equations in several independent variables". In: *Mathematics of the USSR-Sbornik* 10.2 (1970), p. 217.

Definition of the solution

A function $u \in C([0,T]; L^1_{\text{loc}} \cap \mathbf{BV}(\mathbb{R}; [0, u_{\max}]))$ is a **generalised solution** of Eq. (P) if:

$$\forall \varphi \in C_c^1((-\infty, T] \times \mathbb{R}), \quad \int_0^T \int_{\mathbb{R}} \{u \partial_t \varphi + \psi(u) \partial_x \varphi\} + \int_{\mathbb{R}} u_\circ \varphi(0, \cdot) = 0.$$

In addition, $u \in L^\infty([0,T] \times \mathbb{R})$ is called the **Kružkov entropy solution** if for any $k \in \mathbb{R}$, and $\varphi \in C_c^1((-\infty, T] \times \mathbb{R}; \mathbb{R}_+)$,

$$\begin{aligned} \mathcal{R}(u, \varphi, k) \overset{\text{def}}{=} & \int_0^T \int_{\mathbb{R}} |u - k| \, \partial_t \varphi + \int_0^T \int_{\mathbb{R}} \operatorname{sign}(u - k)[\phi(\cdot, \cdot, u) - \phi(\cdot, \cdot, k)] \partial_x \varphi(\cdot, \cdot) \\ & - \int_0^T \int_{\mathbb{R}} \operatorname{sign}(u - k) \partial_x \phi(\cdot, \cdot, u) \varphi + \int_{\mathbb{R}} |u_\circ - k| \, \varphi(0, \cdot) \geq 0. \end{aligned}$$

Well-posedness (Kruzkov, 1970)

Theorem (Existence of the entropic solution). Let an initial condition $u_\circ \in L^\infty(\mathbb{R})$ and assume:

(**H-1**): ϕ, $\partial_u\phi$, $\partial_{ux}\phi$, $\partial_{xx}\phi$ are continuous, and $\partial_u\phi, \partial_x\phi, \partial_{ux}\phi \in L^\infty(\mathcal{U}_T^M;\mathbb{R})$ for any $M > 0$, where $\mathcal{U}_T^M \overset{\text{def}}{=} [0,T]\times\mathbb{R}\times[-M,M]$.

(**H-2**): $\partial_{xu}\phi \in L^\infty(\mathcal{U}_T^M;\mathbb{R})$ and $\int_0^T \int_{\mathbb{R}} \|\partial_x\phi(t,x,\cdot)\|_{L^\infty([-M,M])}\,\mathrm{d}x\,\mathrm{d}t < +\infty$.

Then there exists a unique weak entropy solution

$$u \in L^\infty([0,T];L^1_{\text{loc}}(\mathbb{R};\mathbb{R})).$$

In addition, if $u_\circ \in L^\infty \cap \mathbf{BV}(\mathbb{R};\mathbb{R})$, then the weak entropy solution

$$u \in L^\infty([0,T];L^1 \cap \mathbf{BV}(\mathbb{R};\mathbb{R})).$$

Proof. *Follows from the singular limit of the corresponding viscosity problem*: letting $\varepsilon \to 0$, the solution of the viscous (parabolic) problem

$$\partial_t u + \partial_x[\phi(t,x,u)] = \varepsilon\partial_{xx}u$$

converges to the entropy solution.

Stanislav N Krukov. "First order quasilinear equations in several independent variables". In: *Mathematics of the USSR-Sbornik* 10.2 (1970), p. 217.

Possible generalization to broader class of hyperbolic PDEs

Objective: Problem-specific NN solvers, focused on **hyperbolic** equations.

Motivations:

- Various key applications:
  - Fluid dynamics
  - Gas dynamics
  - Traffic dynamics
- Even when degenerate diffusions are added (applications: fluid dynamics, relativistic heat conduction, plasma physics, traffic flow, ...):

$$\partial_t u + \partial_x [uV(u) - \kappa D(u)\Phi(\partial_x u)] = 0.$$

  Where:
  - $D(u)$: a (nonlinear) diffusion coefficient,
  - $\Phi(\cdot)$: a strictly increasing, bounded, smooth function (e.g. $\tanh(\cdot)$).

  The system still exhibits hyperbolic behaviour (the advection dominates) [Cas13; BD92].

V. Caselles. "Flux limited generalized porous media diffusion equations". In: *Publicacions Matemàtiques* 57.1 (2013), pp. 144–217.
Michiel Bertsch and Roberta Dal Passo. "Hyperbolic phenomena in a strongly degenerate parabolic equation". In: *Archive for rational mechanics and analysis* 117 (1992).

Main goal of the approach

Hyperbolic PDEs:

- **Hyperbolic Property**: Less regular solutions (**BV** class vs. Sobolev class)
- Various existing NN solvers:
  - Perform well on smoother (more regular) solutions.
  - Often inefficient or inaccurate for hyperbolic problems.

We consider a few special classes of flux functions ϕ in (P).

Main goal:

- Show that NN solvers provide **accurate approximations** for hyperbolic PDEs.
- Provide theoretical insight: **convergence to the entropy solution**.

Main contributions

Main contributions:

- A new optimization loss based on the weak formulation (unsupervised).
- A new kind of SOTA, modular, **higher-order (in time)** FVM.

Advantages of the method:

- As easy to implement as any standard FVM.
- Can be trained with or without data (supervised/unsupervised).
- When supervised: no assumptions on the underlying flow.
- **Conservation of mass** is enforced as a hard constraint.
- Easily extensible to larger stencils.
- Can easily incorporate stencils from multiple timesteps.
- Same time complexity as standard FVMs.
- **Significantly more precise than existing FVMs**.

Canonical traffic model (LWR, 1955-1956)

In the context of traffic, the **LWR model** is the reduced form of Eq. (P) to:

$$\partial_t \rho + \partial_x f(\rho) = 0,$$

where:

- $\rho \in L^1([0,T] \times \mathbb{R})$ is the **conserved quantity** and represents **traffic density**,
- $f : \rho \mapsto \rho v(\rho)$ is known as **traffic flux**.

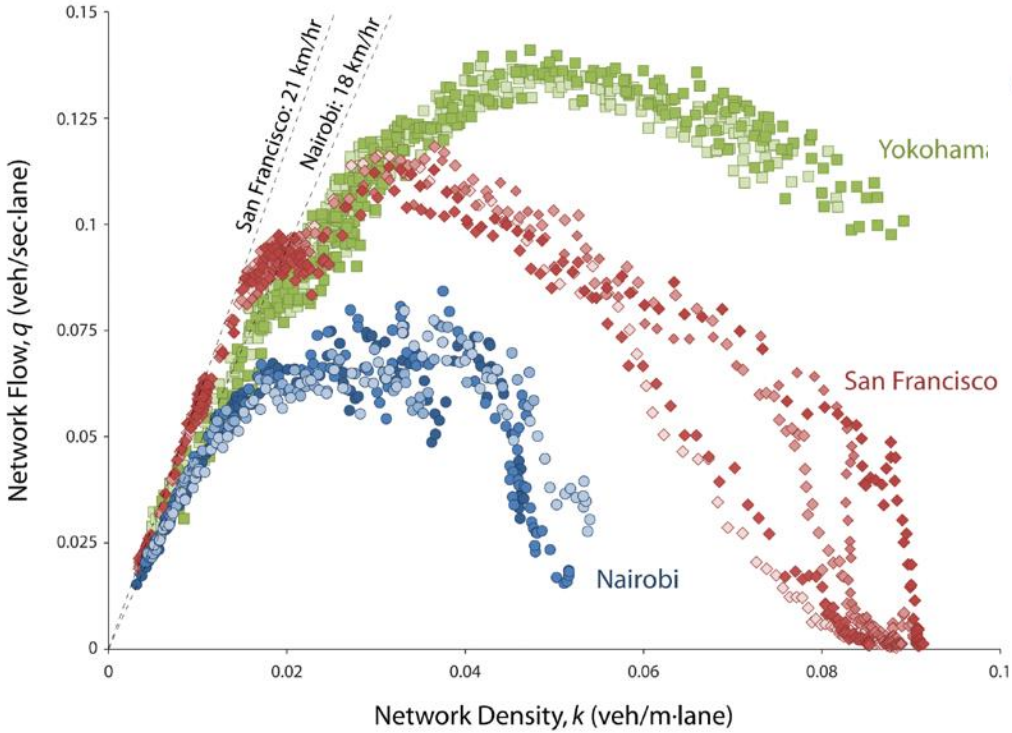
Greenshields Model: Monotonically decreasing and bounded velocity function $v \in C^1([0, \rho_{\max}]; \mathbb{R}_+)$, with $v(0) = v_{\max}$ and $v(\rho_{\max}) = 0$, defined by

$$v : \rho \mapsto v_{\max} \left(1 - \frac{\rho}{\rho_{\max}}\right),$$

which leads to the flux $f \in C^2([0, \rho_{\max}]; \mathbb{R}_+)$, which is genuinely nonlinear (strictly concave in this case) and defined by

$$f : \rho \mapsto \rho v(\rho) = v_{\max} \rho \left(1 - \frac{\rho}{\rho_{\max}}\right).$$

Fundamental diagram fit

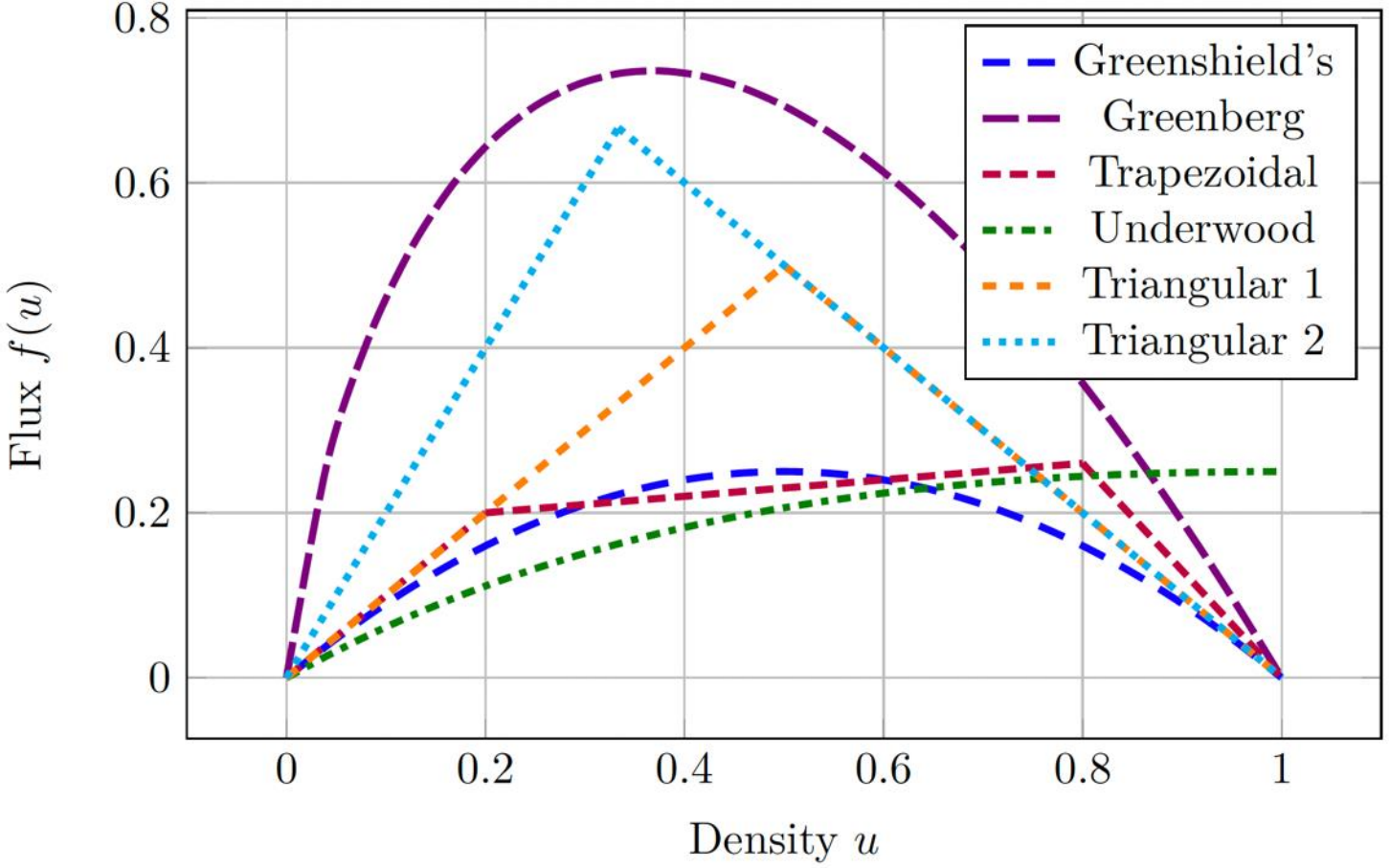


$$\partial_t \rho + \partial_x f(\rho) = 0,$$

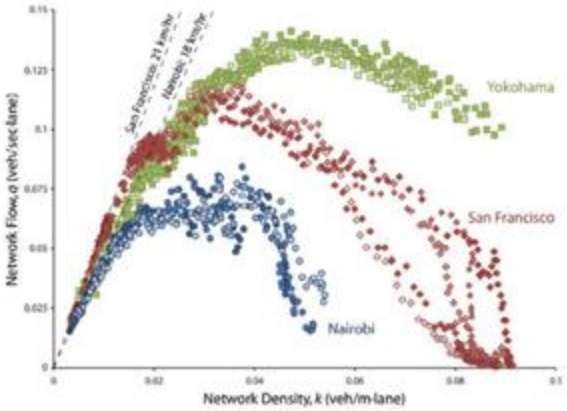
Figure: Network flow vs. density from road data points [AKN16].

Richard Arnott, Anatolii Kokoza, and Mehdi Naji. "Equilibrium traffic dynamics in a bathtub model: A special case". In: *Economics of Transportation* 7-8 (2016), pp. 38–52.

Fundamental diagram fit



Example: traffic: Lighthill Whitham Richards

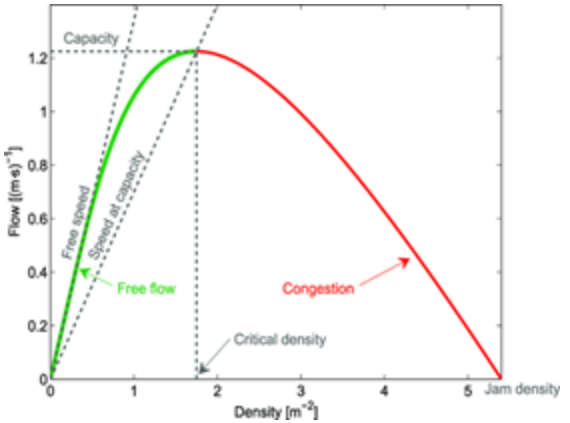


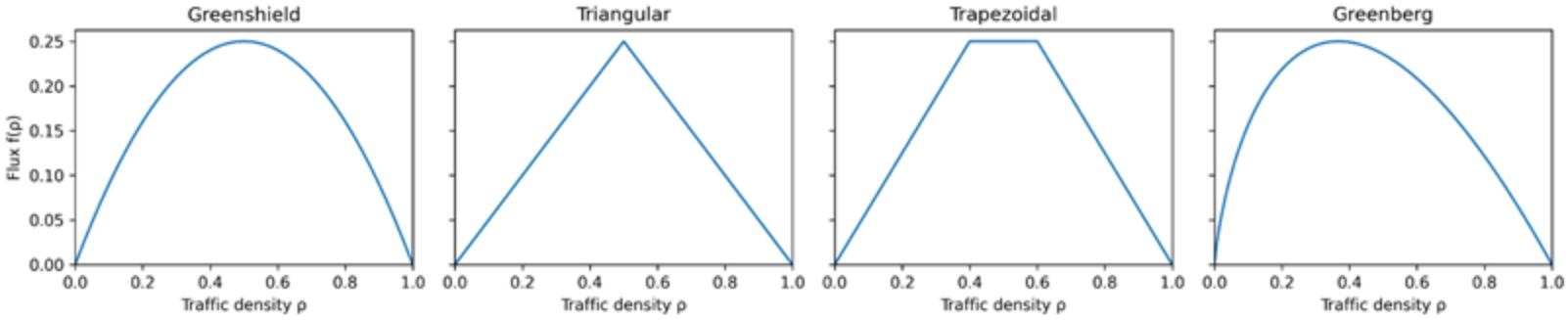
$$q = \rho v$$

q : flow (veh/hr)

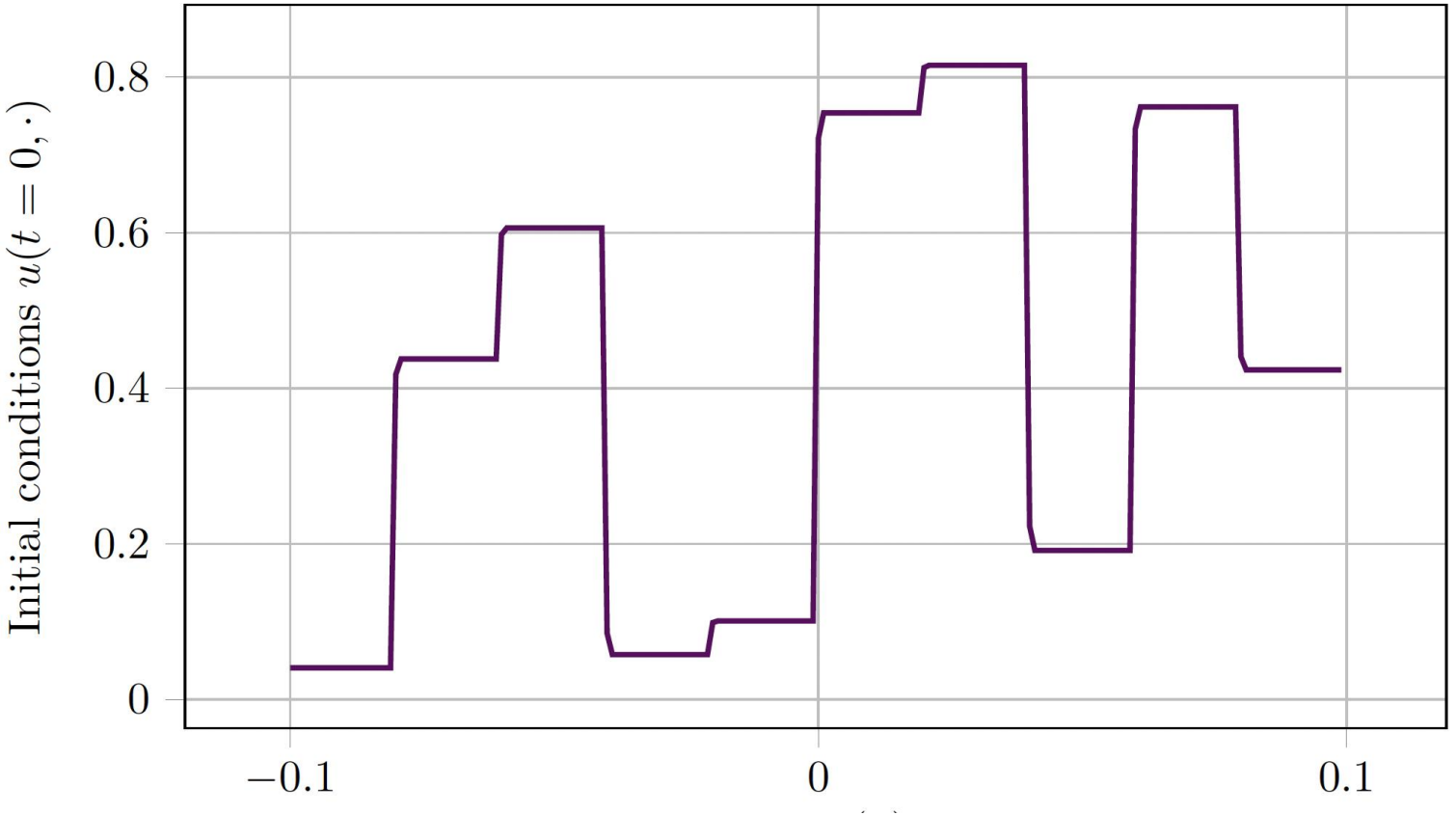
ρ : density (veh/km)

v : speed (km/hr)

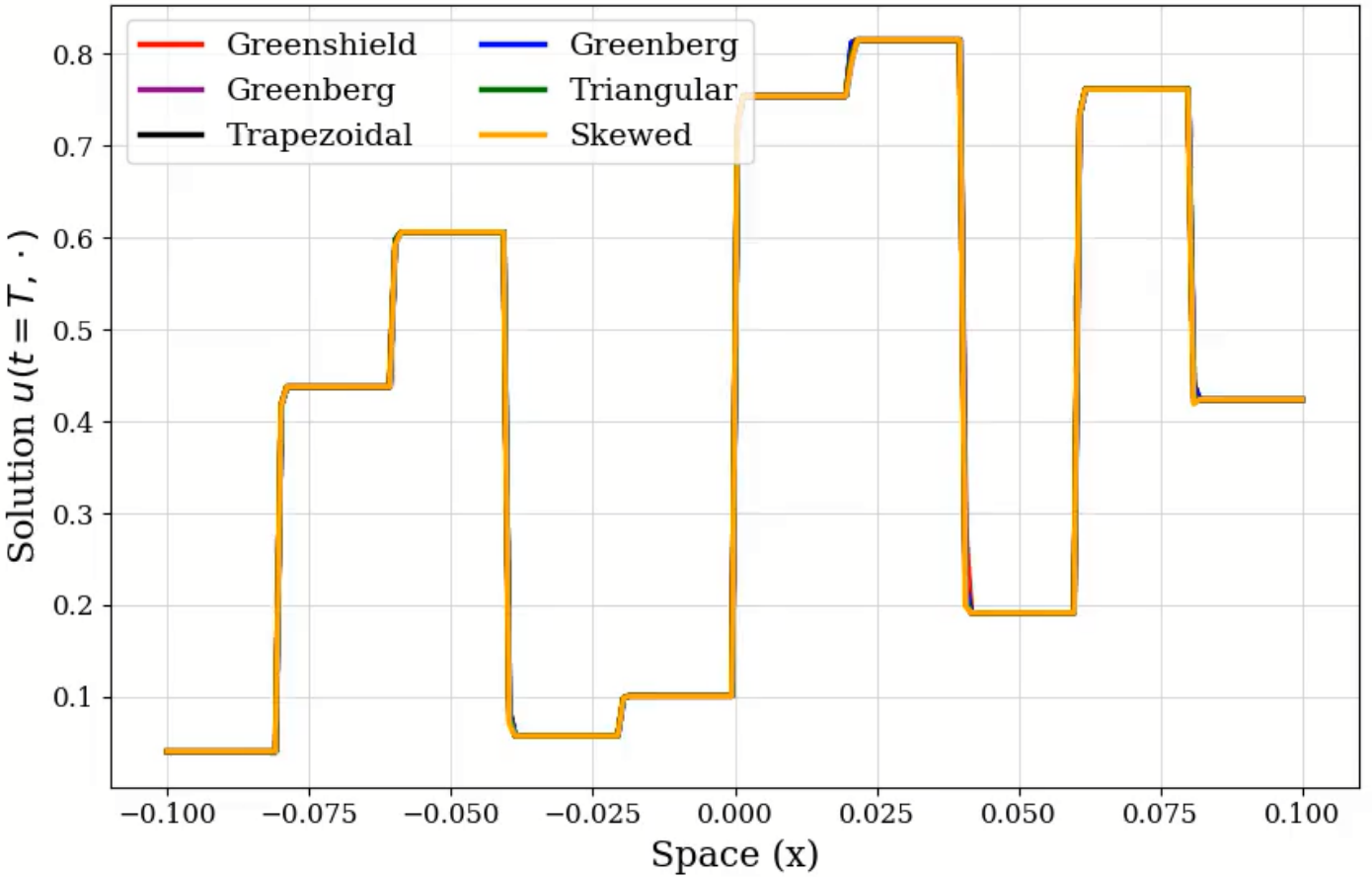




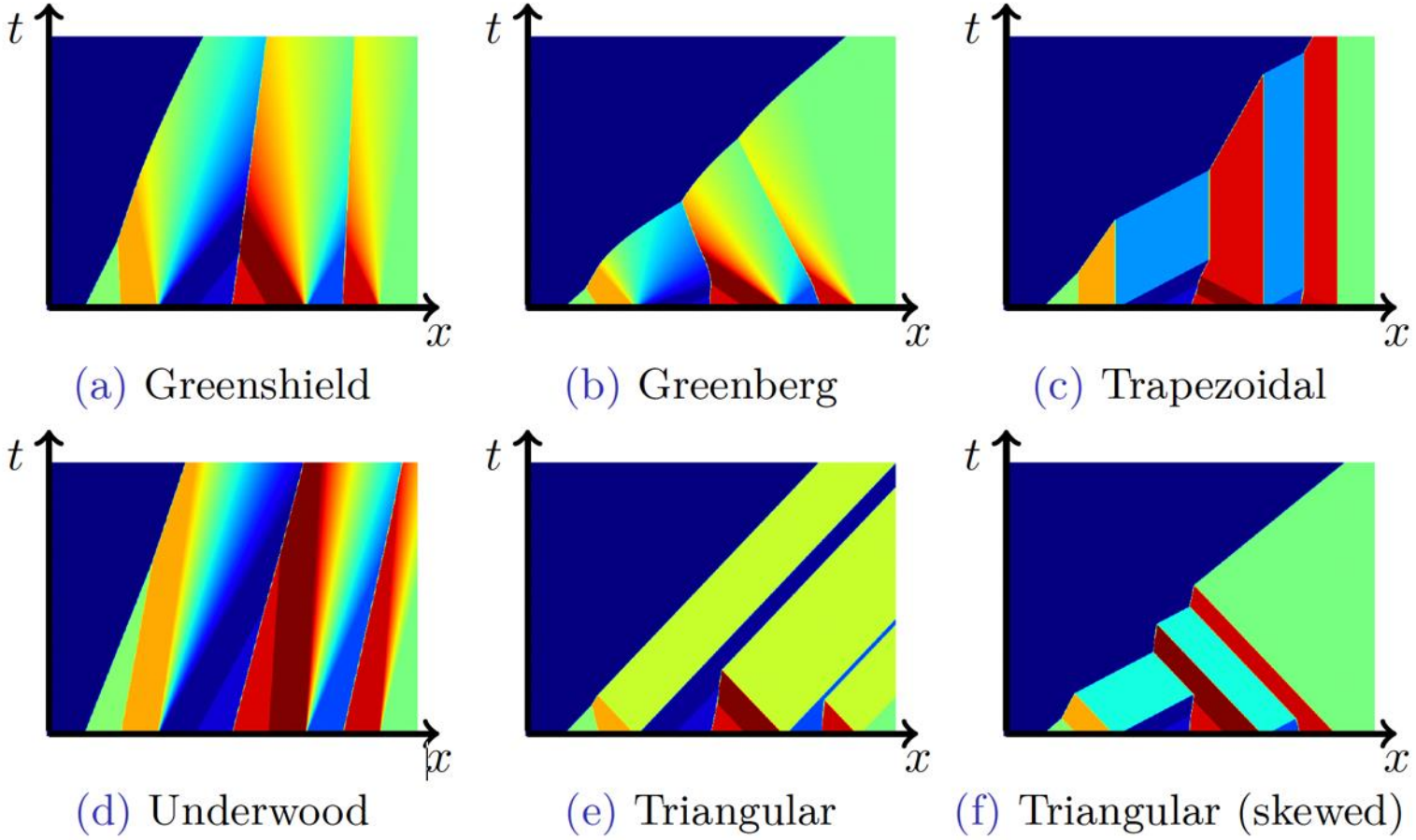
Predictive capabilities of different models



Predictive capabilities of different models



Predictive capabilities of different models



Properties of the solution

Analytical Properties of LWR:

- Let $f \in C^1([0, \rho_{\max}]; \mathbb{R}_+)$, $\rho_\circ \in L^1 \cap L^\infty(\mathbb{R})$, and $T.V.(\rho_\circ) < \infty$.
- Well-posedness conditions are met.
- The solution satisfies: for any $s, t \in [0, T]$, and $x \in \mathbb{R}$

$$\begin{aligned} \|\rho(t) - \rho(s)\|_{L^1(\mathbb{R})} &\leq \boldsymbol{K} \, |t - s| \, \|\rho_\circ\|_{T.V.}, \\ \|\rho(t, \cdot)\|_{L^\infty(\mathbb{R})} &\leq \|\rho_\circ\|_{L^\infty(\mathbb{R})}, \\ \|\rho(t, \cdot)\|_{T.V.(\mathbb{R})} &\leq \|\rho_\circ\|_{T.V.(\mathbb{R})}. \end{aligned}$$

Entropy solution enforcement in NN solver

- The first question: Why not directly impose the entropy condition in the loss function, in the unsupervised case? i.e.,

$$\mathcal{L}(\theta) \overset{\text{def}}{=} \arg\min_{\theta} \max_{\varphi \in W_{\circ}^{1,\infty}([0,T]\times\mathcal{U}),k\in\mathbb{R}} \underbrace{[-\mathcal{R}(\hat{\rho}_\theta,\varphi,k)]}_{\text{residual, entropy condition}} .$$

$$\mathcal{R}(u,\varphi,k) \overset{\text{def}}{=} \int_0^T\int_{\mathbb{R}} |u-k|\,\partial_t\varphi + \int_0^T\int_{\mathbb{R}} \text{sign}(u-k)[\phi(\cdot,\cdot,u)-\phi(\cdot,\cdot,k)]\partial_x\varphi(\cdot,\cdot)$$
$$- \int_0^T\int_{\mathbb{R}} \text{sign}(u-k)\partial_x\phi(\cdot,\cdot,u)\varphi + \int_{\mathbb{R}} |u_\circ - k|\,\varphi(0,\cdot) \geq 0.$$

Entropy solution enforcement in NN solver

- The first question: Why not directly impose the entropy condition in the loss function, in the unsupervised case? i.e.,

$$\mathcal{L}(\theta) \overset{\text{def}}{=} \arg\min_{\theta} \max_{\varphi \in W_{\circ}^{1,\infty}([0,T]\times\mathcal{U}),k\in\mathbb{R}} \underbrace{[-\mathcal{R}(\hat{\rho}_\theta,\varphi,k)]}_{\text{residual, entropy condition}} .$$

- Let $p, q \in [1, +\infty]$ conjugate exponents, ρ the entropy solution. Then for any function $v \in L^q([0,T]\times\mathcal{U})$, we have

$$-\mathcal{R}(v,\varphi,k) \lesssim \|\varphi\|_{W^{1,p}([0,T]\times\mathcal{U})} \|\rho - v\|_{L^q([0,T]\times\mathcal{U})} .$$

Theorem (Global NN-approx). For every $u \in \mathbf{BV}([0,T]\times\mathcal{U})$ and $\varepsilon > 0$, there is a NN solution $\hat{u}$ with $\mathcal{O}(\varepsilon^{-2})$ number of neurons such that

$$\|u - \hat{u}\|_{L^1([0,T]\times\mathcal{U})} \leq \varepsilon.$$

Tim De Ryck, Ameya D Jagtap, and Siddhartha Mishra. "Error estimates for physics-informed neural networks approximating the Navier–Stokes equations". In: *IMA Journal of Numerical Analysis* 44.1 (2024), pp. 83–119.

Entropy solution enforcement in NN solver

Let $\mathcal{S} \overset{\text{def}}{=} \{(t_i, x_i)\}_{i\in\{1,\cdots,M\}} \subset [0,T]\times\mathbb{R}$. For fixed $\theta \in \Theta$, φ and $k \in \mathbb{R}$, the training error is denoted by $\mathcal{E}_T(\theta, \mathcal{S}, \varphi, k)$ and

$$\mathcal{E}_T(\theta^*_{\mathcal{S}}, \mathcal{S}, \varphi^*_{\mathcal{S}}, k^*_{\mathcal{S}}) \approx \min_{\theta}\max_{\varphi}\max_{k} \mathcal{E}_T(\theta, \mathcal{S}, \varphi, k).$$

which is defined based on the NN-solution $\hat{\rho}_{\theta^*}$. In addition, let

$$\mathcal{E}(\theta) \overset{\text{def}}{=} \int_{\mathbb{R}} |\hat{\rho}_\theta(T,x) - \rho(T,x)|\, dx.$$

Theorem [DJM24]. With probability of at least $1-\delta$, we have that

$$\mathcal{E}(\theta^*_{\mathcal{S}}) \lesssim \mathcal{E}_T(\theta^*_{\mathcal{S}}, \mathcal{S}, \varphi^*_{\mathcal{S}}, k^*_{\mathcal{S}}) + \boldsymbol{K}_1\sqrt{\ln\left(\frac{\boldsymbol{K}_2}{\delta\varepsilon^3}\right)} + (1+\|\hat{\rho}_{\theta^*}\|_{C^1})\ln(\varepsilon^{-3})\varepsilon$$

where the constants $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ depends on M (number of samples), L (number of hidden layers) and W (width).

Tim De Ryck, Ameya D Jagtap, and Siddhartha Mishra. "Error estimates for physics-informed neural networks approximating the Navier–Stokes equations". In: *IMA Journal of Numerical Analysis* 44.1 (2024), pp. 83–119.

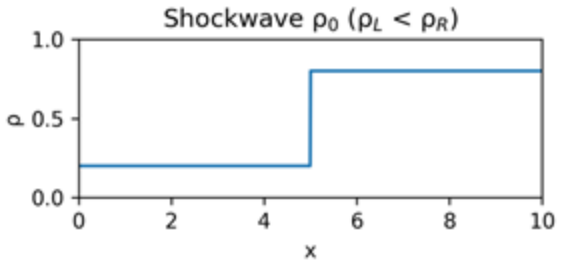
Imposing entropy solutions

Imposing Entropy Condition:

- Theoretical convergence to the entropy solution is guaranteed.
- In practice, however, the problem becomes a min-max optimization problem.
- Computationally inefficient and prone to getting trapped in locally-optimal solutions, leading to oscillatory behaviour.

Building blocks for supervised learning: Riemann problems

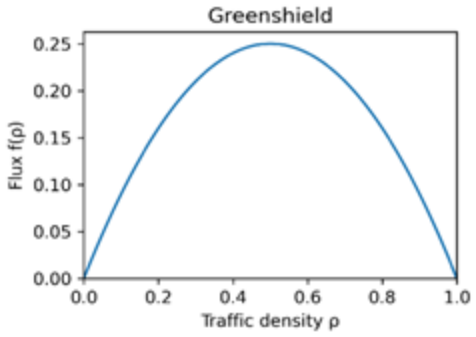
”Fundamental” problems for this PDE: Riemann probems

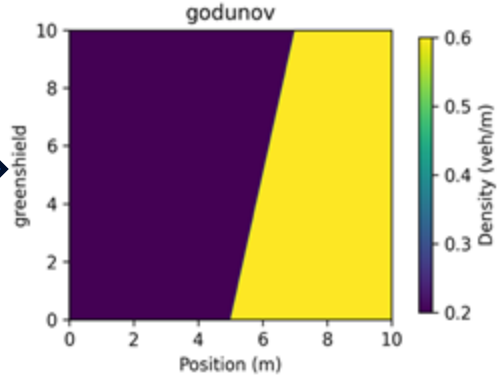


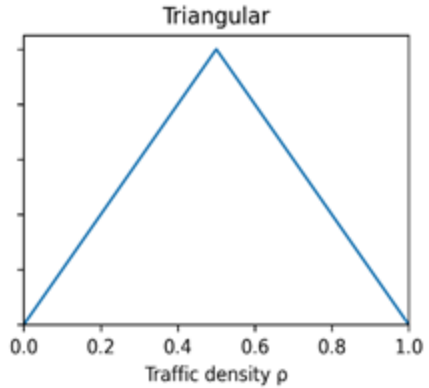
The solution to the Riemann problem depends on the flux function

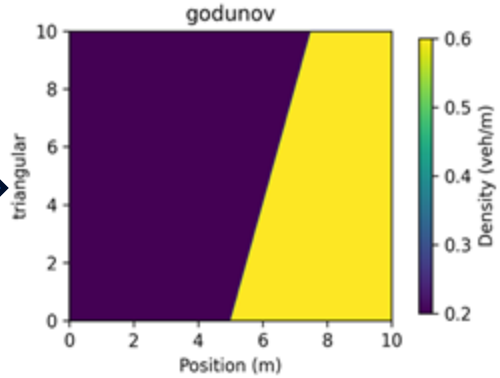
$$\rho_0(x) = \begin{cases} \rho_l & if\ x < 0 \\ \rho_r & if\ x > 0. \end{cases}$$

$$\rho(t,x) = \begin{cases} \rho_l & if\ x < \frac{f(\rho_l)-f(\rho_r)}{\rho_l-\rho_r} \\ \rho_r & if\ x > \frac{f(\rho_l)-f(\rho_r)}{\rho_l-\rho_r}. \end{cases}$$



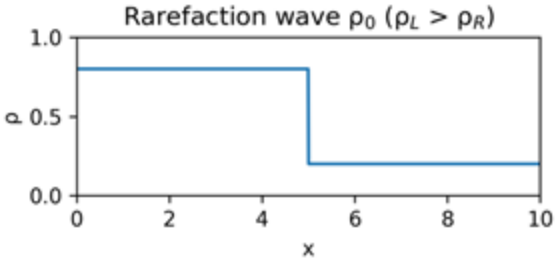






Building blocks for supervised learning: Riemann problems

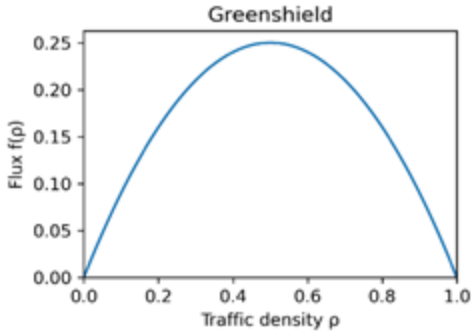
”Fundamental” problems for this PDE: Riemann probems

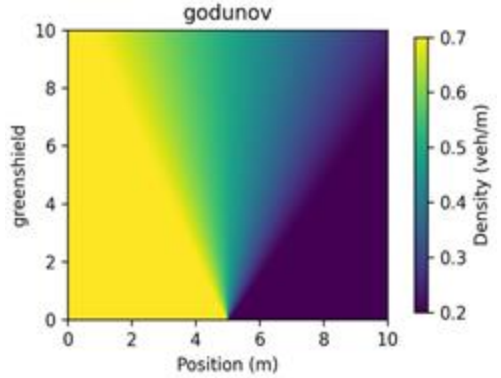


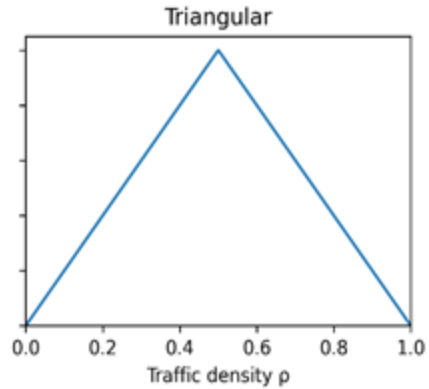
The solution to the Riemann problem depends on the flux function

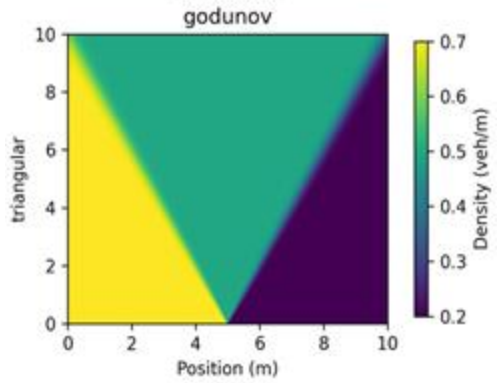
$$\rho_0(x) = \begin{cases} \rho_l & if\ x < 0 \\ \rho_r & if\ x > 0. \end{cases}$$

$$\rho(t,x) = \begin{cases} \rho_l & if\ x < f'(\rho_l)t, \\ (f')^{-1}\left(\frac{x}{t}\right) & if\ f'(\rho_l)t < x < f'(\rho_r)t, \\ \rho_r & if\ x > f'(\rho_r)t, \end{cases}$$

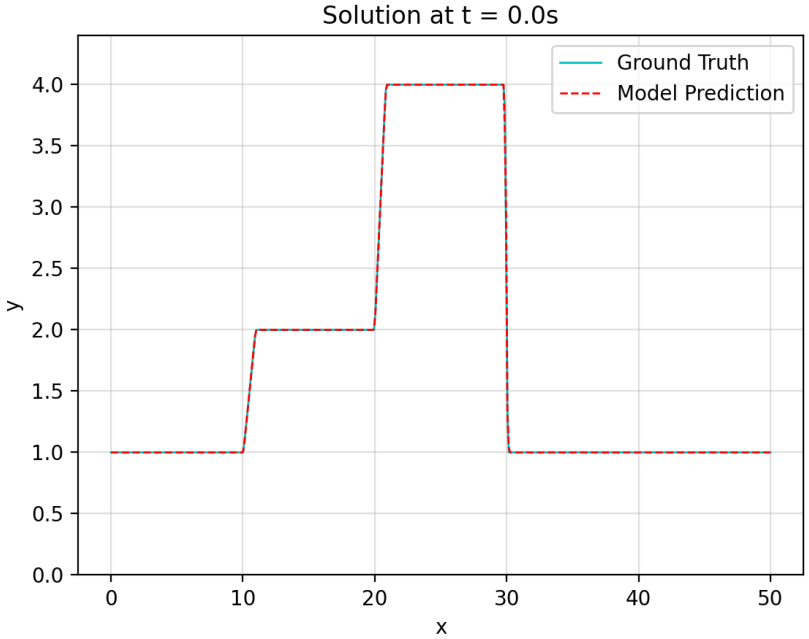


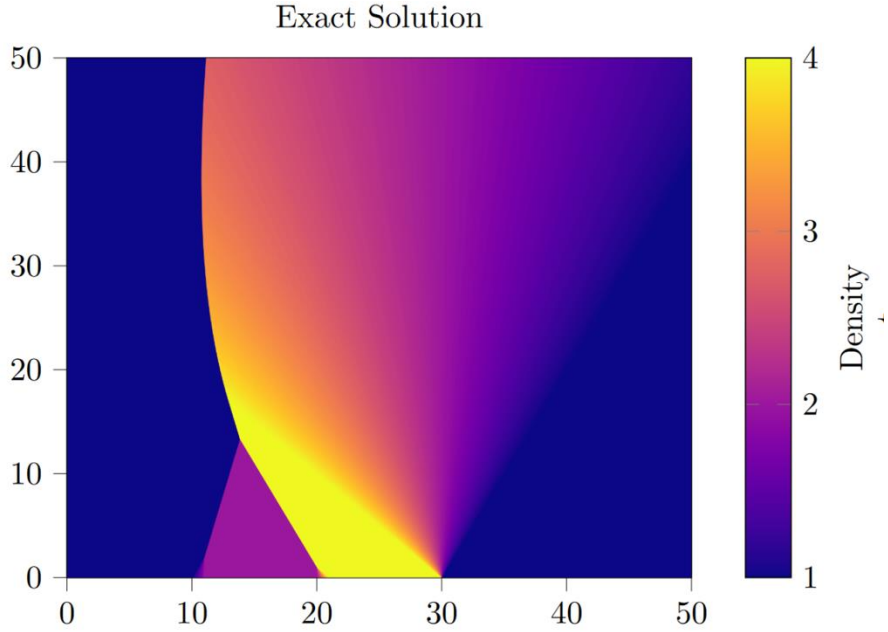






General solution as combination of Riemann problems

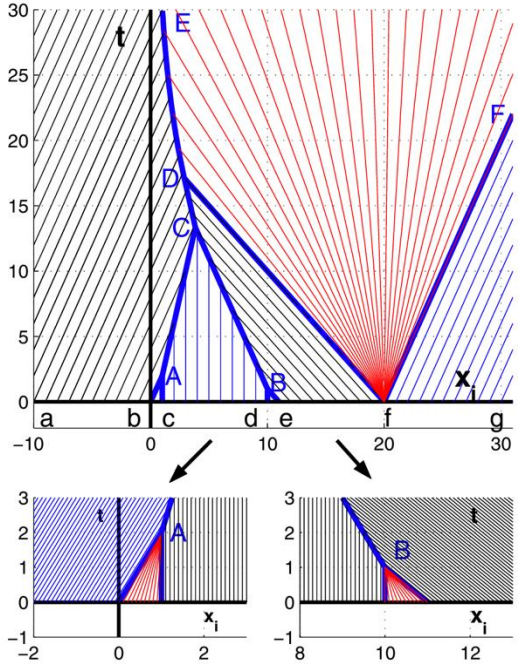




$$\rho_0(x) = \begin{cases} 1 & if\ x \leq 0 \\ 1+x & if\ x \in (0,1] \\ 2 & if\ x \in (1,10] \\ 2+2(x-10) & if\ x \in (10,11] \\ 4 & if\ x \in (11,20] \\ 1 & if\ x > 20. \end{cases}$$

General solution as combination of Riemann problems

Computations of solutions: combinations of fundamental solutions
For simple flux functions: these solutions can be computed semi-explicitly (explicit language)



$$\rho_0(x) = \begin{cases} 1 & \text{if } x \le 0 \\ 1+x & \text{if } x \in (0,1] \\ 2 & \text{if } x \in (1,10] \\ 2+2(x-10) & \text{if } x \in (10,11] \\ 4 & \text{if } x \in (11,20] \\ 1 & \text{if } x > 20. \end{cases}$$

$$\rho(t,x) = \begin{cases} 1 & \text{if } \begin{cases} (x > 20 + \frac{1}{2}t) \\ \vee\ (t \le 2 \ \wedge\ x \le \frac{1}{2}t) \\ \vee\ (2 \le t \le \frac{40}{3} \ \wedge\ x \le \frac{t+2}{4}) \\ \vee\ (\frac{40}{3} \le t \le \frac{154}{9} \ \wedge\ x \le \frac{43}{6} - \frac{t}{4}) \\ \vee\ \left(t \ge \frac{154}{9} \ \wedge\ x \le \frac{t}{2} - \sqrt{\frac{77}{2}t} + 20\right) \end{cases} \\ 1 + \frac{2x-t}{2-t} & \text{if } \left(t \le 2 \ \wedge\ \frac{t}{2} \le x \le 1\right) \\ 2 + \frac{2x-20}{1-t} & \text{if } (t \le 1 \ \wedge\ 10 \le x \le 11-t) \\ 2 & \text{if } \begin{cases} (t \le 1 \ \wedge\ 1 \le x \le 10) \\ \vee\ \left(1 \le t \le 2 \ \wedge\ 1 \le x \le \frac{21}{2} - \frac{t}{2}\right) \\ \vee\ \left(2 \le t \le \frac{40}{3} \ \wedge\ \frac{t+2}{4} \le x \le \frac{21-t}{2}\right) \end{cases} \\ 4 & \text{if } \begin{cases} (t \le 2 \ \wedge\ 11 - t \le x \le 20 - t) \\ \vee \left(1 \le t \le \frac{40}{3} \ \wedge\ \frac{21-t}{2} \le x \le 20 - t\right) \\ \vee \left(\frac{40}{3} \le t \le \frac{154}{9} \ \wedge\ \frac{43}{6} - \frac{t}{4} \le x \le 20 - t\right) \\ \vee (t \le 1 \ \wedge t \le 20 - x \ \wedge t \ge 11 - x) \end{cases} \\ 2 - \frac{2x-40}{t} & \text{if } \begin{cases} \left(t \le \frac{154}{9} \ \wedge\ 20 - t \le x \le \frac{t}{2} + 20\right) \\ \vee\ \left(t \ge \frac{154}{9} \ \wedge\ \frac{t}{2} - \sqrt{\frac{77}{2}t} + 20 \le x \le \frac{t}{2} + 20\right) \end{cases} \end{cases}$$

Supervised vs. unsupervised learning for PDEs

Classical Numerical Methods for PDEs:

- Finite Difference/Volume/Element methods [LeV02; Muf22; CD15].

Supervised Neural Network Approaches:

- Train NNs using labeled PDE solutions: CNNs [ZZ18], GNNs [BWW22], Neural Operators (FNO [Li+20], DeepONet [Lu+21]).
- Advantage: Can approximate PDEs from data without explicit PDE.
- Limitations: Large dataset required; no guarantees for physical properties (entropy, conservation).

Unsupervised Neural Network Approaches:

- PINNs enforce PDE residuals in loss functions without training data [RPK19].
- Weak formulations (test functions) handle irregularities like shocks more robustly [YF21; SWS23].
- Advantage: No explicit dataset required; physics embedded.
- Limitations: Standard PINNs fail for hyperbolic PDEs with shocks [MM23; DJM24]; need special weak-formulation techniques.

Supervised vs. unsupervised learning for PDEs

Challenges of Existing NN Solvers:

- Typical NN methods struggle with preserving:
  - Regularity (functional space of existence)
  - Entropy conditions (uniqueness/stability)
  - Conservation laws (mass, momentum)
  - Physical constraints (bounds, realistic behaviors)

Unique Difficulties with Hyperbolic PDEs:

- Solutions exhibit **singularities** such as shocks and expansion waves.
- **High sensitivity** to discretization errors, especially near discontinuities.

Supervised vs. unsupervised learning for PDEs

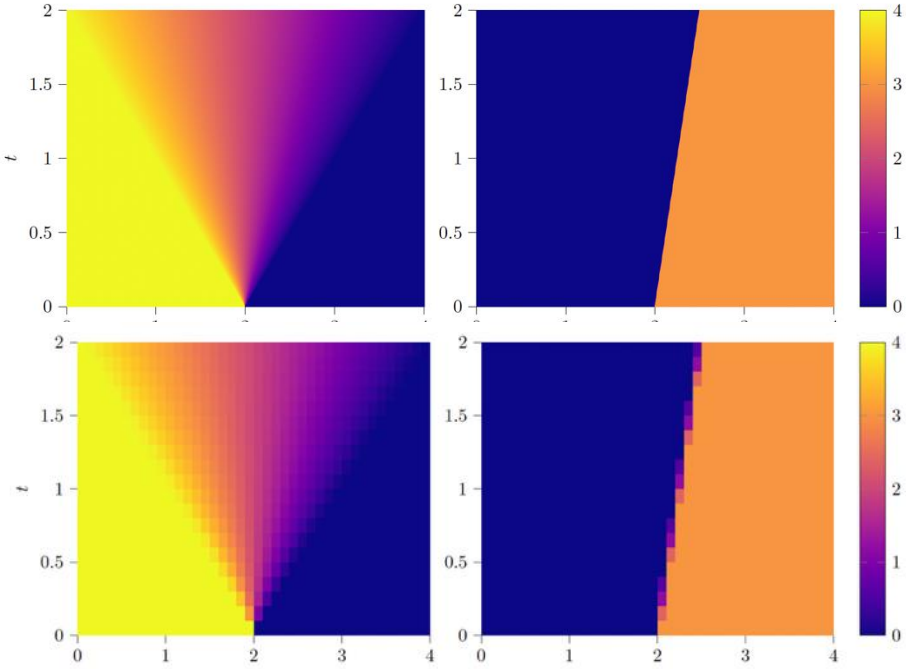
Our Solution: Neural Finite Volume Method (NFVM):

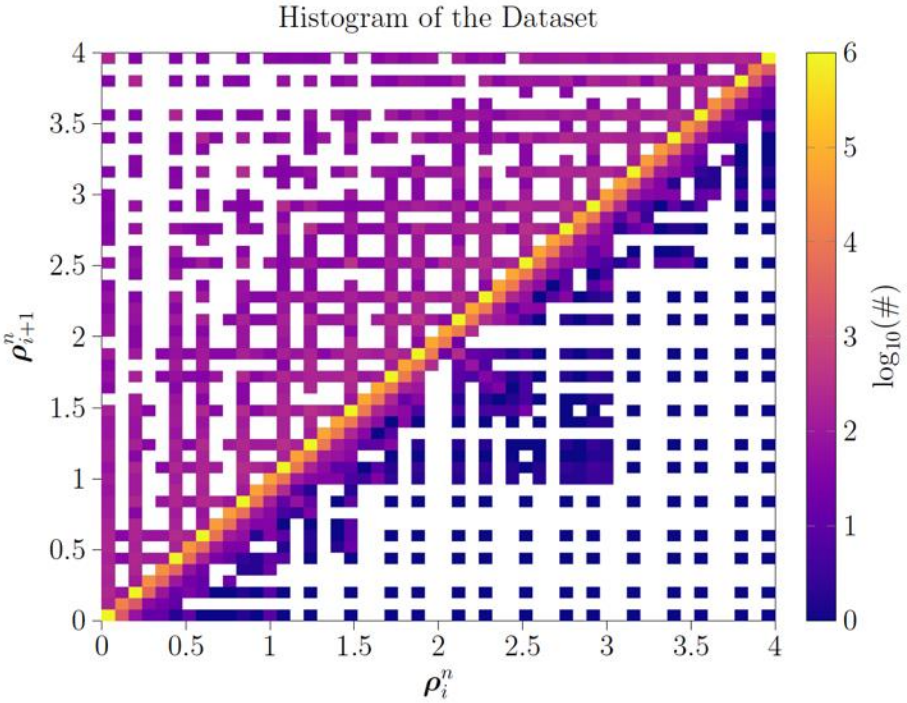
- Specifically designed NN architecture preserving essential PDE properties:
  - Entropy conditions to ensure physical relevance.
  - Stability and conservation laws inherently satisfied.
- Mitigate discretization accumulation errors found in classical schemes.
- Improves accuracy and stability of NN-based solvers for hyperbolic PDEs.

Goal: Bridge the gap between PDE theory and NN approximation, demonstrating NN's capability to accurately capture critical solution behaviors of hyperbolic systems.

Training set

Training set is composed of large number of solutions of Riemann problems :





Other canonical example: Burgers equation

The Burgers' equation is a PDE used in various domains such as fluid mechanics, nonlinear acoustics, gas dynamics, and traffic flow. It is a **conservation law**:

$$\partial_t u + \frac{1}{2}\partial_x u^2 = 0.$$

It is opposite to LWR because it has a **strictly convex** flux. As with LWR:

- Solutions may only exist in the weak formulation.
- Smooth initial conditions may lead to non-differentiable solutions.
- An entropic condition gives unicity to weak solutions.

Building blocks for a supervised learning (Burgers)

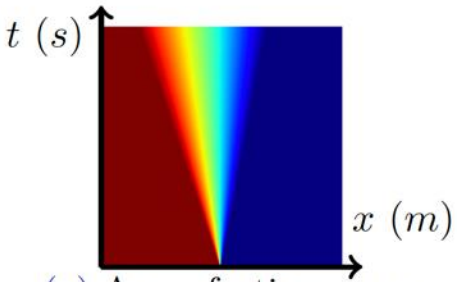
Consider the scalar conservation laws with the following initial condition (Riemann problem):

$$u_\circ(x) = u_L \mathbb{1}_{\{u_L < u_R\}} + u_R \mathbb{1}_{\{u_L \geq u_R\}}$$

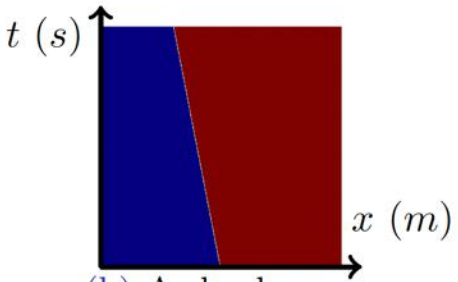
Then the Riemann solution (entropic solution) reads:

$$u(t,x) = \begin{cases} u_L, & x/t < \lambda_1 \\ f'^{-1}(x/t), & x/t \in [\lambda_1, \lambda_2) \\ u_R, & x/t \geq \lambda_2. \end{cases}$$

If $u_L < u_R$, the solution u represents a **shock wave** with $\lambda_1 = \lambda_2$ defined by the Rankine-Hugoniot condition, otherwise it is a **rarefaction (expansion) wave** with $\lambda_1 = f'(u_L)$ and $\lambda_2 = f'(u_R)$.



(a) A rarefaction wave.



(b) A shock wave.

Validation set: Lax Hopf formula

- The selected problems (LWR, Burgers) have exact solutions which helps evaluating the performance of the proposed NN solvers against traditional schemes.
- Let us consider a convex flux function f. Let us recall the H-J reformulation.
- Considering the potential function $u(t,x) = \partial_x \xi(t,x)$, the problem can be presented in the H-J framework by

$$\partial_t \xi(t,x) + H(\partial_x \xi(t,x)) = 0.$$

- The Lagrangian (dual):

$$L(u) = \sup_{v \in \mathbb{R}} \{uv - f(v)\}$$

  then, $H = L^*$ and $H^* = L$.
- The Lagrangian L is convex and super-linear.

Validation set: Lax Hopf formula

[Lax57] provides an explicit solution for the H-J equations:

$$\xi(t,x) \overset{\text{def}}{=} \inf_{y\in\mathbb{R}} \left\{ tL\left(\frac{x-y}{t}\right) + \xi_\circ(y) \right\}.$$

For Burgers', $f(u) = \frac{u^2}{2}$ and $L(v) = \frac{v^2}{2}$. Therefore, the explicit form becomes

$$\xi(t,x) \overset{\text{def}}{=} \inf_{y\in\mathbb{R}} \left\{ \frac{(x-y)^2}{2t} + \xi_\circ(y) \right\}.$$

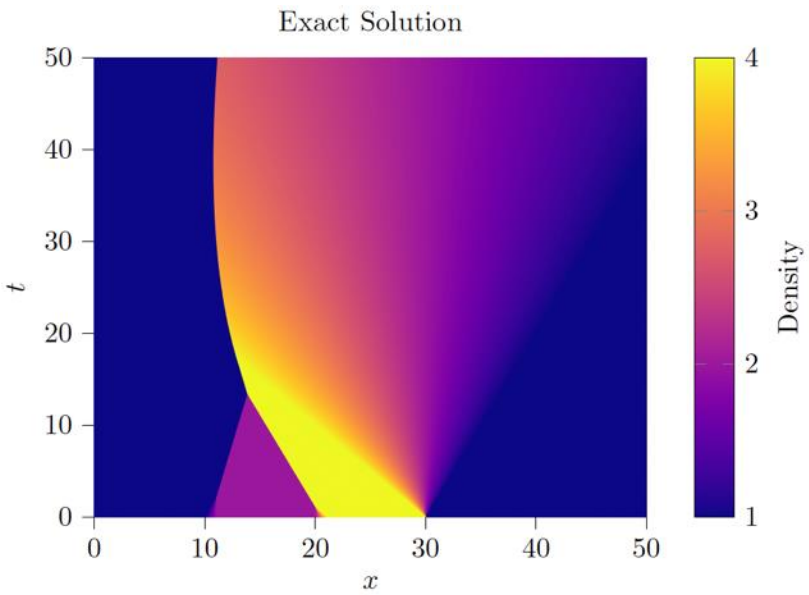
[CB10] proposed an explicit algorithm to calculate the solution of Lax-Hopf equations for any piecewise-constant initial condition.

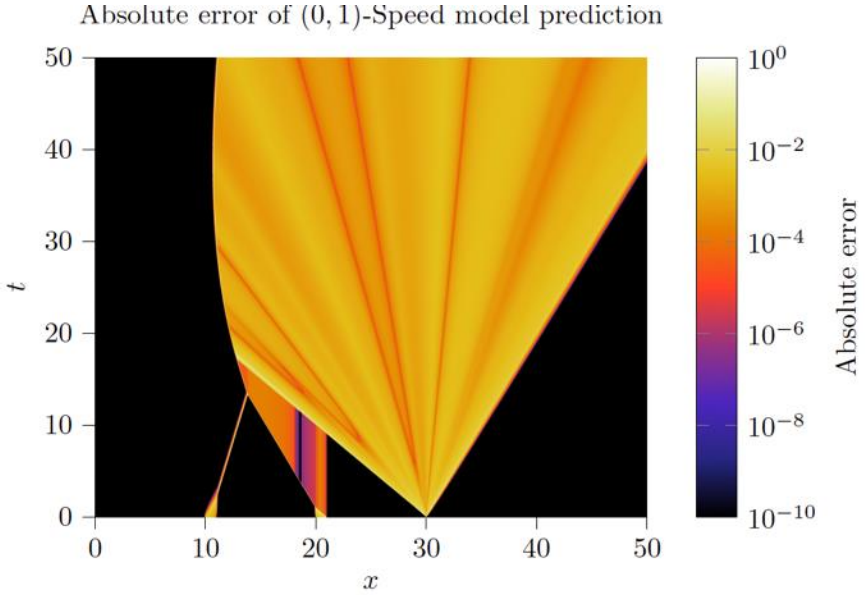
PD Lax. "Hyperbolic systems of conservation laws II". In: *Communications on Pure and Applied Mathematics* 10.4 (1957), pp. 537–566.
Christian G. Claudel and Alexandre M. Bayen. "Lax–Hopf Based Incorporation of Internal Boundary Conditions Into Hamilton-Jacobi Equation. Part II: Computational Methods". In: *IEEE Transactions on Automatic Control* 55.5 (2010), pp. 1158–1174.

Nature of the error and illustration

Errors mostly occurs

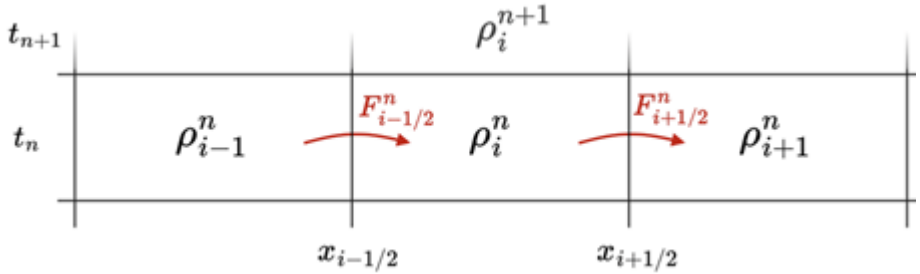
- Close to the shocks (artefact of discretization)
- In the fans of the expansion waves





Finite volume approximation

Define cells by discretizing time and space:



$$x_{i\pm1/2} = i\Delta x \pm \frac{\Delta x}{2}$$

$$t_n = n\Delta t$$

Average density in a cell: $$\rho_{\mathbf{i}}^{\mathbf{n}} = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \rho(t_n, x)dx$$

Average flux going between cells: $$\mathbf{F}_{\mathbf{i+1/2}}^{\mathbf{n}} = \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f(\rho(x_{i+1/2}, t))dt$$

Finite volume method

Then, integrating the PDE over cells as follows:

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \int_{t_n}^{t_{n+1}} \left(\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}(f(\rho(x,t))) \right) dt\, dx = 0$$

yields the following relation, describing the average cell density in the next time step as a function of the average flows entering and exiting that cell over time:

$$\rho_{\mathbf{i}}^{\mathbf{n+1}} = \rho_{\mathbf{i}}^{\mathbf{n}} - \frac{\Delta t}{\Delta x} \underbrace{\left(\mathbf{F}_{\mathbf{i+1/2}}^{\mathbf{n}} - \mathbf{F}_{\mathbf{i-1/2}}^{\mathbf{n}} \right)}$$

In classical numerical analysis:
approximated for forward computation

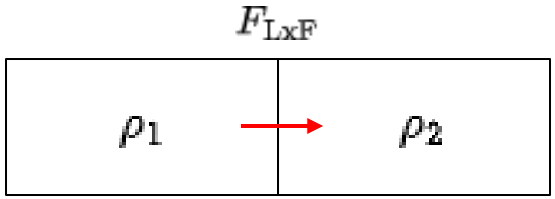
Different Numerical Schemes

Classical methods:
- Lax-Friedrichs (Lax's PhD with Friedrichs: 1949)
- Godunov (1956)
- Discontinuous Galerkin (DG)

More recent methods (post 1980')
- Engquist-Osher
- Essentially Non Oscillatory (ENO)
- Weighted Essentially Non Oscillatory (WENO)

Lax-Friedrichs (made at NYU)

$$\mathcal{F}_{\mathrm{LxF}} : (\rho_1, \rho_2) \mapsto \frac{1}{2}\left(f(\rho_1) + f(\rho_2) - \frac{\Delta x}{\Delta t}(\rho_2 - \rho_1)\right)$$

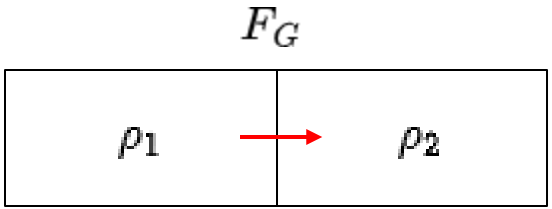


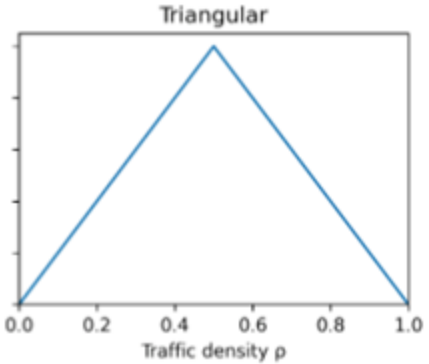




Godunov

$$\mathcal{F}_G(\rho_1, \rho_2) = \begin{cases} f(\rho_2) & \text{if } \rho_c < \rho_2 < \rho_1 \\ f(\rho_c) & \text{if } \rho_2 < \rho_c < \rho_1 \\ f(\rho_1) & \text{if } \rho_2 < \rho_1 < \rho_c \\ \min(f(\rho_1), f(\rho_2)) & \text{if } \rho_1 \leq \rho_2 \end{cases}$$

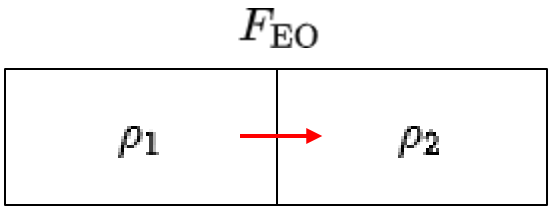


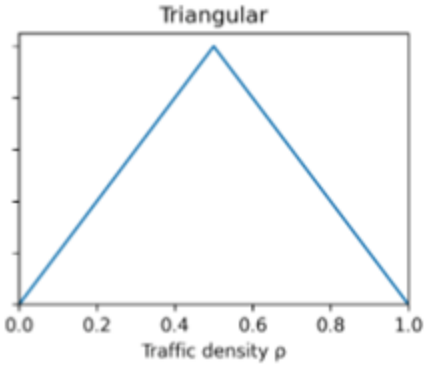




Engquist-Osher

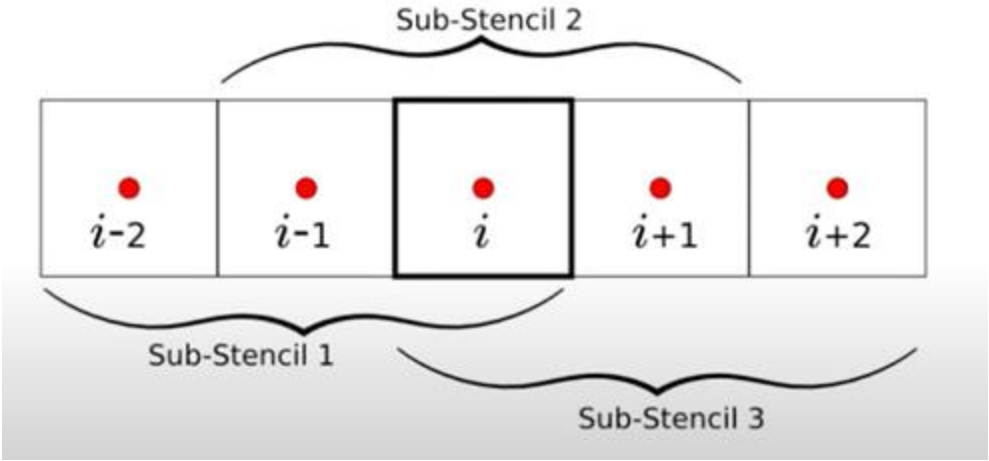
$$\mathcal{F}_{\mathrm{EO}}(\rho_1, \rho_2) = \begin{cases} f(\rho_1) & \text{if } \rho_2, \rho_1 \leq \rho_c \\ f(\rho_2) & \text{if } \rho_2, \rho_1 > \rho_c \\ f(\rho_c) & \text{if } \rho_2 \leq \rho_c < \rho_1 \\ f(\rho_1) + f(\rho_2) - f(\rho_c) & \text{if } \rho_1 \leq \rho_c < \rho_2. \end{cases}$$







Weighted Essentially Non Oscillatory (WENO)



$$u_{i+1/2}=\frac{1}{30}\bar{u}_{i-2}-\frac{13}{60}\bar{u}_{i-1}+\frac{47}{60}\bar{u}_{i}+\frac{9}{20}\bar{u}_{i+1}-\frac{1}{20}\bar{u}_{i+2}. \tag{12}$$

Notice that this is a fifth-order accurate approximation if the function $u(x)$ is smooth in the large stencil S. Further investigation on the fifth-order approximation $u_{i+1/2}$ in (12) and the three third-order approximations $u^{(1)}_{i+1/2}, u^{(2)}_{i+1/2}$ and $u^{(3)}_{i+1/2}$, defined by (5), (7) and (8) reveals the following linear combination relationship:

$$u_{i+1/2}=\gamma_1 u^{(1)}_{i+1/2}+\gamma_2 u^{(2)}_{i+1/2}+\gamma_3 u^{(3)}_{i+1/2}, \tag{13}$$

where the constants γ_1, γ_2 and γ_3, satisfying $\gamma_1 + \gamma_2 + \gamma_3 = 1$, are called the linear weights. In this case they take values

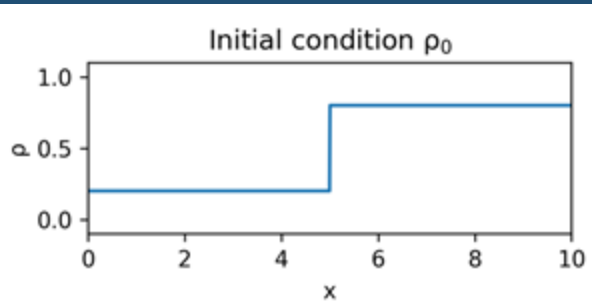
$$\gamma_1=\frac{1}{10}, \quad \gamma_2=\frac{3}{5}, \quad \gamma_3=\frac{3}{10}. \tag{14}$$

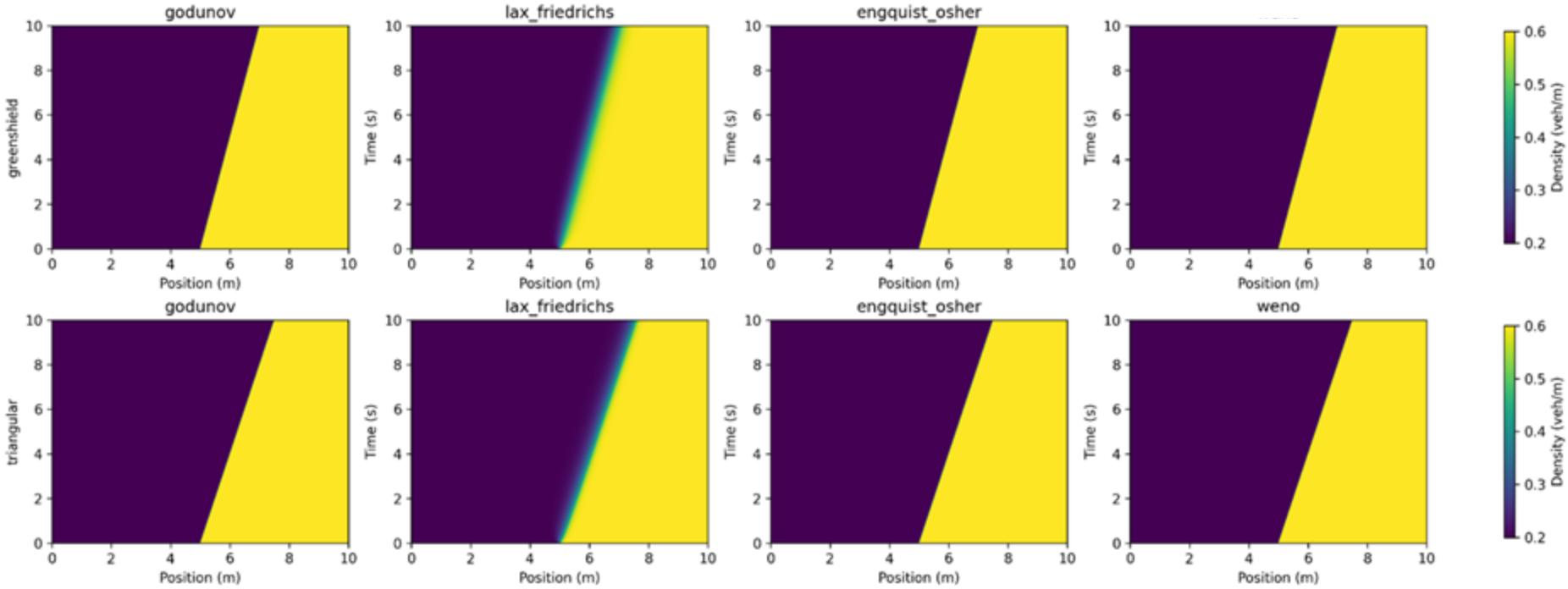
Idea:

Instead of using only one of the candidate stencils to form the reconstruction, one uses a convex combination of all of them.

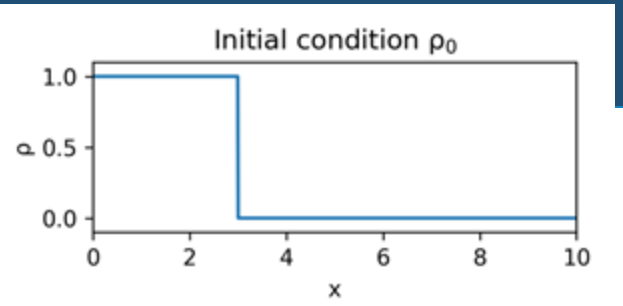


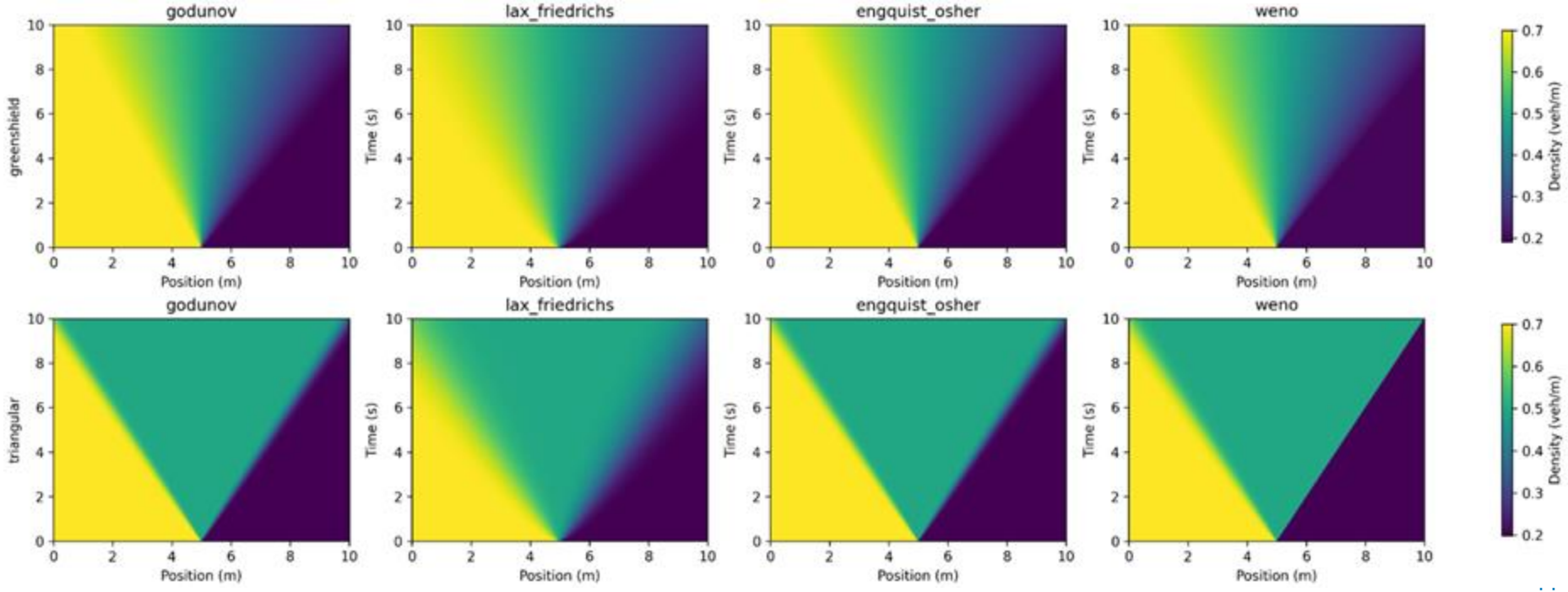
Comparison of some of the schemes



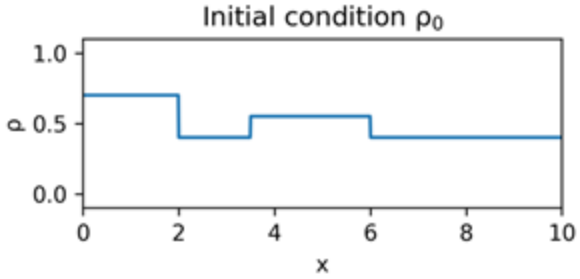


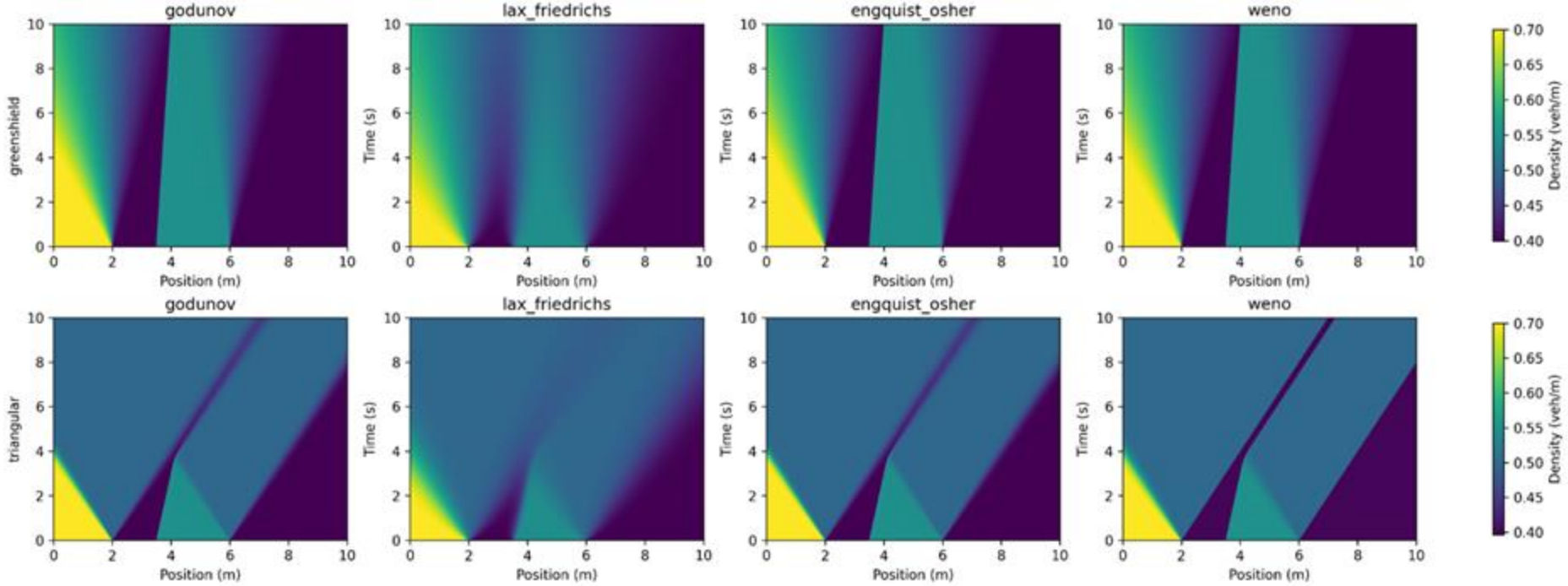
Comparison of some of the schemes





Comparison of some of the schemes





Forward computation (FVM)

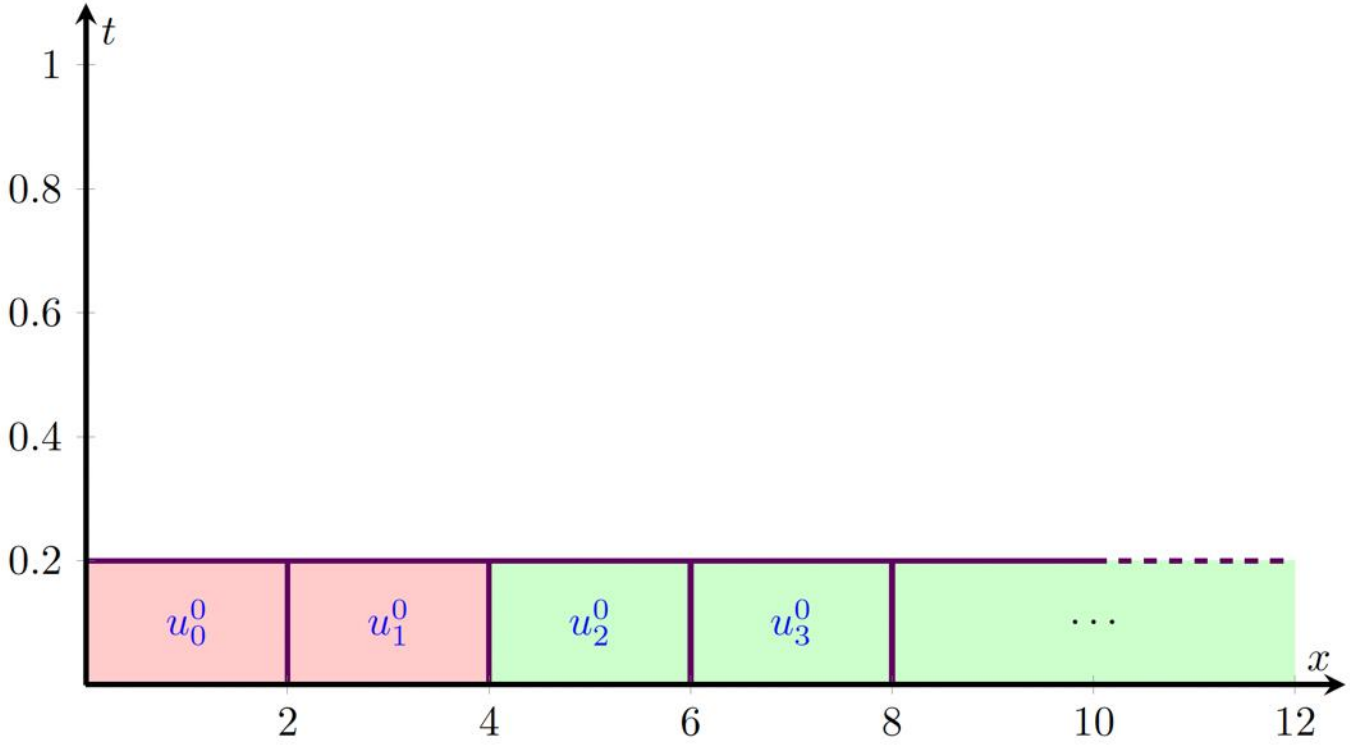


Figure: Illustration of the finite volume method.

Forward computation (FVM)

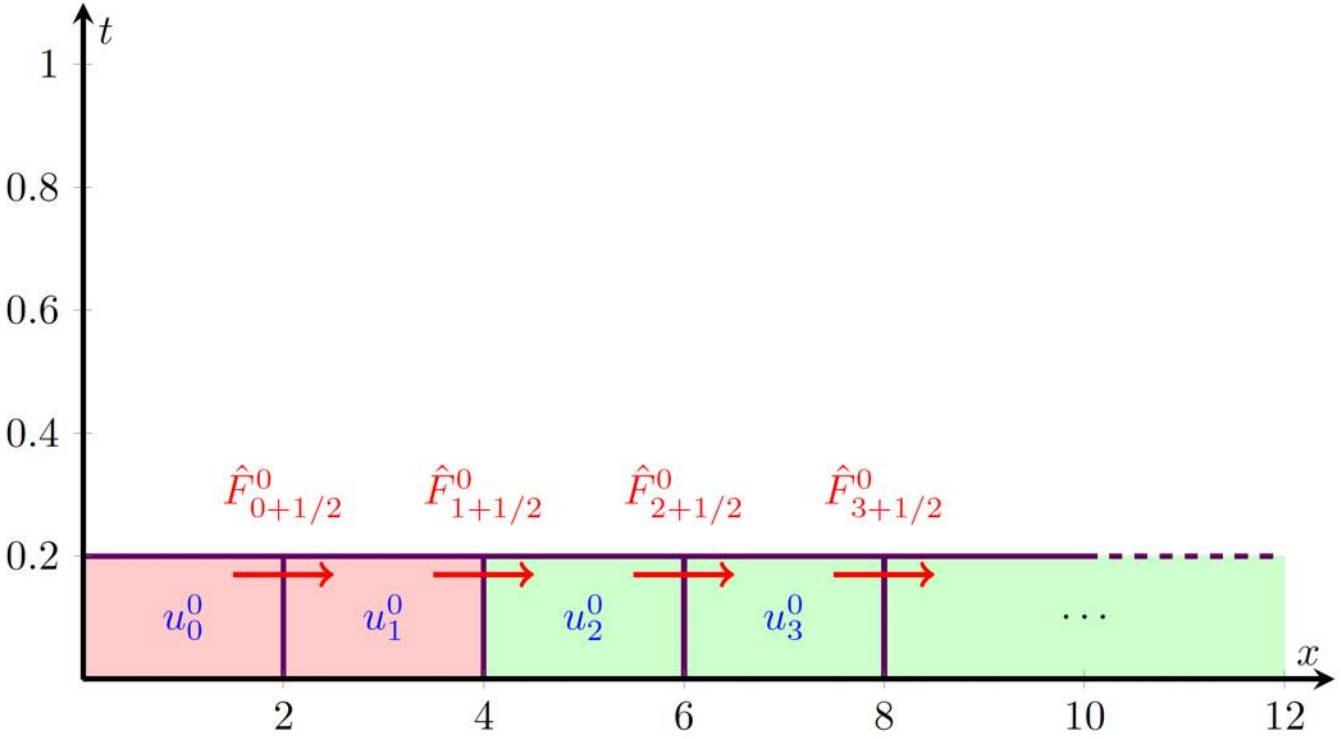


Figure: Illustration of the finite volume method.

Forward computation (FVM)

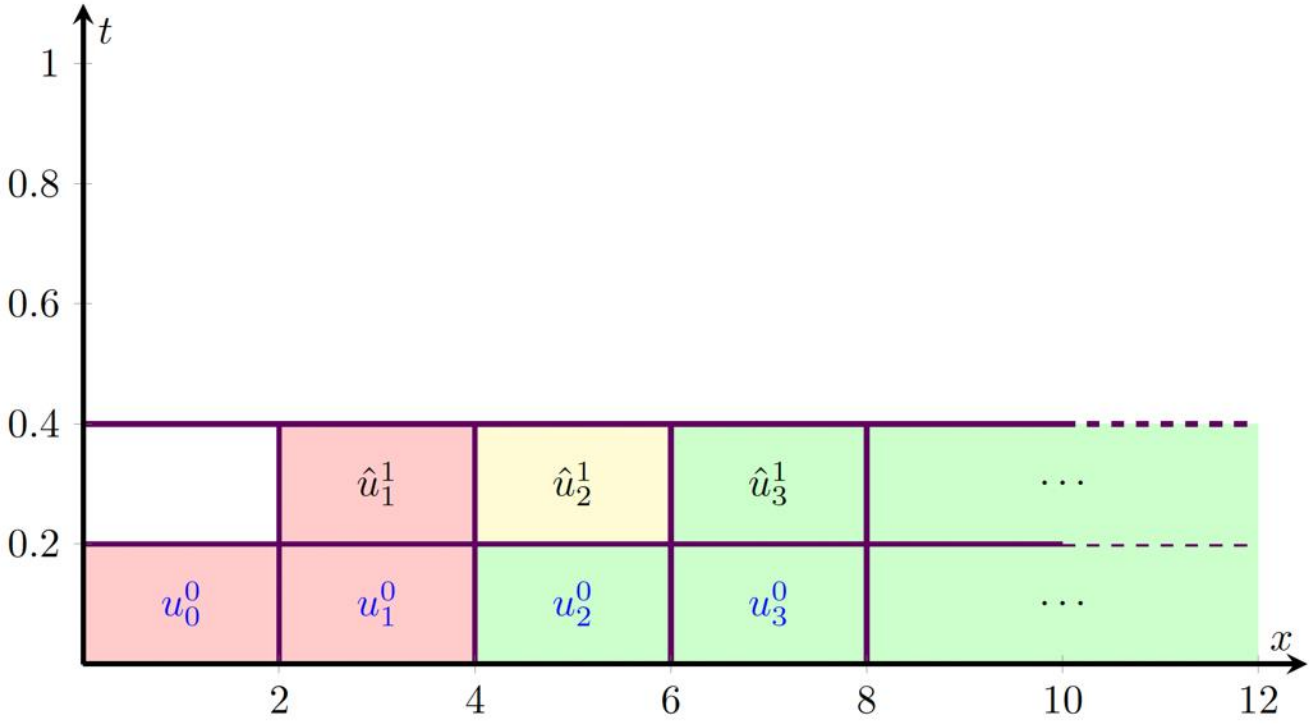


Figure: Illustration of the finite volume method.

Forward computation (FVM)

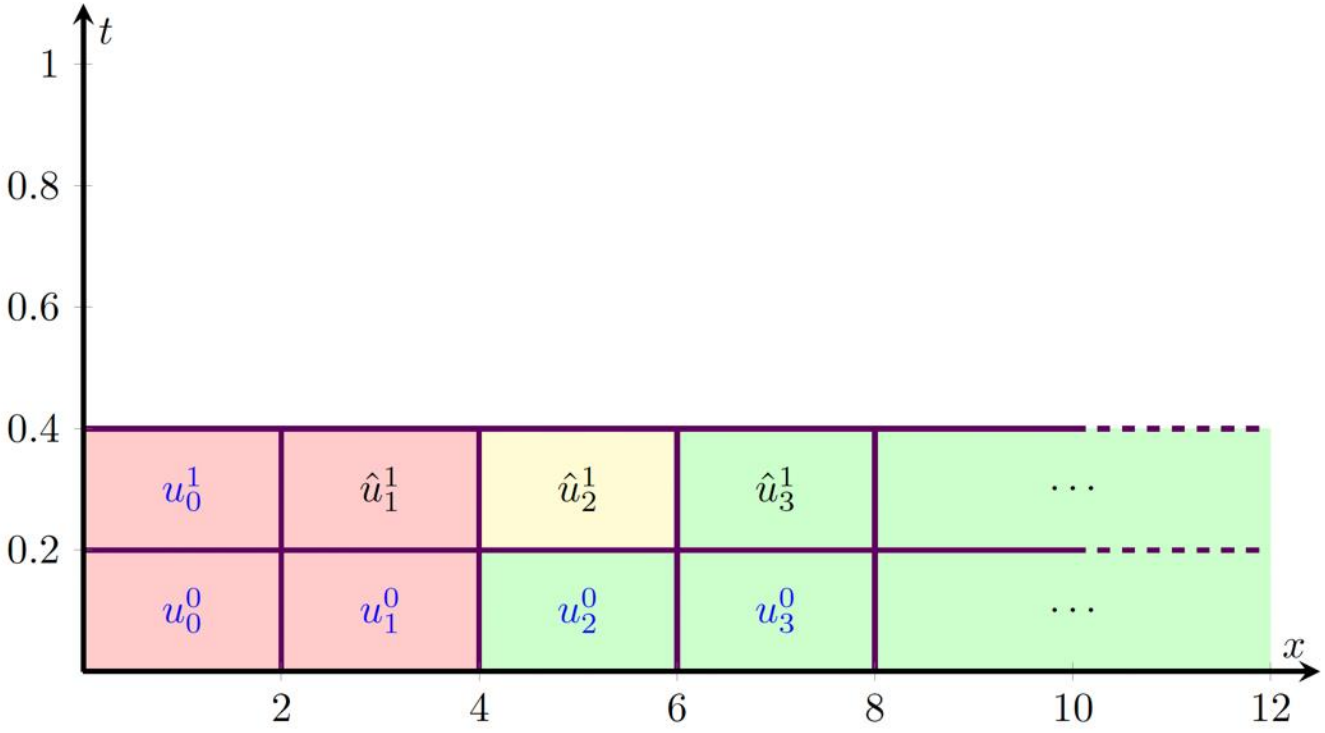


Figure: Illustration of the finite volume method.

Forward computation (FVM)

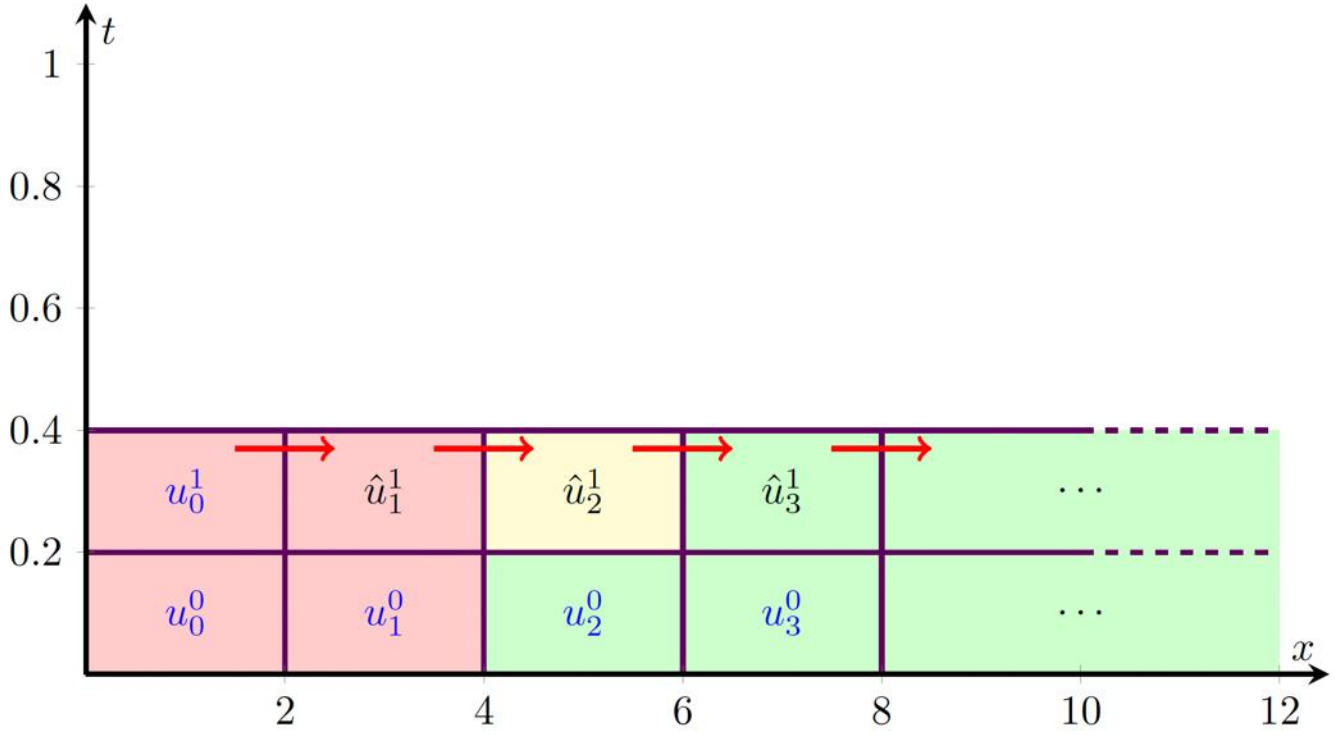


Figure: Illustration of the finite volume method.

Forward computation (FVM)

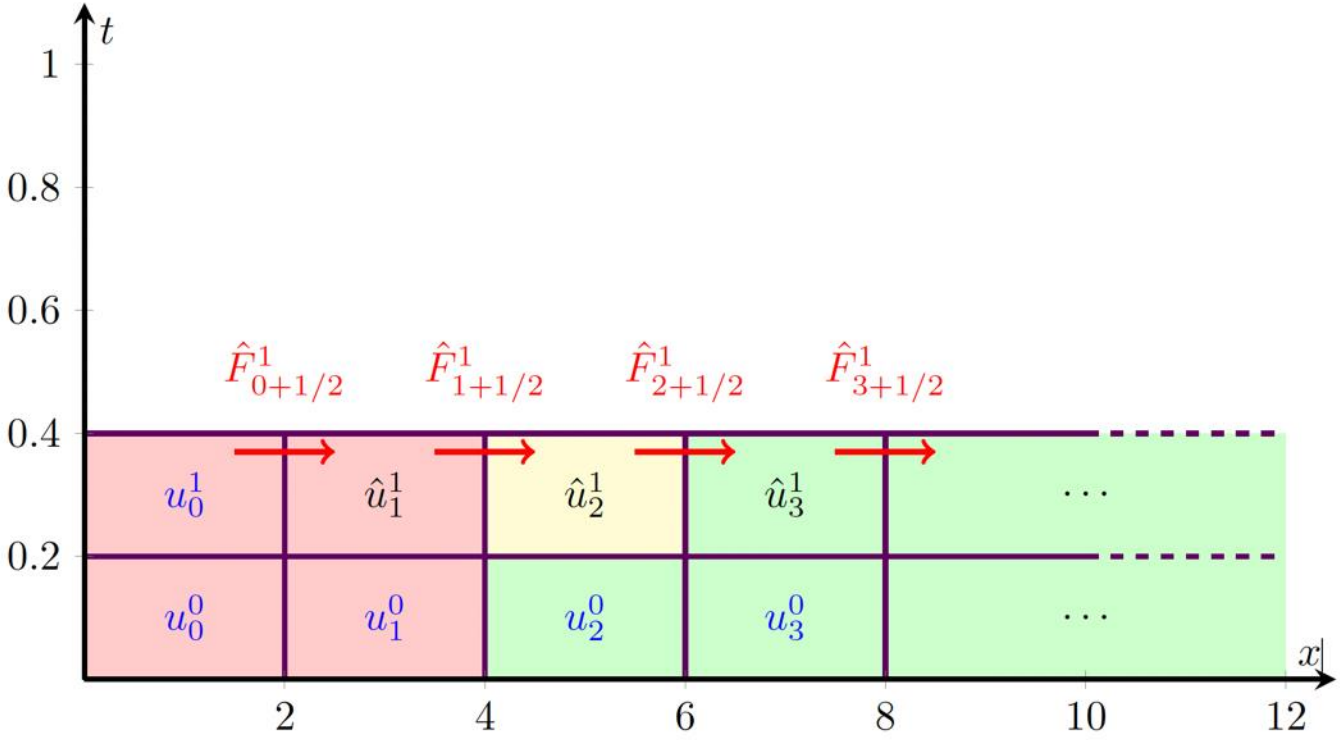


Figure: Illustration of the finite volume method.

Forward computation (FVM)

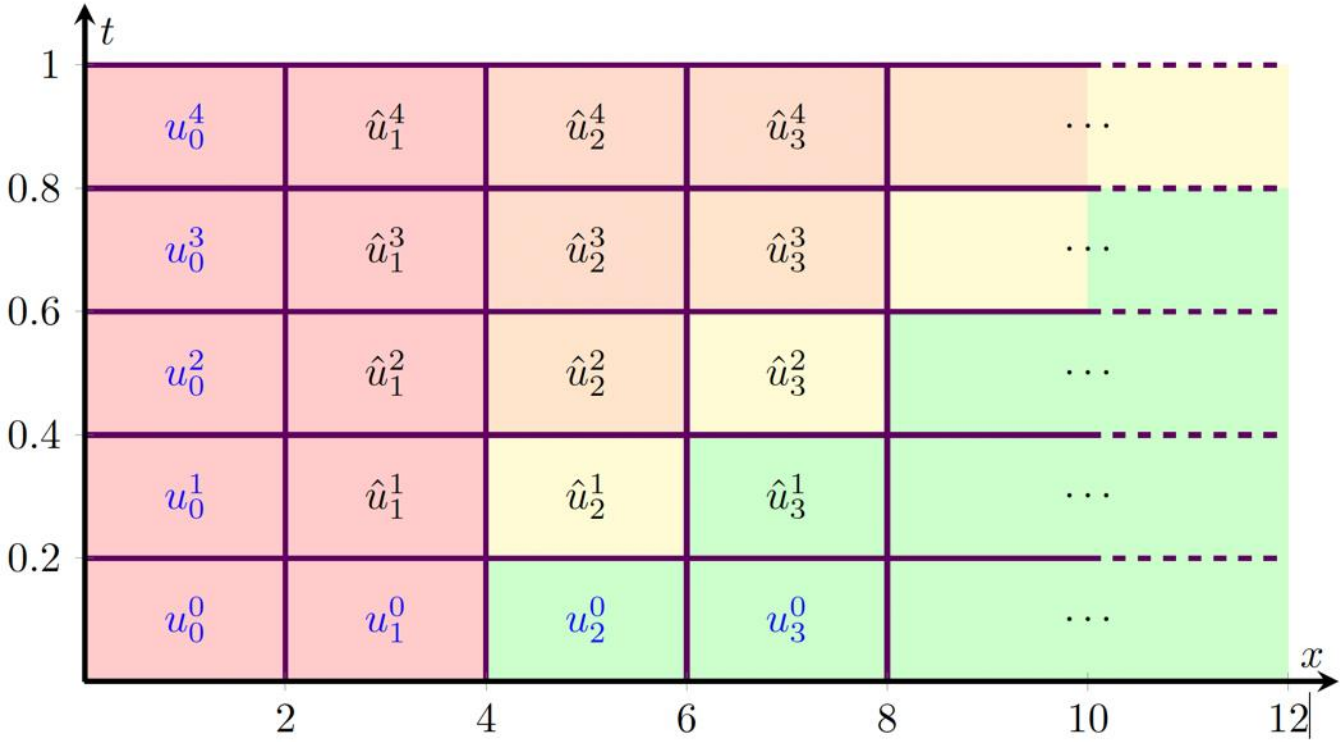
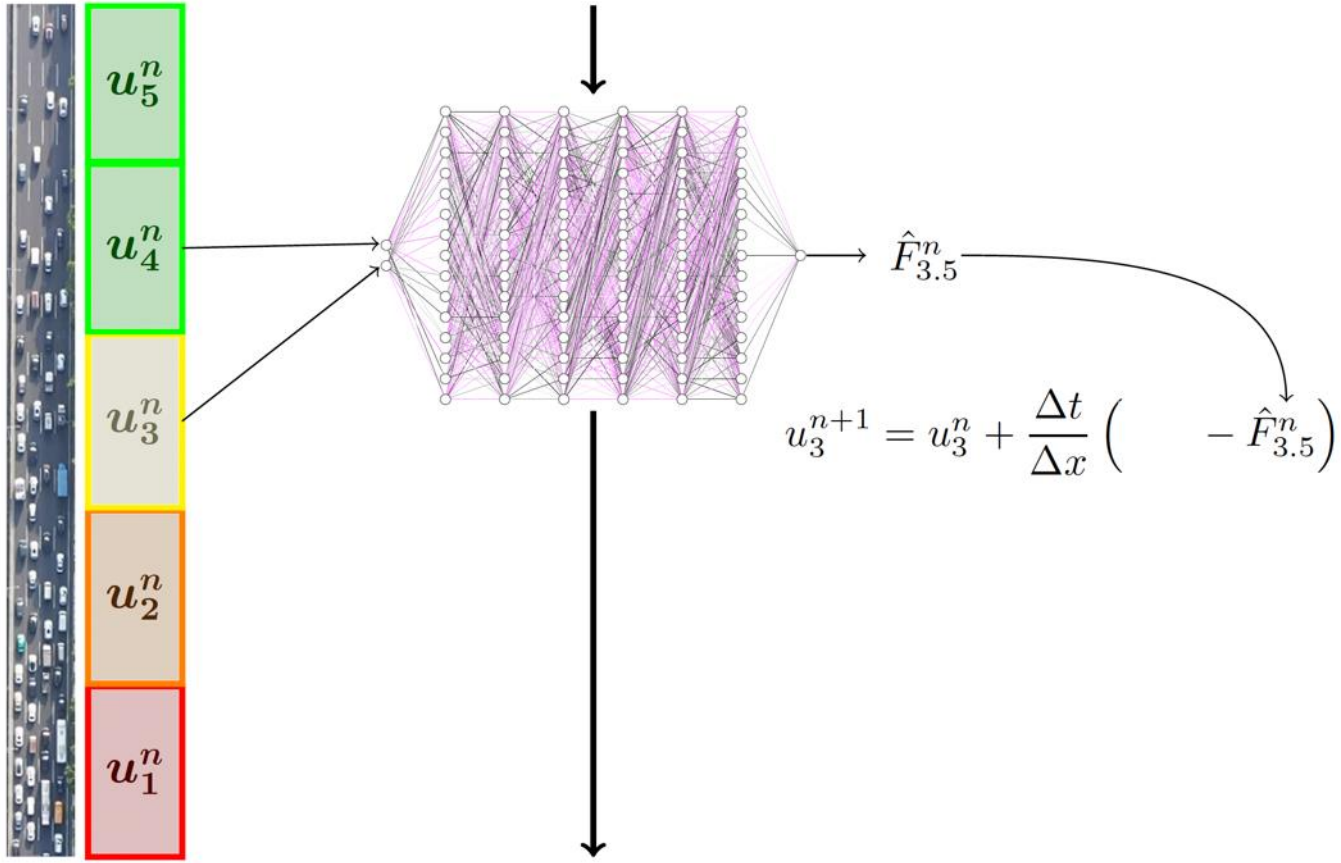
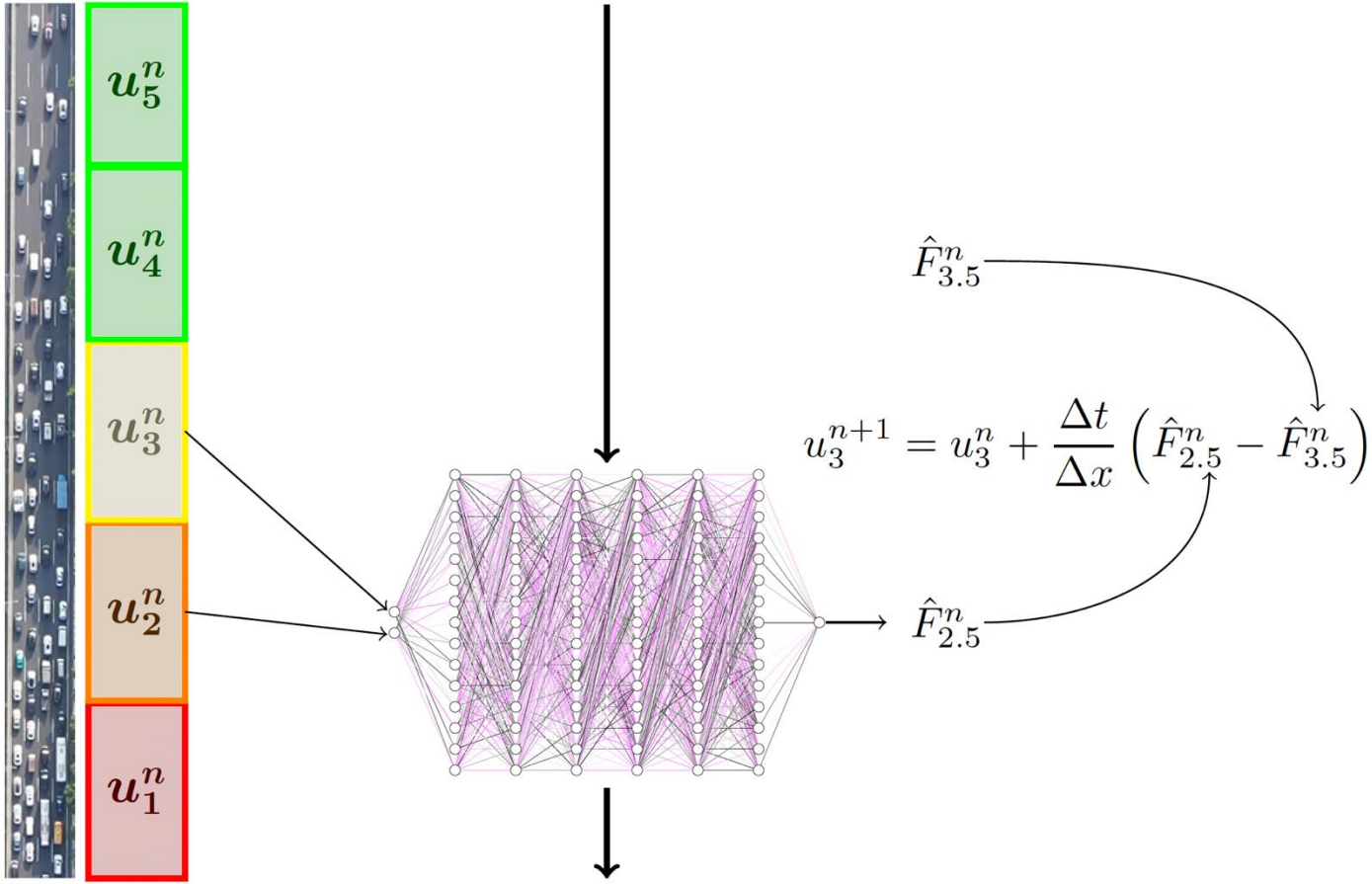


Figure: Illustration of the finite volume method.

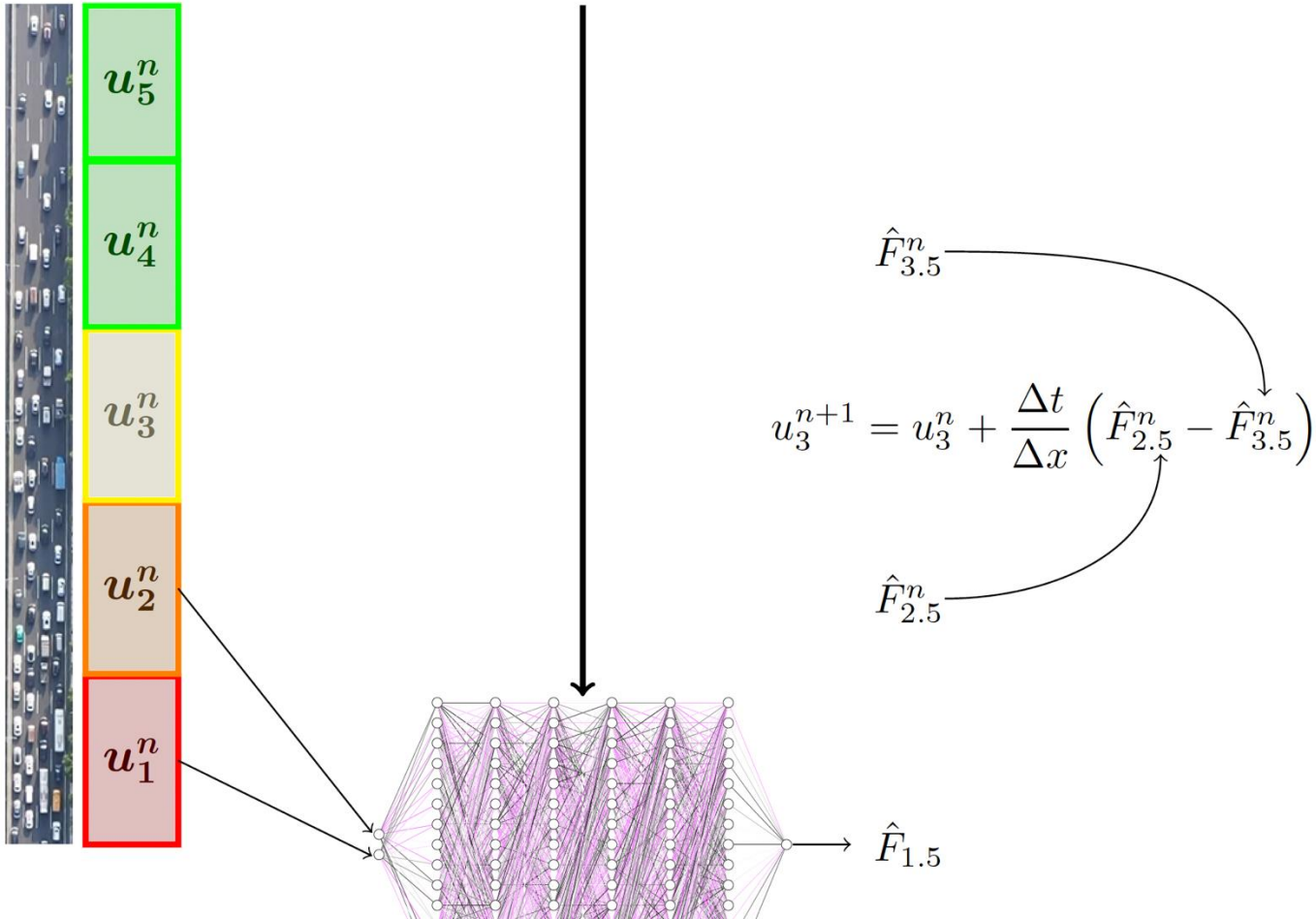
Neural “approximation” of FVM schemes



Neural “approximation” of FVM schemes



Neural “approximation” of FVM schemes



NN architecture

Our method: use a neural network as numerical flux.
Architecture: 1D CNN

- Kernel size: number of stencils used (2 for 1st order) for the first layer.
- Channels: 1 per timestep (Total is 1 for classic FVM).
- Layers: 6
- Hidden layer size: 15
- Learnt parameters: < 2000

The prediction updates $\hat{u}^{n+1}$ given $\hat{u}^n$ is as follows: $\forall j, m \in [\![1, 15]\!] \times [\![2, 5]\!]$,

$$\begin{cases} \boldsymbol{y_{(1,j)}} = \tanh\left(b_{(1,j)} + \boldsymbol{w_{(1,j,1)}} \star \hat{\boldsymbol{u}}^{\boldsymbol{n}}\right) \\ \boldsymbol{y_{(m,j)}} = \tanh\left(b_{(m,j)} + \sum_{k=1}^{15} w_{(m,j,k)} \boldsymbol{y_{(m-1,k)}}\right) \\ \hat{\boldsymbol{F}}^{\boldsymbol{n}}_{\cdot+\boldsymbol{1/2}} = b_{(6,1)} + \sum_{k=1}^{15} w_{(6,1,k)} \boldsymbol{y_{(5,k)}} \\ \hat{\boldsymbol{u}}^{\boldsymbol{n+1}} = \hat{\boldsymbol{u}}^{\boldsymbol{n}} + \dfrac{\Delta t}{\Delta x}\left(\hat{\boldsymbol{F}}^{\boldsymbol{n}}_{\cdot-\boldsymbol{1/2}} - \hat{\boldsymbol{F}}^{\boldsymbol{n}}_{\cdot+\boldsymbol{1/2}}\right) \end{cases}$$

where $\boldsymbol{w_{(1,j,1)}} \in \mathbb{R}^2$, $w_{(m,j,k)}, w_{(6,1,k)} \in \mathbb{R}$, and $b_{(1,j)}, b_{(m,j)}, b_{(6,1)} \in \mathbb{R}$ are the trained neural network parameters.

Unsupervised learning approach

Our method: use a neural network as numerical flux.
Architecture: 1D CNN

- Kernel size: number of stencils used (2 for 1st order) for the first layer.
- Channels: 1 per timestep (Total is 1 for classic FVM).
- Layers: 6
- Hidden layer size: 15
- Learnt parameters: < 2000

Supervised:

$$\mathcal{L} = \mathbb{E}_{u \sim \mathcal{D}} ||u - \hat{u}||_2^2.$$

Where:

- $\mathcal{D}$ is a dataset of solutions to Riemann problems.
- u is a solution from $\mathcal{D}$.
- $\hat{u}$ is a solution computed using the NN on the same initial conditions as u.
- While the solution is defined in L_1, we optimise the solution in L_2 over a bounded domain, which is justified by uniqueness; in order to enable strictly convex optimisation.

Supervised learning approach (weak-PINN)

Our method: use a neural network as numerical flux.
Architecture: 1D CNN

- Kernel size: number of stencils used (2 for 1st order) for the first layer.
- Channels: 1 per timestep (Total is 1 for classic FVM).
- Layers: 6
- Hidden layer size: 15
- Learnt parameters: < 2000

Supervised:

$$\mathcal{L} = \mathbb{E}_{u\sim\mathcal{D}}||u - \hat{u}||_2^2.$$

Unsupervised:

$$\mathcal{L} = \mathbb{E}_{\varphi\in\Phi, u_0\sim\mathcal{R}} \int_T \left(\int_X \partial_t \hat{u}(t,\cdot)\varphi(\cdot) + \int_X f(\hat{u}(t,\cdot))\varphi'(\cdot) \right)^2 \mathrm{d}t$$

where Φ is a set of suitable test functions, $\mathcal{R}$ is a distribution of random Riemann initial conditions and X and T are the space and time domains.

Tradeoffs

	NFVM	UNFVM	FVM	DG
Doesn't need the equation	✓	✗	✗	✗
Doesn't need data	✗	✓	✓	✓
Computationally cheap	✓	✓	✓	✗
Easy to implement	✓	✓	∼*	✗
Easy to scale up stencil size	✓	✓	✗	✗
Conservation of mass is guaranteed	✓	✓	✓	✓

*Some FVM methods are trivial to implement (*e.g.* Godunov), while higher-order methods can be more complicated (*e.g.* ENO/WENO).

Predictive capabilities of the NN approach

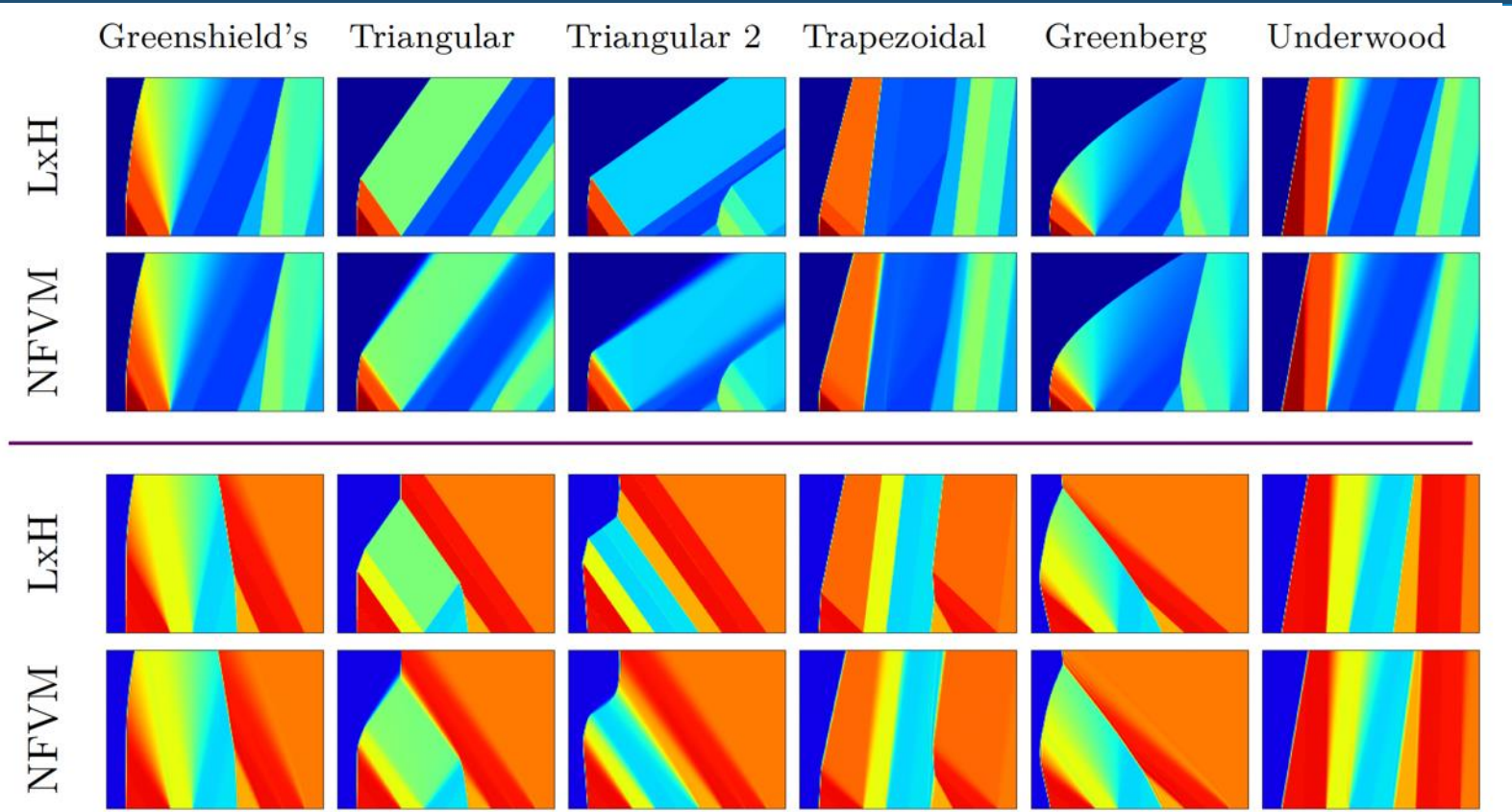
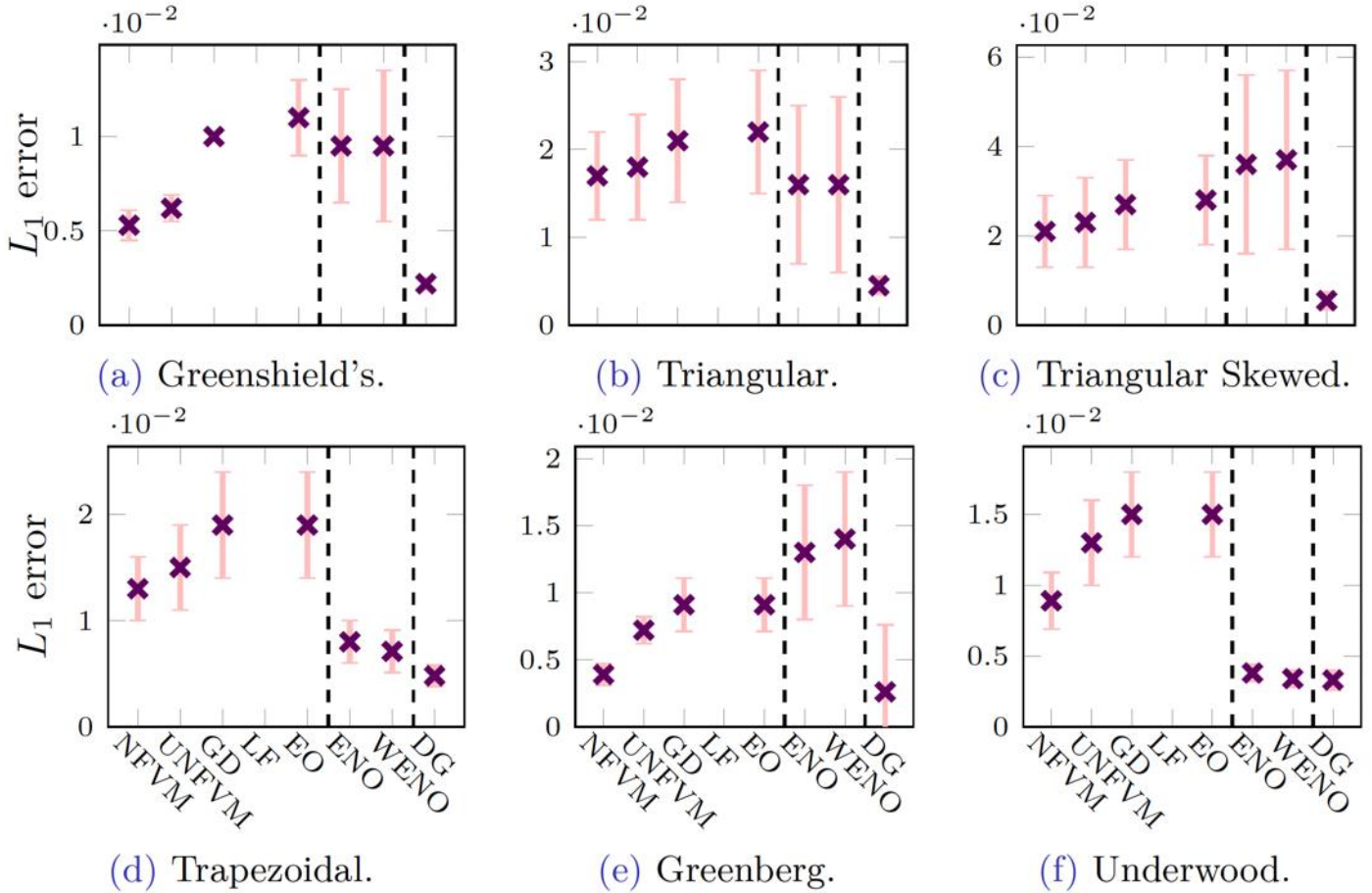


Figure: Ground truth vs. NFVM prediction on 2 different initial conditions.

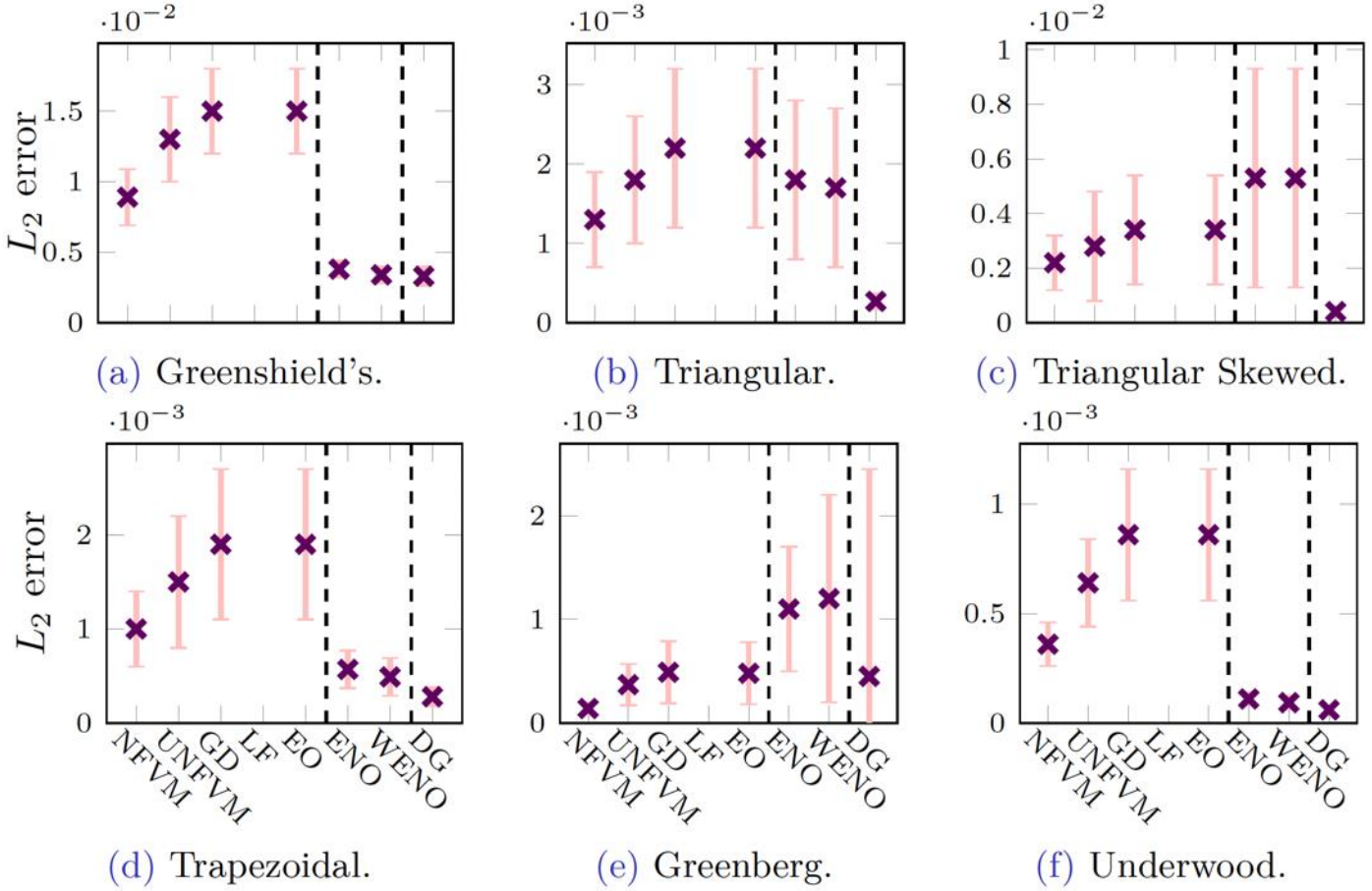
Predictive capabilities of the NN approach



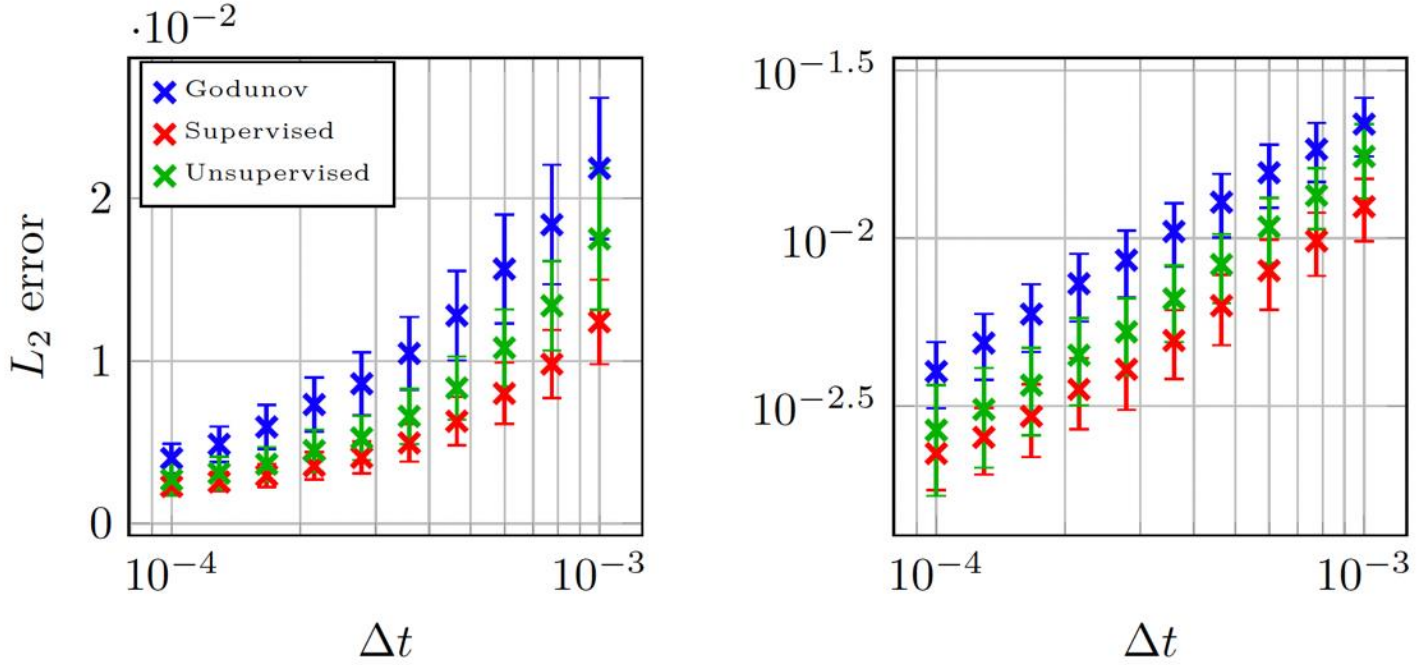
(a) Greenshield's. (b) Triangular. (c) Triangular Skewed.

(d) Trapezoidal. (e) Greenberg. (f) Underwood.

Predictive capabilities of the NN approach



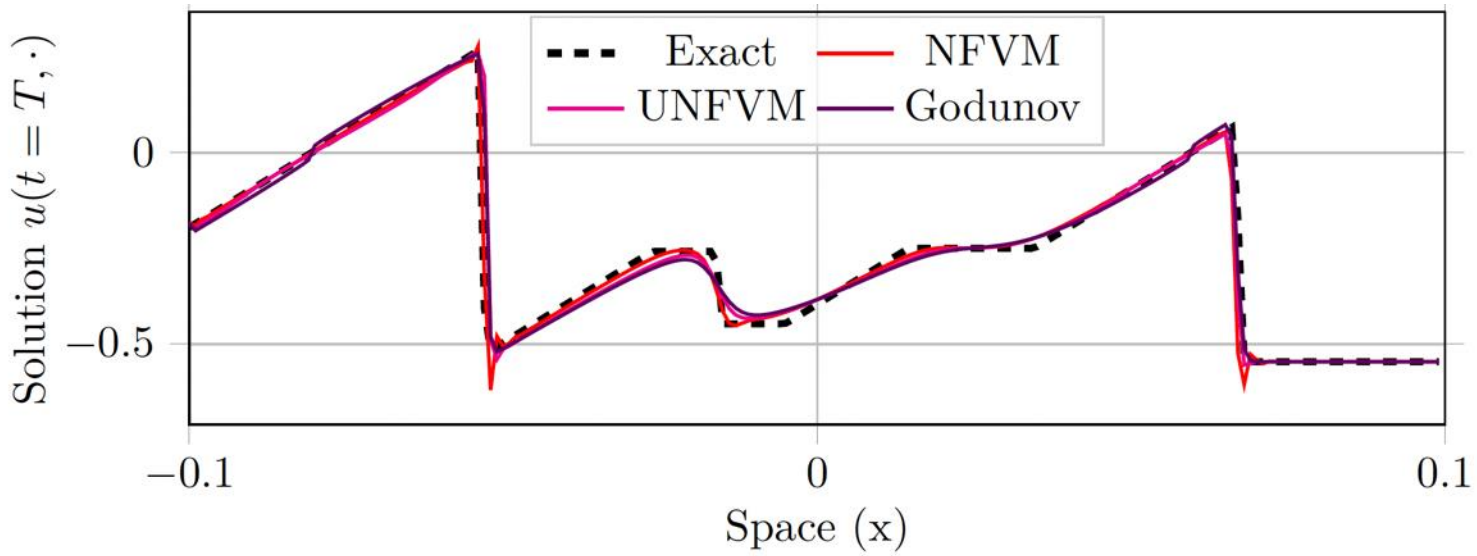
Predictive capabilities of the NN approach



(a) Log-Linear scale. (b) Log-Log scale.

Figure: **Convergence plot for the Greenshield's flux.** Standard deviation is shown as error bars. The ratio $\Delta t/\Delta x = 0.1$ is fixed.

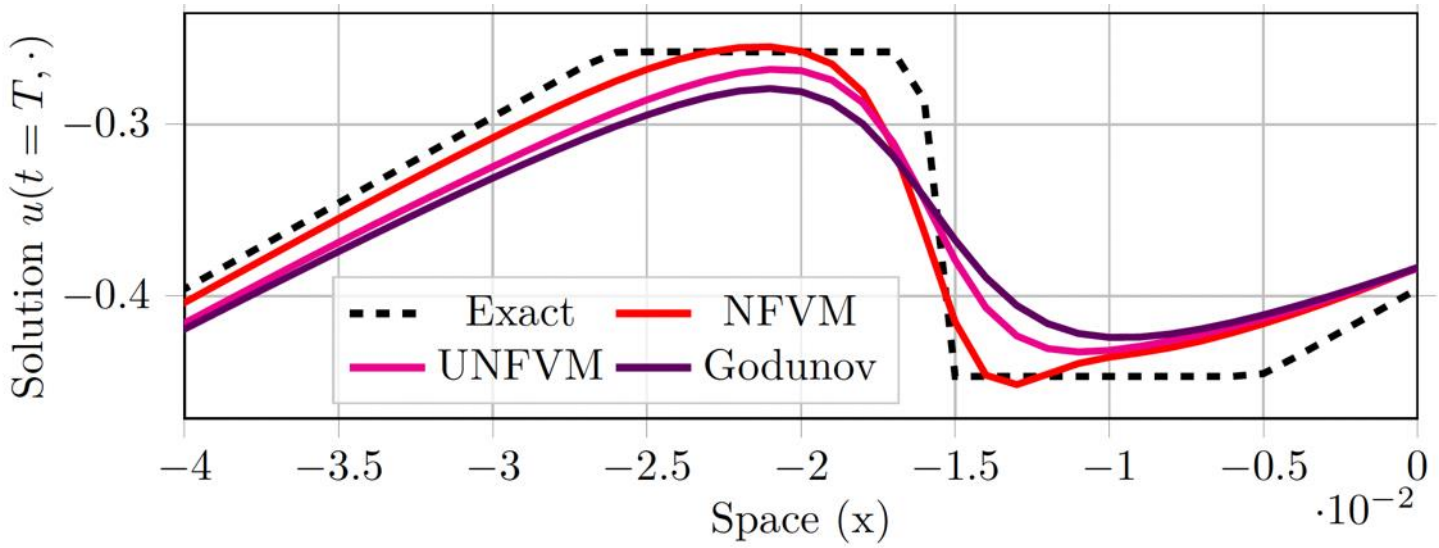
Local features of solutions



	1st order FVM				Higher order FVM		FEM
	NFVM	**UNFVM**	GD	EO	ENO	WENO	DG
L1	$\mathbf{1.2e^{-2}}_{\pm 2e^{-3}}$	$1.6e^{-2}_{\pm 3e^{-3}}$	$2.1e^{-2}_{\pm 4e^{-3}}$	$2.1e^{-2}_{\pm 4e^{-3}}$	$1.9e^{-2}_{\pm 7e^{-3}}$	$2.0e^{-4}_{\pm 7e^{-3}}$	$1.9e^{-3}_{\pm 4e^{-4}}$
L2	$\mathbf{8.5e^{-4}}_{\pm 3e^{-4}}$	$1.3e^{-3}_{\pm 6e^{-4}}$	$1.9e^{-3}_{\pm 7e^{-4}}$	$1.8e^{-3}_{\pm 6e^{-4}}$	$2.7e^{-3}_{\pm 1e^{-3}}$	$2.8e^{-3}_{\pm 1e^{-3}}$	$1.0e^{-4}_{\pm 4e^{-5}}$
Rel.	$\mathbf{5.9e^{-2}}_{\pm 1e^{-2}}$	$8.5e^{-2}_{\pm 2e^{-2}}$	$1.6e^{-1}_{\pm 4e^{-2}}$	$1.5e^{-1}_{\pm 4e^{-2}}$	$1.5e^{-1}_{\pm 5e^{-2}}$	$1.5e^{-2}_{\pm 5e^{-2}}$	$1.1e^{-2}_{\pm 5e^{-3}}$

Table: Results on Burgers with piecewise-constant initial conditions.

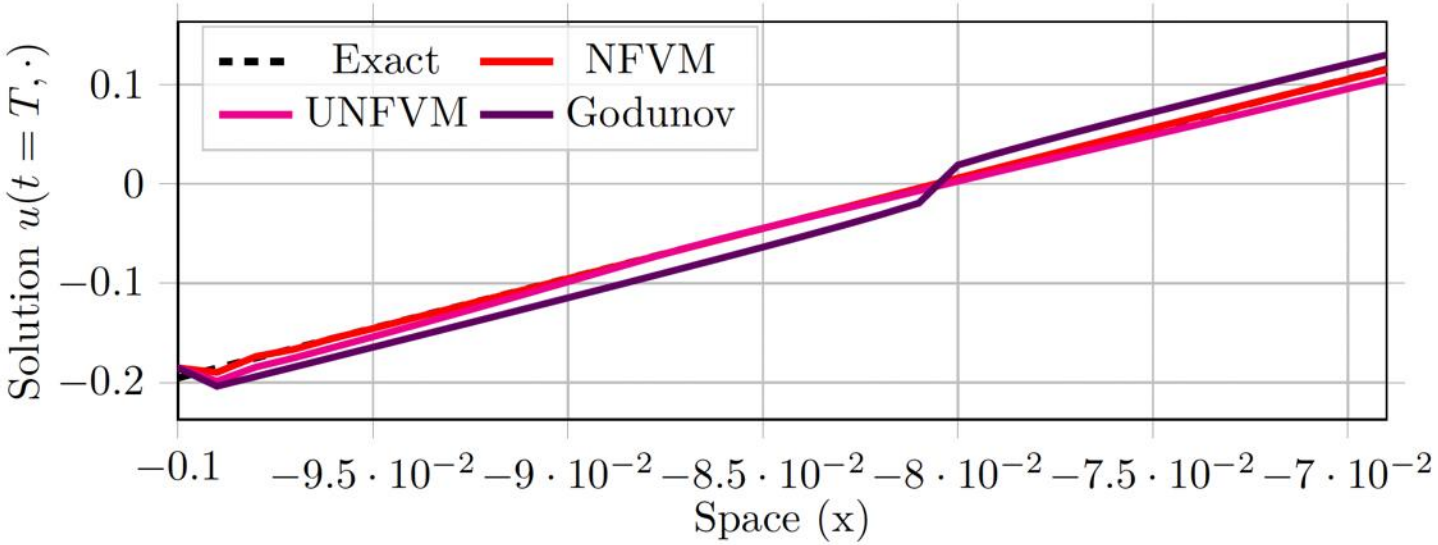
Local features of solutions



	1st order FVM				Higher order FVM		FEM
	NFVM	**UNFVM**	GD	EO	ENO	WENO	DG
L1	$\mathbf{1.2e^{-2}}_{\pm 2e^{-3}}$	$1.6e^{-2}_{\pm 3e^{-3}}$	$2.1e^{-2}_{\pm 4e^{-3}}$	$2.1e^{-2}_{\pm 4e^{-3}}$	$1.9e^{-2}_{\pm 7e^{-3}}$	$2.0e^{-4}_{\pm 7e^{-3}}$	$1.9e^{-3}_{\pm 4e^{-4}}$
L2	$\mathbf{8.5e^{-4}}_{\pm 3e^{-4}}$	$1.3e^{-3}_{\pm 6e^{-4}}$	$1.9e^{-3}_{\pm 7e^{-4}}$	$1.8e^{-3}_{\pm 6e^{-4}}$	$2.7e^{-3}_{\pm 1e^{-3}}$	$2.8e^{-3}_{\pm 1e^{-3}}$	$1.0e^{-4}_{\pm 4e^{-5}}$
Rel.	$\mathbf{5.9e^{-2}}_{\pm 1e^{-2}}$	$8.5e^{-2}_{\pm 2e^{-2}}$	$1.6e^{-1}_{\pm 4e^{-2}}$	$1.5e^{-1}_{\pm 4e^{-2}}$	$1.5e^{-1}_{\pm 5e^{-2}}$	$1.5e^{-2}_{\pm 5e^{-2}}$	$1.1e^{-2}_{\pm 5e^{-3}}$

Table: Results on Burgers with piecewise-constant initial conditions.

Local features of solutions



	1st order FVM				Higher order FVM		FEM
	NFVM	**UNFVM**	GD	EO	ENO	WENO	DG
L1	$\mathbf{1.2e^{-2}}_{\pm 2e^{-3}}$	$1.6e^{-2}_{\pm 3e^{-3}}$	$2.1e^{-2}_{\pm 4e^{-3}}$	$2.1e^{-2}_{\pm 4e^{-3}}$	$1.9e^{-2}_{\pm 7e^{-3}}$	$2.0e^{-4}_{\pm 7e^{-3}}$	$1.9e^{-3}_{\pm 4e^{-4}}$
L2	$\mathbf{8.5e^{-4}}_{\pm 3e^{-4}}$	$1.3e^{-3}_{\pm 6e^{-4}}$	$1.9e^{-3}_{\pm 7e^{-4}}$	$1.8e^{-3}_{\pm 6e^{-4}}$	$2.7e^{-3}_{\pm 1e^{-3}}$	$2.8e^{-3}_{\pm 1e^{-3}}$	$1.0e^{-4}_{\pm 4e^{-5}}$
Rel.	$\mathbf{5.9e^{-2}}_{\pm 1e^{-2}}$	$8.5e^{-2}_{\pm 2e^{-2}}$	$1.6e^{-1}_{\pm 4e^{-2}}$	$1.5e^{-1}_{\pm 4e^{-2}}$	$1.5e^{-1}_{\pm 5e^{-2}}$	$1.5e^{-2}_{\pm 5e^{-2}}$	$1.1e^{-2}_{\pm 5e^{-3}}$

Table: Results on Burgers with piecewise-constant initial conditions.

Complexity

NFVM architecture: 1D CNN

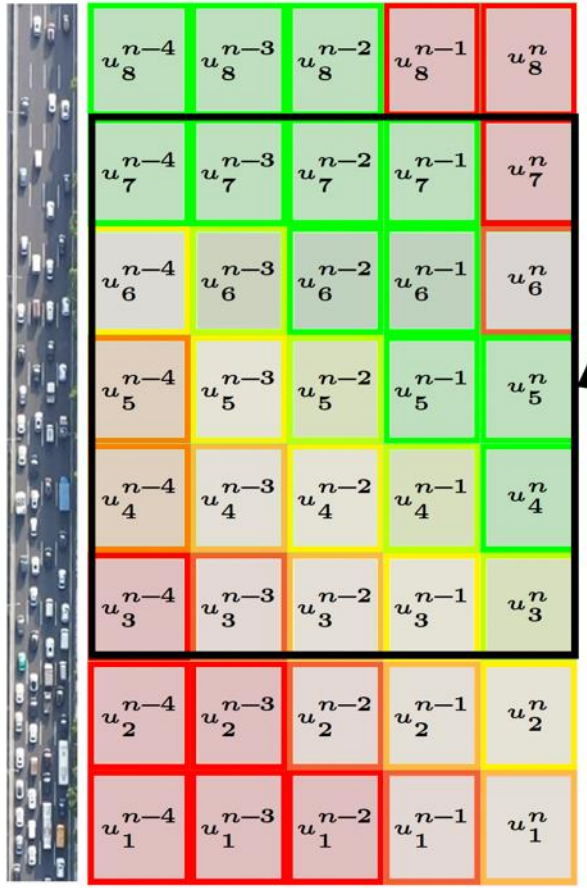
- Input size (space): 2 cells (same as 1st order FVMs)
- Input size (time): 1 timestep (same as classical FVMs)
- Layers: 6 hidden layers of size 15
- Learnt parameters: ≈ 1200

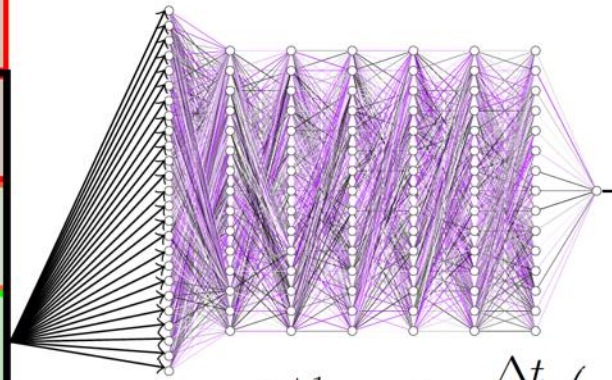
NFVM_5^5 architecture: 1D CNN

- Input size (space): **5 cells**
- Input size (time): **5 timesteps**
- Layers: 6 hidden layers of size 15
- Learnt parameters: ≈ 1600

The CNN-based architecture can easily be scaled up to larger inputs; designing explicit numerical methods with several input timesteps is much more complex.

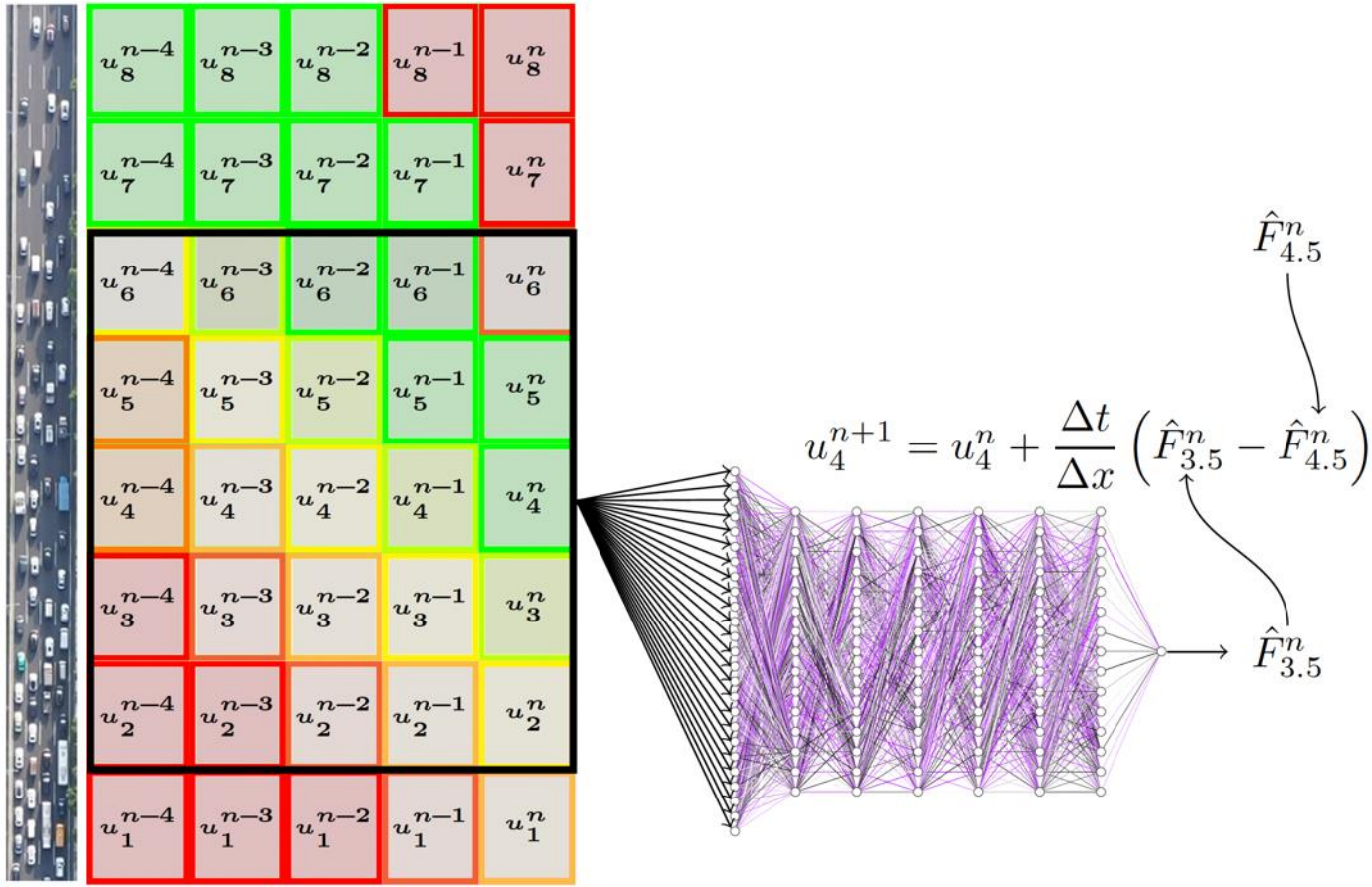
Unleashing stencils



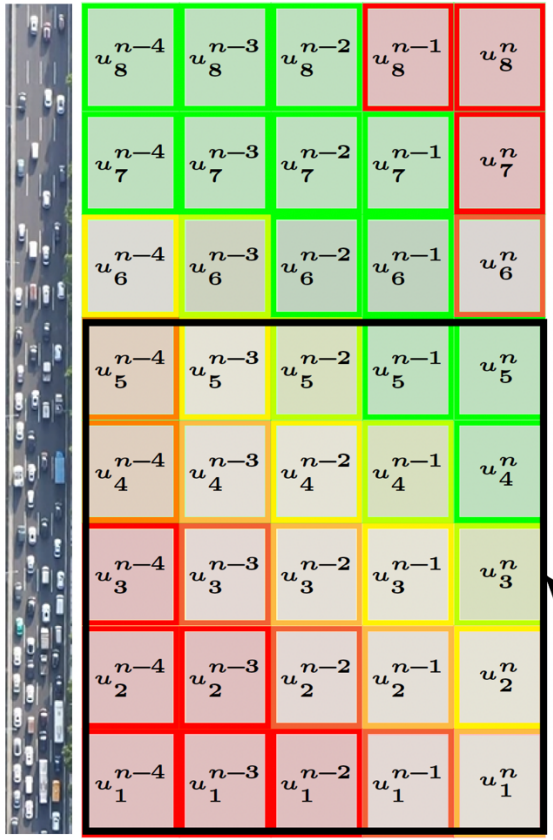


$$u_4^{n+1} = u_4^n + \frac{\Delta t}{\Delta x}\left(\quad - \hat{F}_{4.5}^n \right)$$

Unleashing stencils



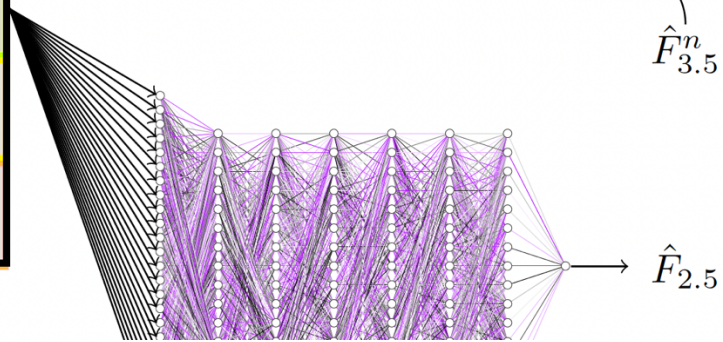
Unleashing stencils



$$u_4^{n+1} = u_4^n + \frac{\Delta t}{\Delta x}\left(\hat{F}_{3.5}^n - \hat{F}_{4.5}^n\right)$$

$\hat{F}_{4.5}^n$

$\hat{F}_{3.5}^n$



Local features of the 5X5 stencil

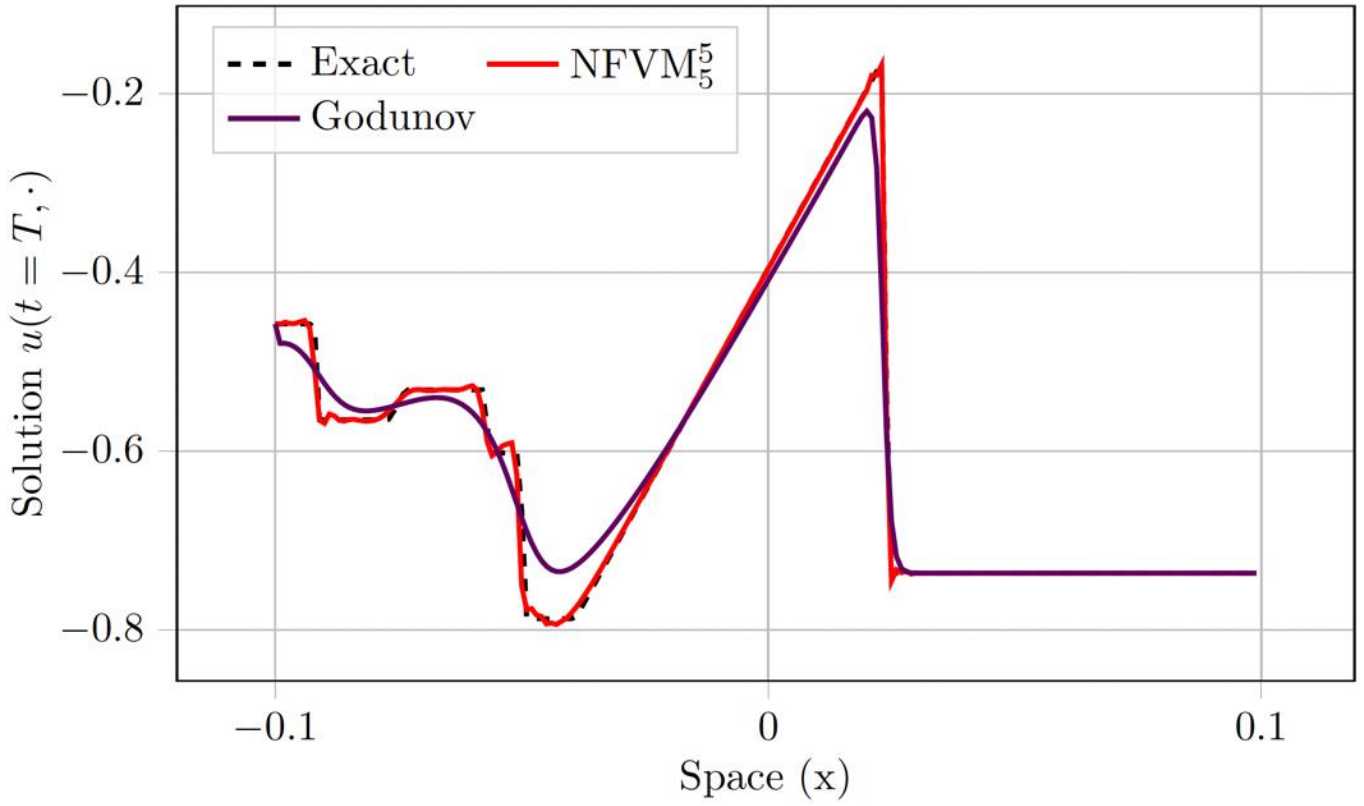


Figure: Burgers.

Local features of the 5X5 stencil

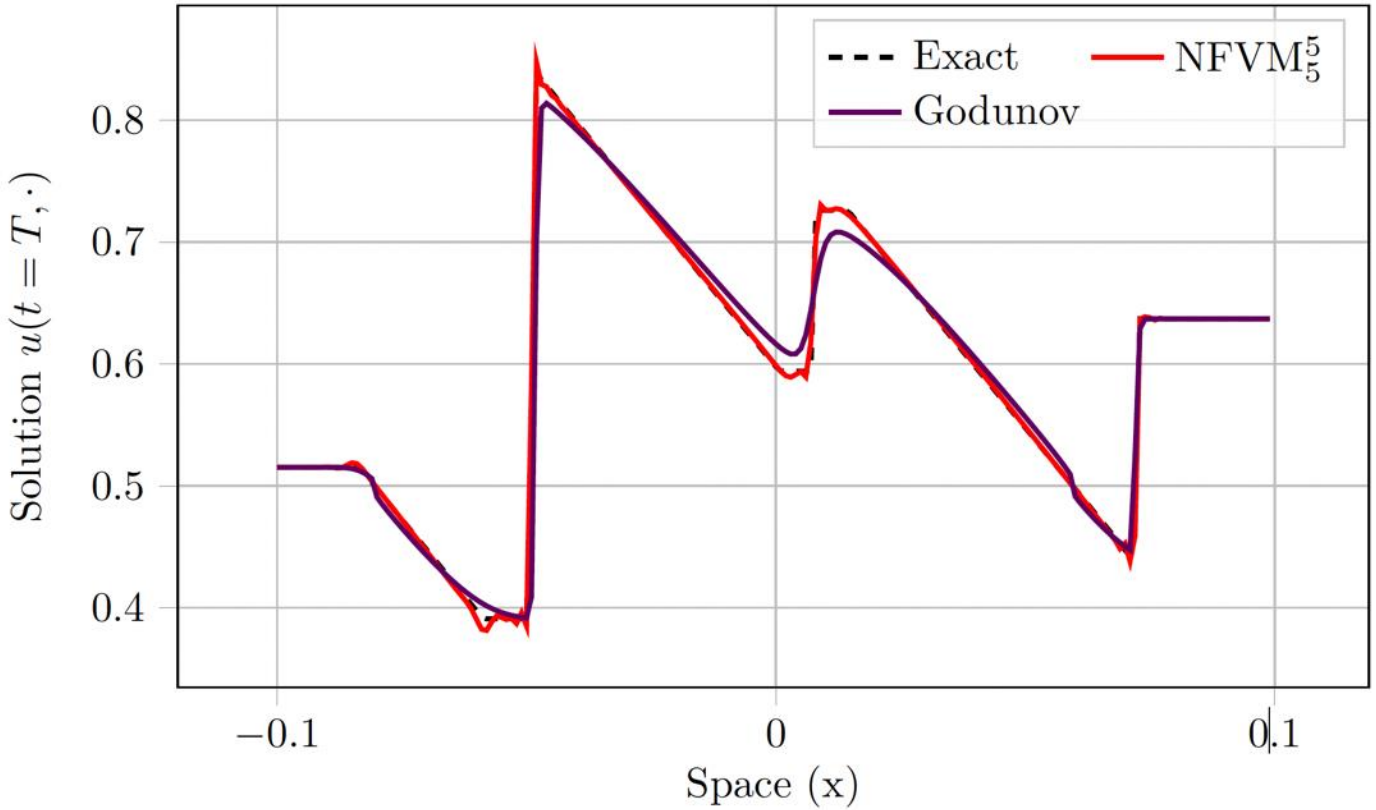


Figure: LWR with Greenshields flow.

Local features of the 5X5 stencil

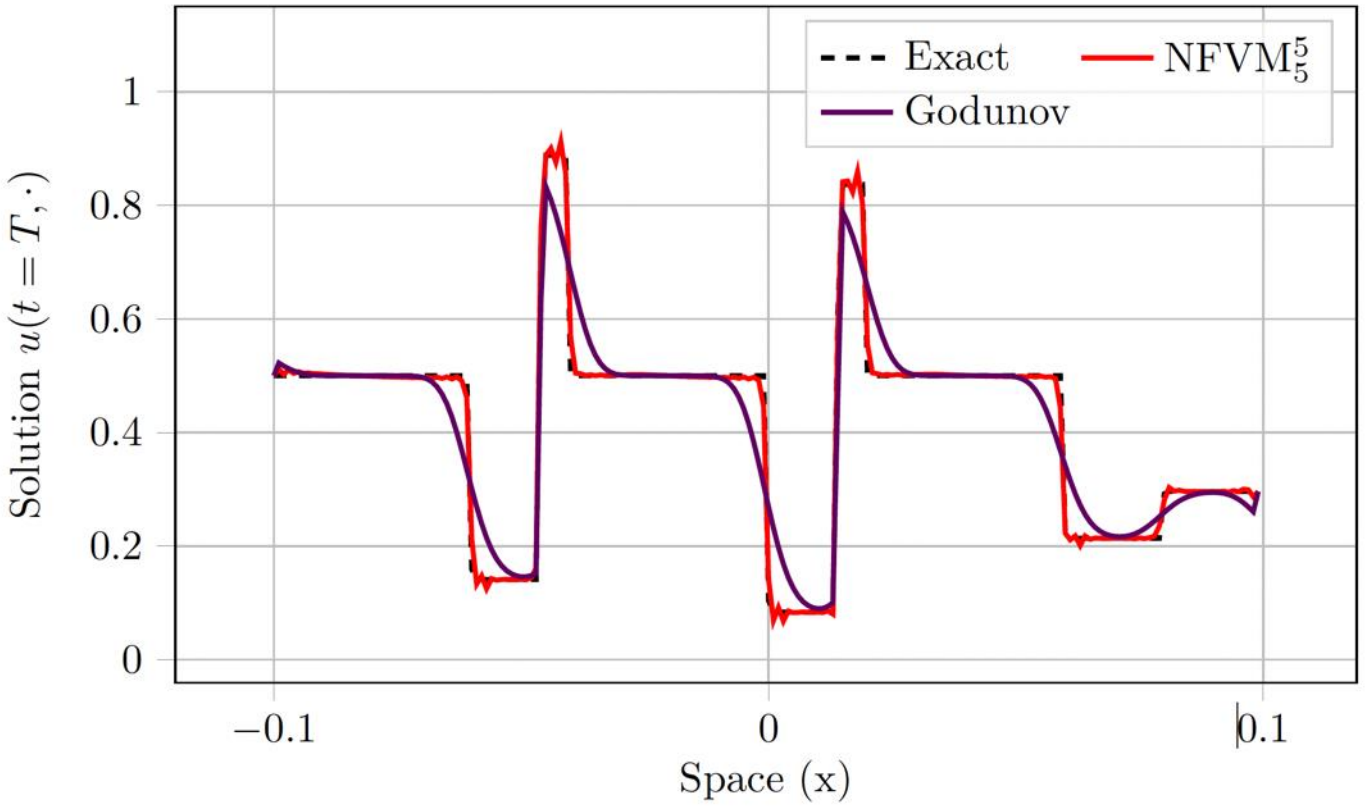
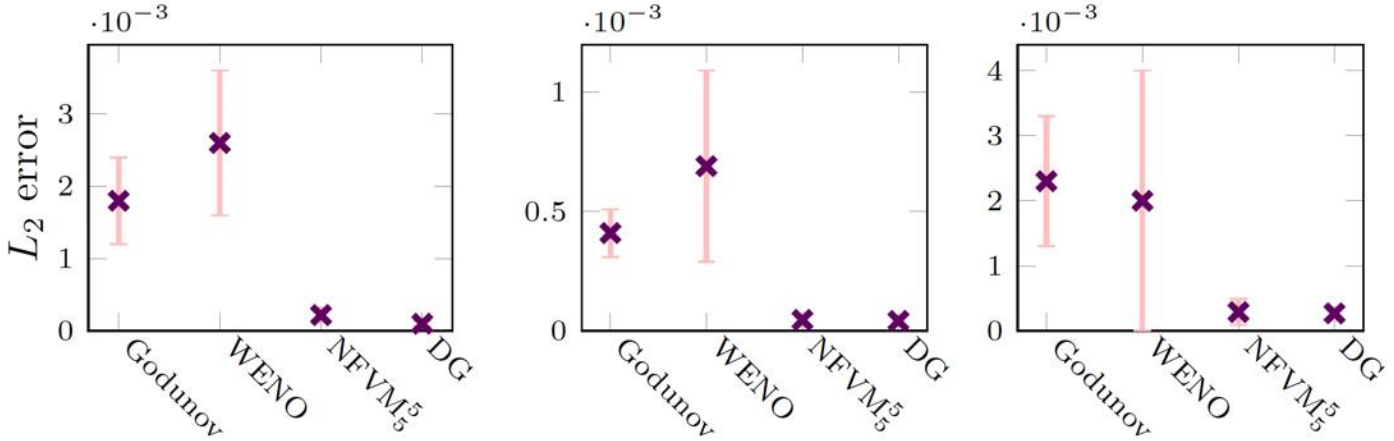


Figure: LWR with Triangular flow.

Numerical errors



(a) Burgers. (b) Greenshield's. (c) Triangular.

	Godunov	WENO	**$NFVM_5^5$**	DG
Burgers (L2)	$1.8e^{-3}_{\pm 6e^{-4}}$	$2.6e^{-3}_{\pm 1e^{-3}}$	$2.2e^{-4}_{\pm 1e^{-4}}$	$\mathbf{1.0e^{-4}}_{\pm 4e^{-5}}$
Greenshield's (L2)	$4.1e^{-4}_{\pm 1e^{-4}}$	$6.9e^{-4}_{\pm 4e^{-4}}$	$4.6e^{-5}_{\pm 3e^{-5}}$	$\mathbf{4.2e^{-5}}_{\pm 2e^{-5}}$
Triangular (L2)	$2.2e^{-3}_{\pm 1e^{-3}}$	$2.0e^{-3}_{\pm 2e^{-3}}$	$2.9e^{-4}_{\pm 2e^{-4}}$	$\mathbf{2.7e^{-4}}_{\pm 1e^{-4}}$

Supervised learning on field data

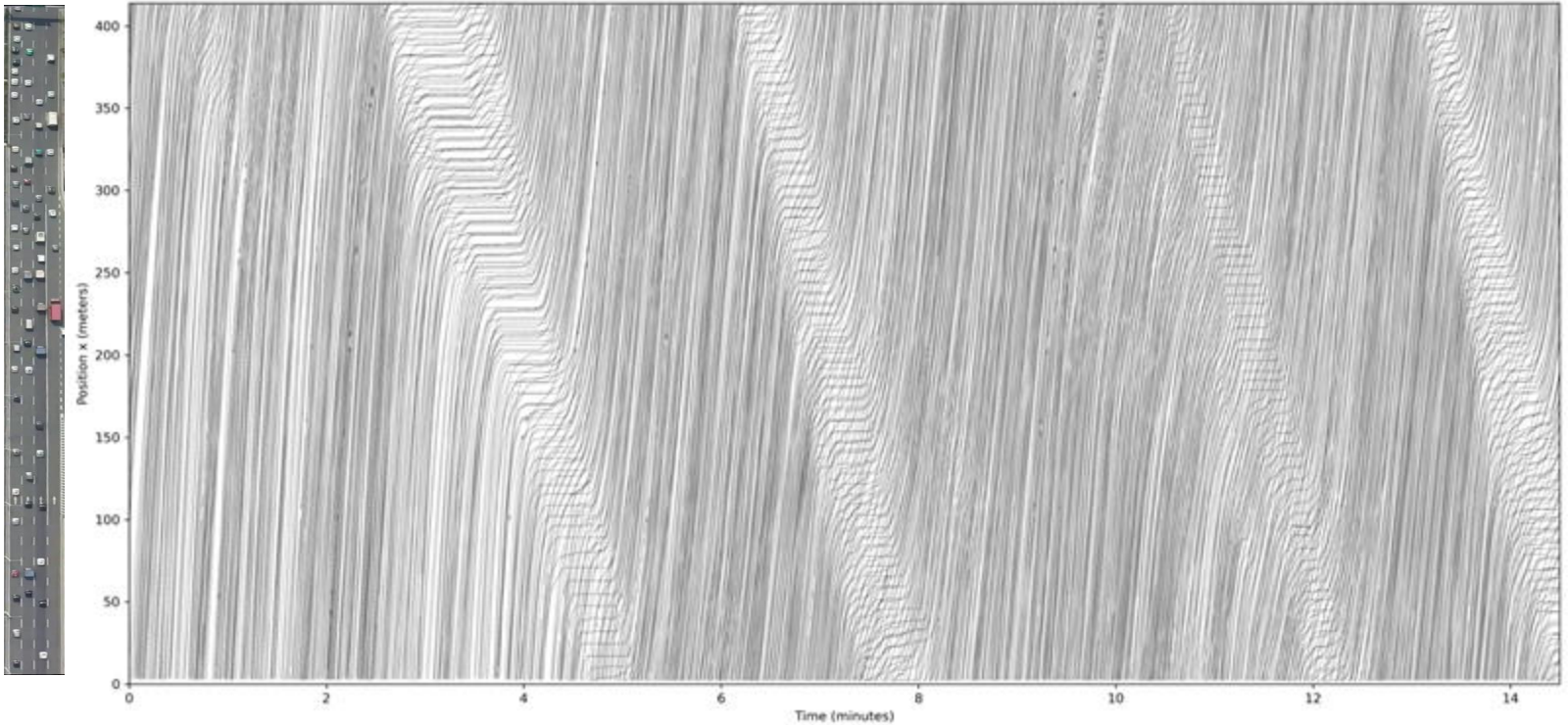


Figure: Trajectories are extracted from drone-recorded highway video data.

- 12 videos recorded by a drone above a highway (4K, 30 FPS).
- Videos annotated using a vehicle detection and tracking algorithm.
- 15 minutes of data.
- 4 visible stop-and-go waves.

Fangyu Wu et al. "Decentralized vehicle coordination: The Berkeley deepdrive drone dataset". In: *arXiv preprint arXiv:2209.08763* (2022).

Drone data set



Training and validation set

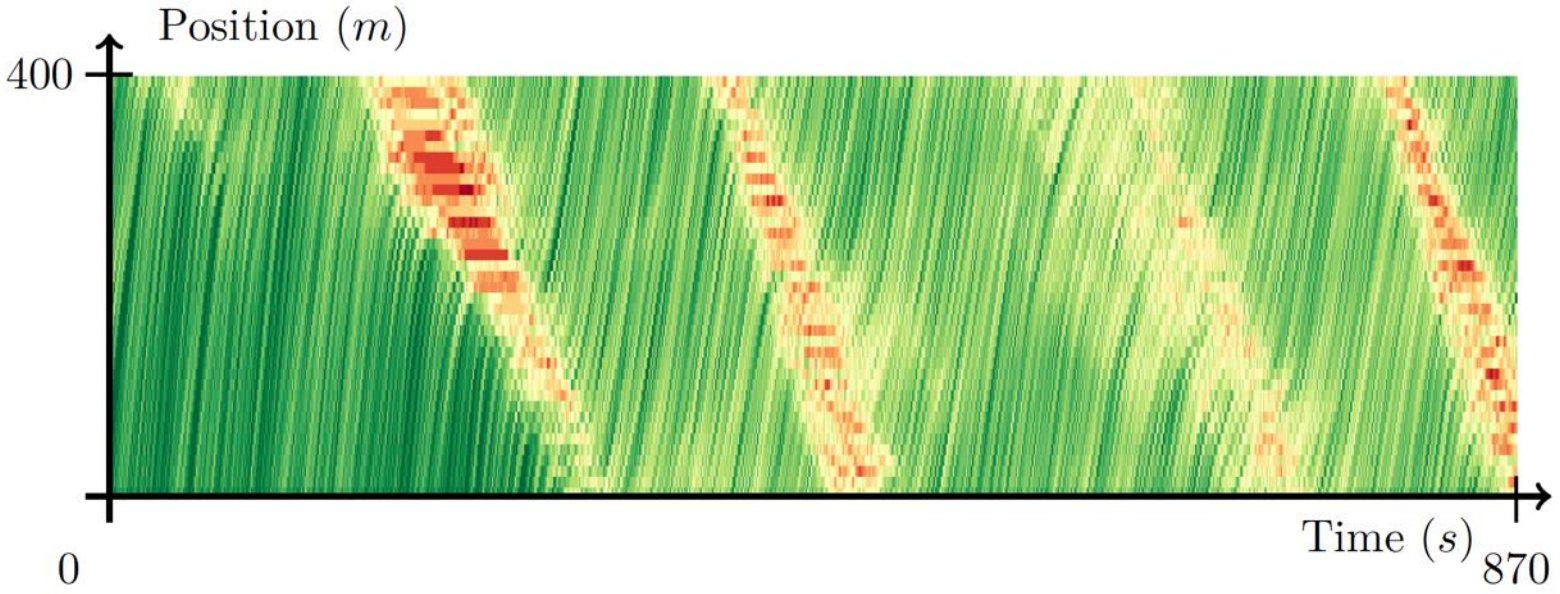


Figure: Time-space diagram of vehicle trajectories extracted from the drone video, color-coded by density.

Training and validation set

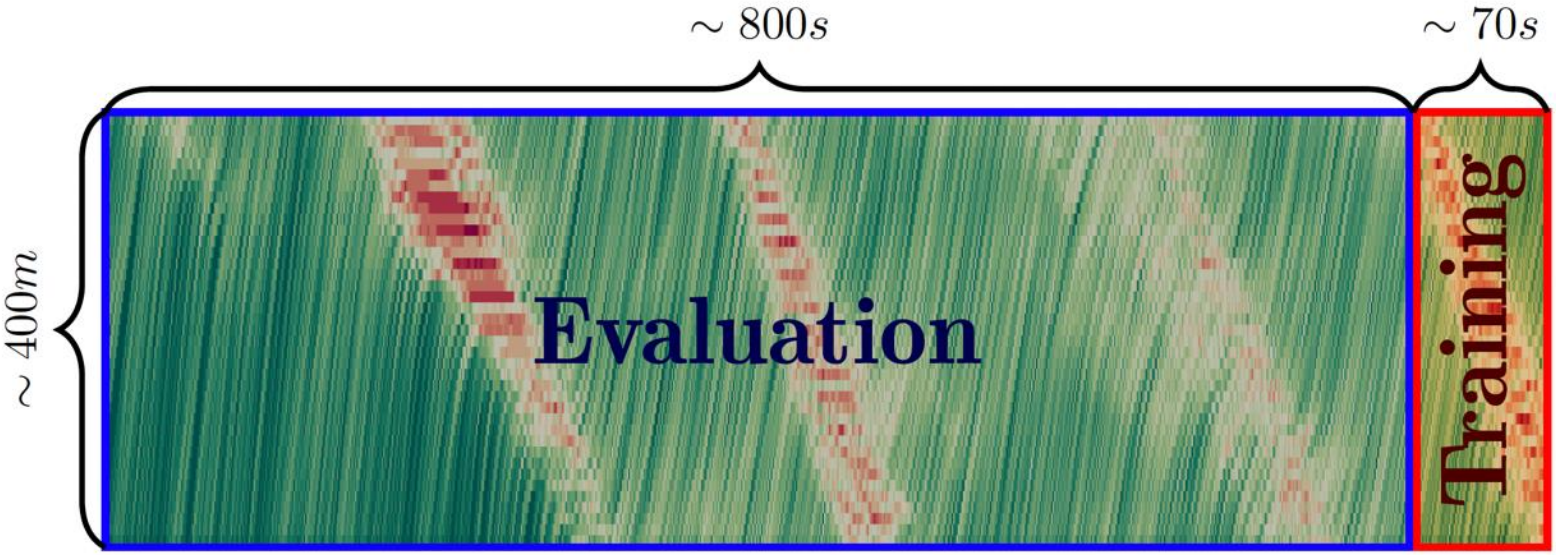


Figure: Time-space diagram of vehicle trajectories extracted from the drone video, color-coded by density.

Training and validation set

- Flow f_ϕ (e.g. Greenshields, Triangular...) with parameters ϕ ($v_{\max}, k_{max}$...).
- Numerical scheme ψ (e.g. Godunov, ENO...).
- Let $\hat{u}_{\psi,f_\phi}$ be the forward prediction using scheme ψ and flow f_ϕ.
- Optimal flow parameter are found by solving the following problem:

$$\arg\min_{\phi} \|u - \hat{u}_{\psi,f_\phi}\|_2.$$

  where u is the ground truth (training set).
- We solve this problem by grid search for each of 5 flows and 5 schemes, resulting in 25 different predictions.

Forward computation with regular FVM schemes

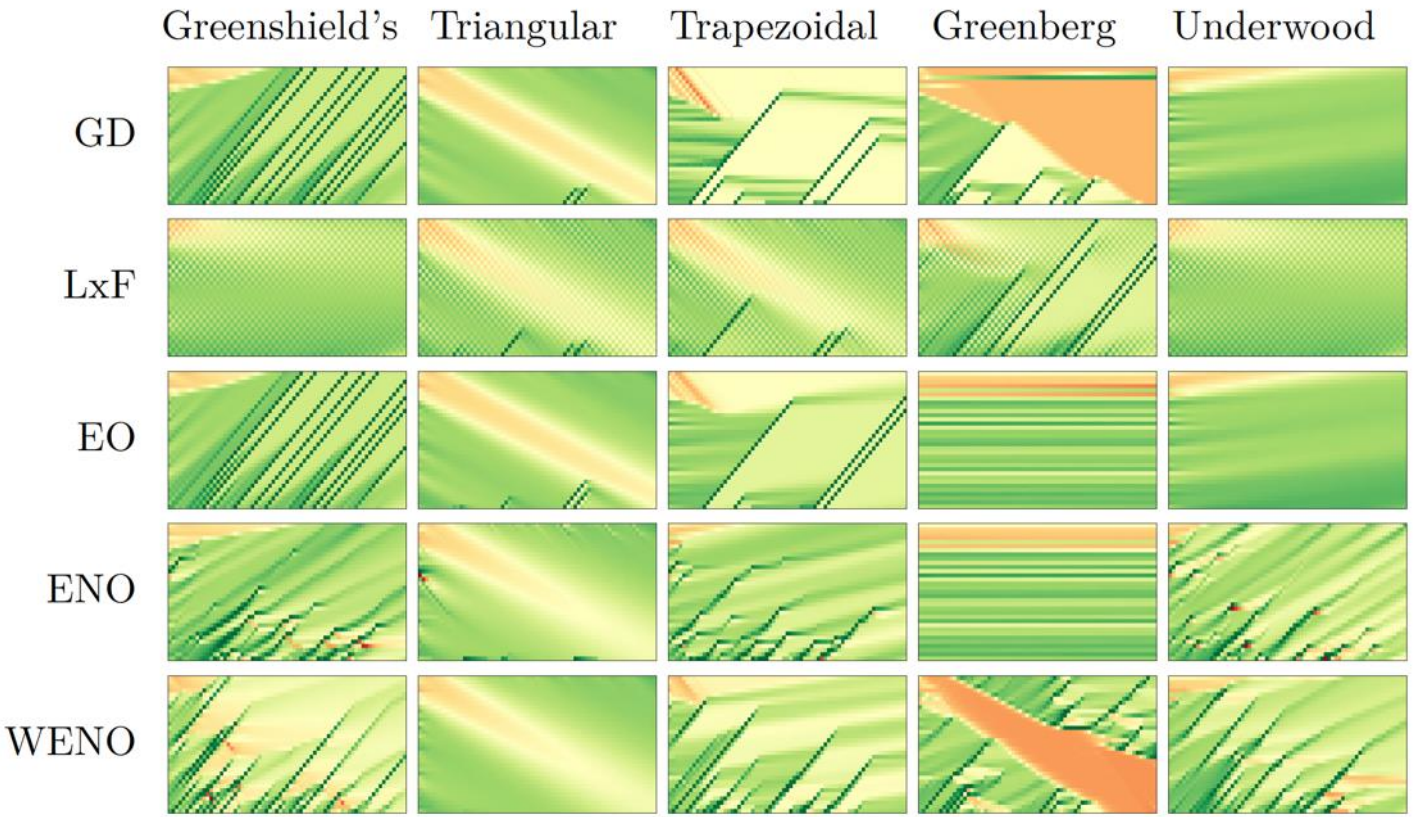


Figure: Forward prediction on training data with best flow parameters.

Forward computation with regular FVM schemes

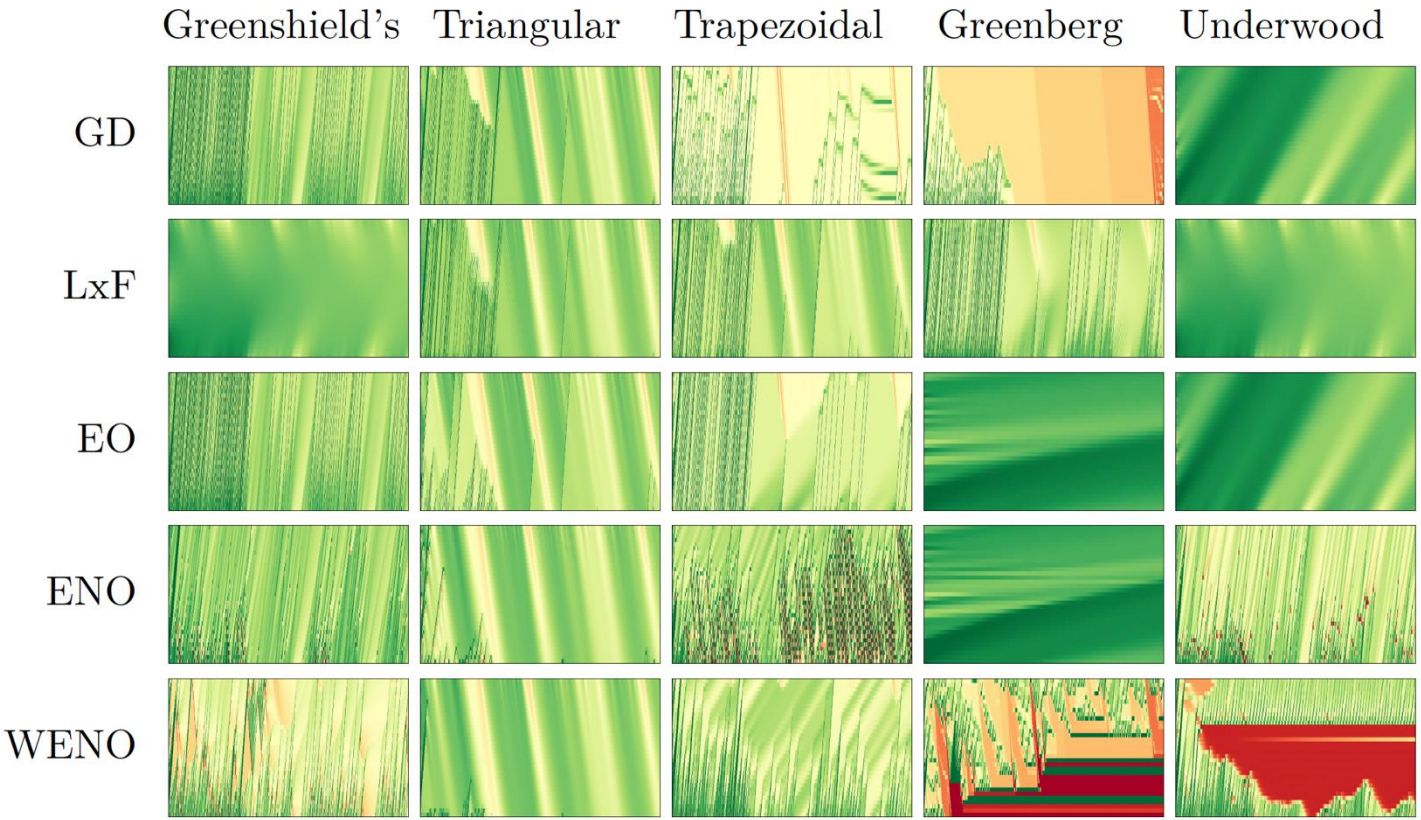
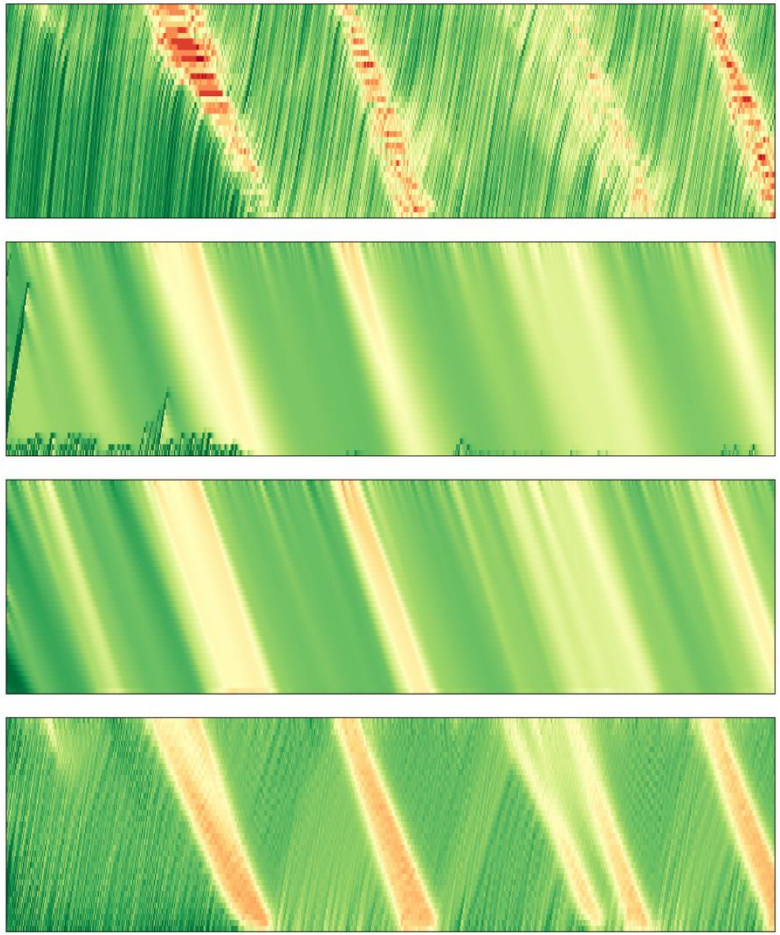


Figure: Forward prediction on the whole dataset data with best flow parameters.

Winners comparison



Future work

- The first natural generalization of scalar conservation laws equations.
- Considering the problem of the form:

$$\begin{cases} \partial_t \rho(t,x) + \partial_x [f(\gamma(x), \rho(t,x))] = 0, \\ \rho(0,x) = \rho_\circ(x). \end{cases}$$

- Why this PDE:
  - The Riemann solution is not as simple.
  - Loses one of the main requirements: No bounded variation!!

$$\begin{aligned} \partial_t \rho + \partial_x [f(\gamma, \rho)] &= 0 \\ \partial_t \gamma &= 0 \end{aligned}$$

The Jacobian is degenerate (coinciding eigenvalues). The system is resonant and hence no total variation.

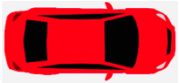
  - Can we still learn the entropy solution?

What is mixed autonomy?

The coexistence of CAVs and human driven vehicles



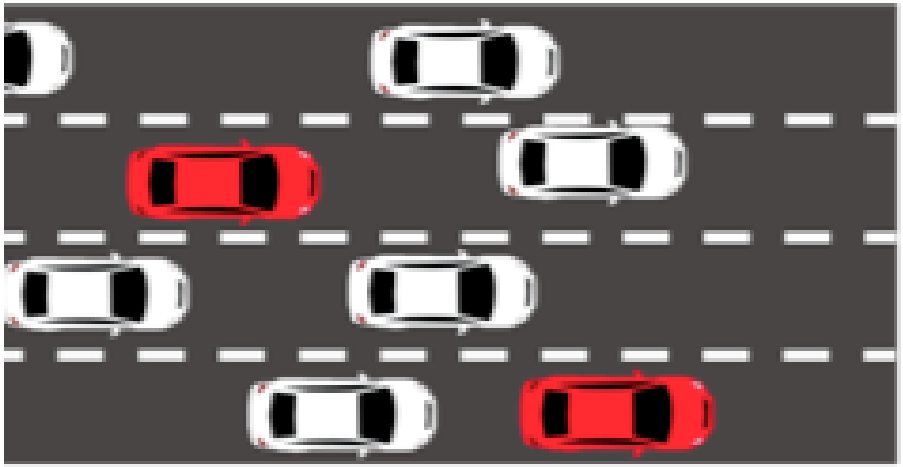
Prof. Cathy Wu, MIT

CAV: 

Autonomous,
Onboard policy (learned)
Connected to other CAVs

Other vehicle 

Following human dynamics
(car following model)

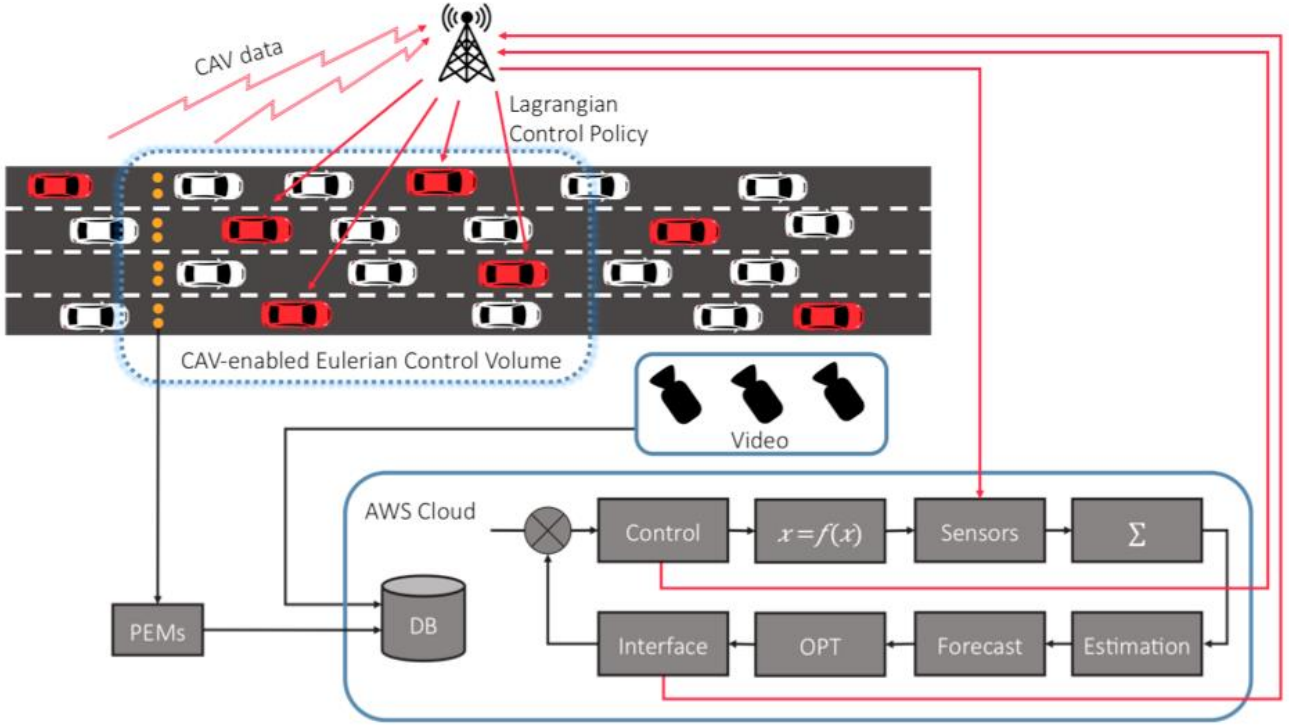


How can mixed autonomy be used?

CAVs can be used to control traffic

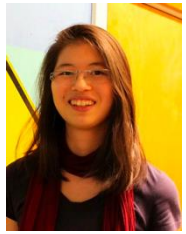


Prof. Cathy Wu, MIT

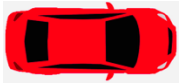


Local sensing

CAVs can sense vehicles around them



Prof. Cathy Wu, MIT

CAV: 

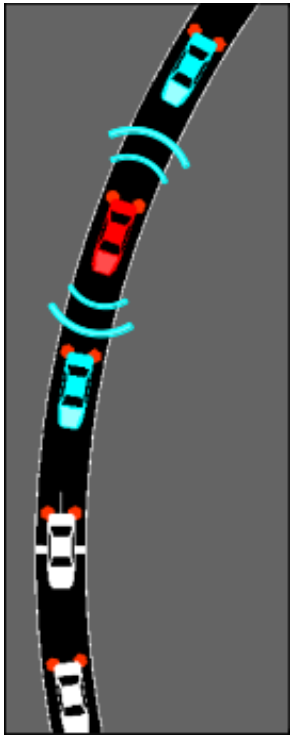
Autonomous,
Onboard policy (learned)
Connected to other CAVs

Sensed vehicle: 

Sensed by CAV proximity
Or other [C]onnected vehicle

Other vehicle 

Following human dynamics
(car following model)



Longitudinal control



Prof. C. Wu,
MIT

SCENARIO 1: Single Lane(4 AVs, 7 Human)

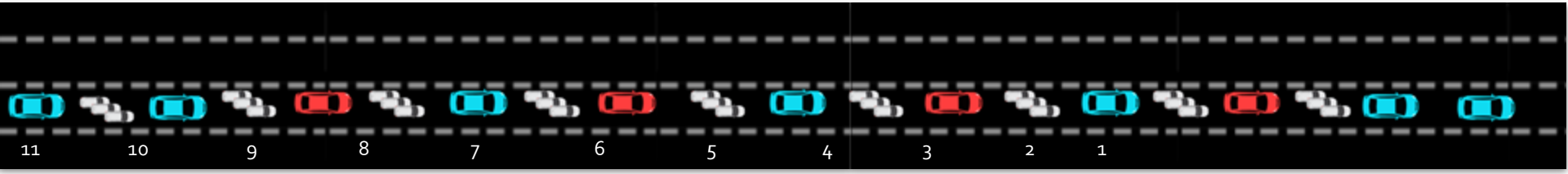
All tests on Day 2 (Wednesday), and most test on Day 3 (Friday)



Rental 11, Rental10, VU-1, Rental 8, UCB-2 , Rental 6, UCB-1, Rental 4, AZ-3, Rental 2, Rental 1

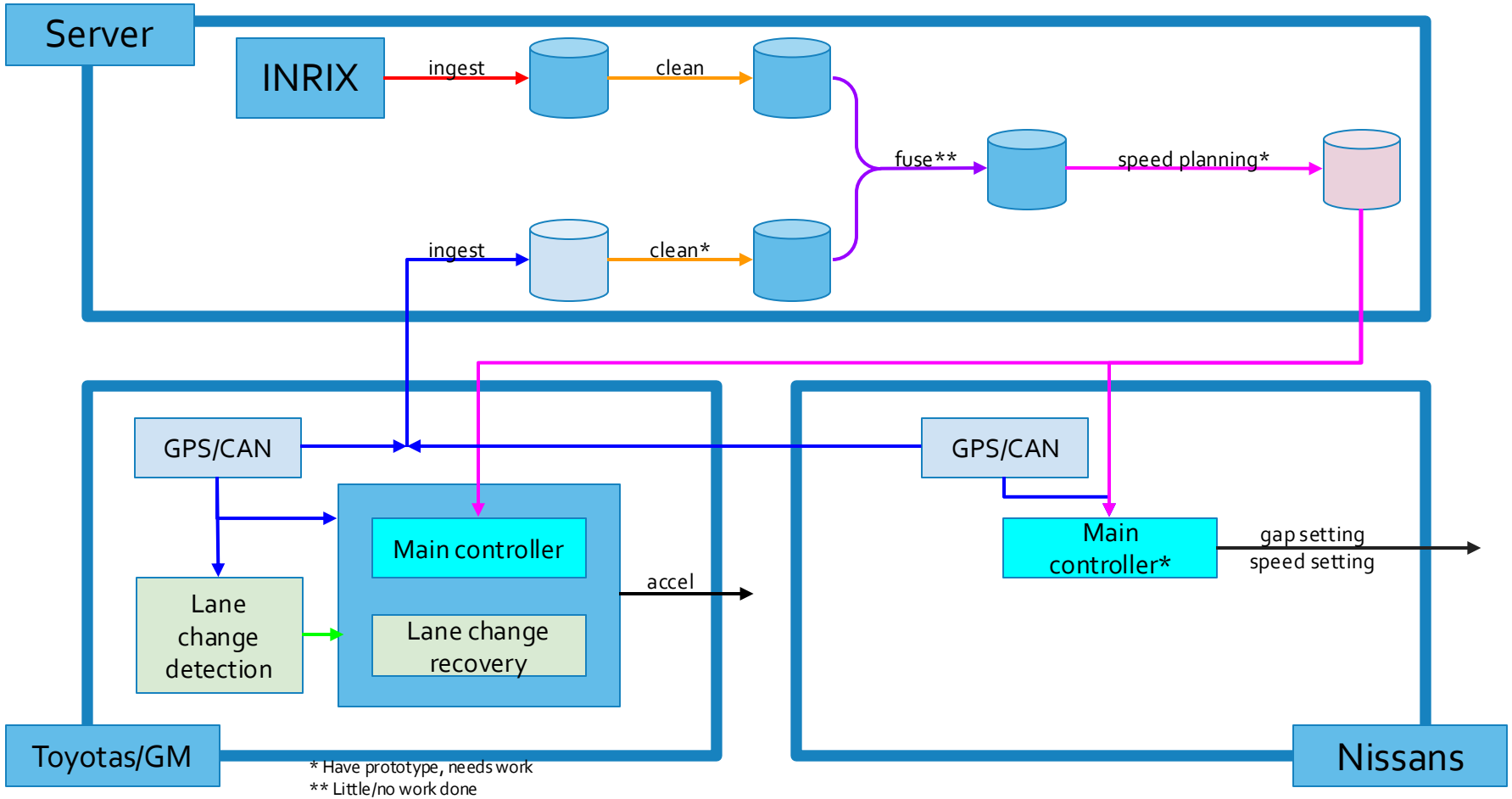
SCENARIO 1: Single Lane(4 AVs, 7 Human)

Last Test on Day 3 (Friday) was on Lane 3

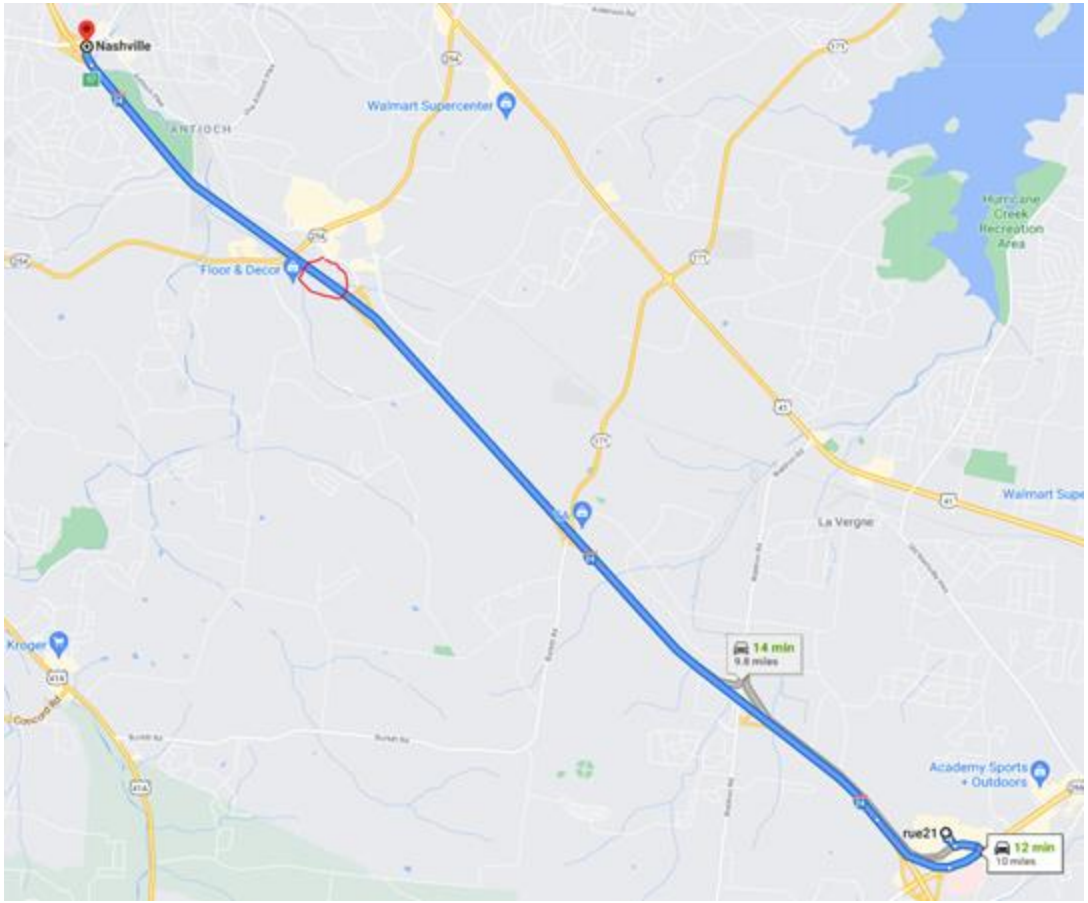


Rental 11, Rental10, VU-1, Rental 8, UCB-2 , Rental 6, UCB-1, Rental 4, AZ-3, Rental 2, Rental 1

MegaController Overview



VanderTest: Testbed, exit 66 to exit 57 (Summer 2021)

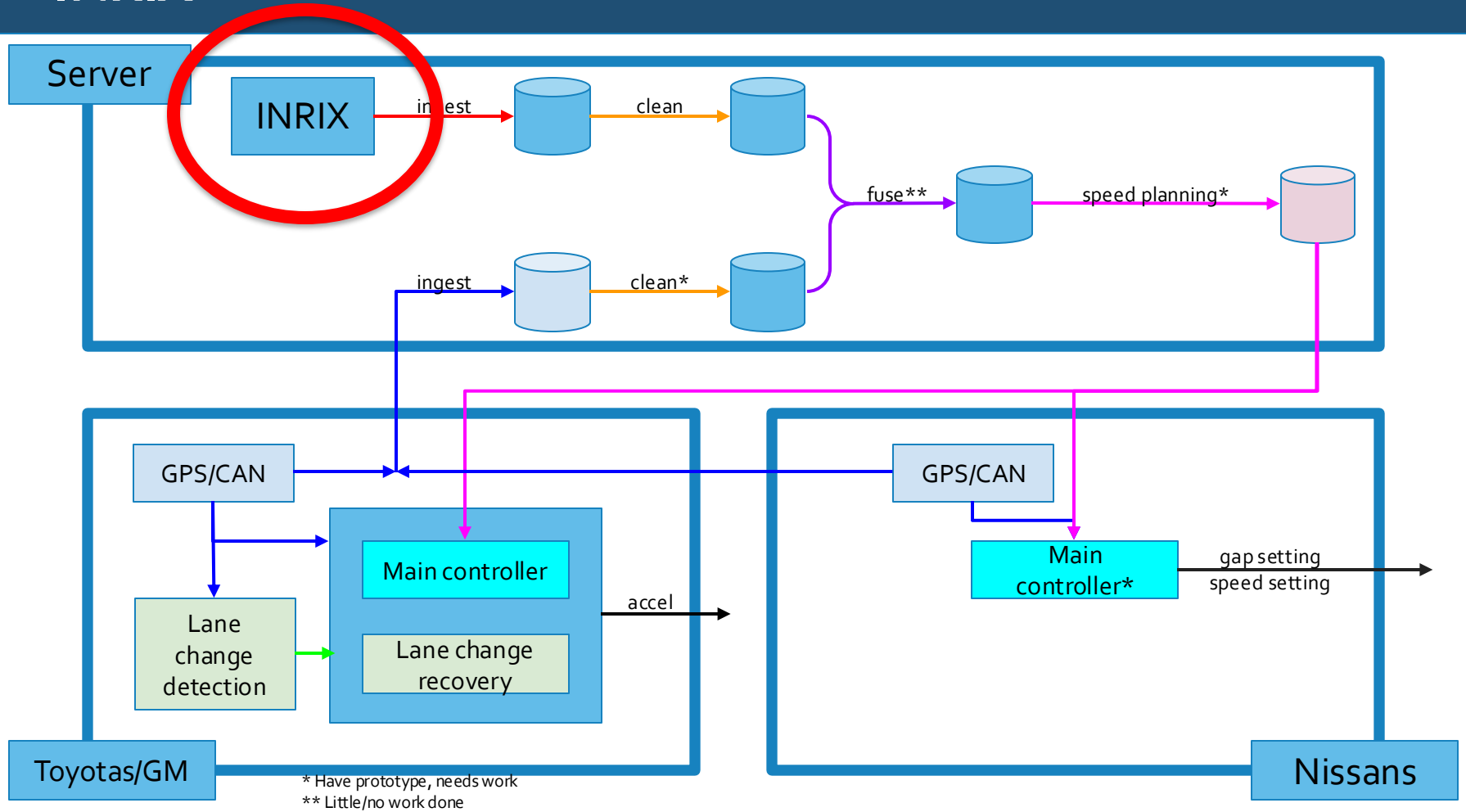


Our test ran from a retail space near "Rue 21." We entered I-24 at exit 66 and this ran Westbound to Haywood lane (exit 57). The red line shows the location of Murad piloting the drone. This was between exit 60 and 59.

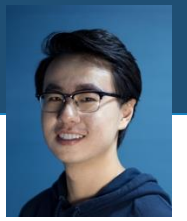
Initially, we signaled Murad with a text from the passenger of the lead vehicle as he passed mile 64; it was determined that we should signal sooner.

We started letting Murad know as we entered the road at mile 66

INRIX


Server
INRIX
ingest
clean
fuse**
speed planning*
ingest
clean*
GPS/CAN
Main controller
Lane change detection
Lane change recovery
accel
GPS/CAN
Main controller*
gap setting
speed setting
Toyotas/GM
Nissans
* Have prototype, needs work
** Little/no work done

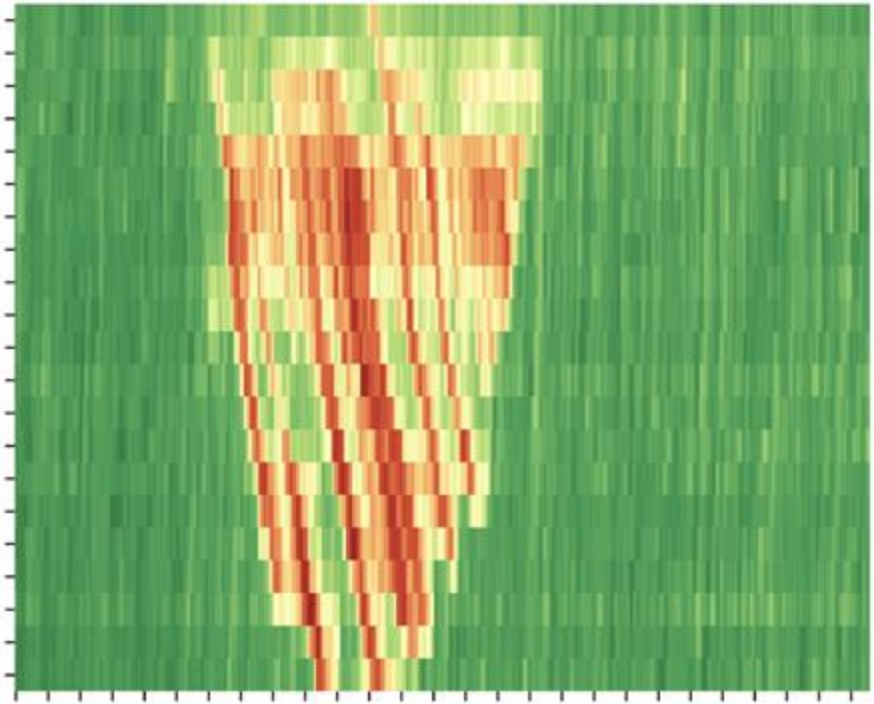
Bottleneck Identification +



H. Wang
UC Berkeley

Distinguish shockwave and standing bottleneck - Real INRIX

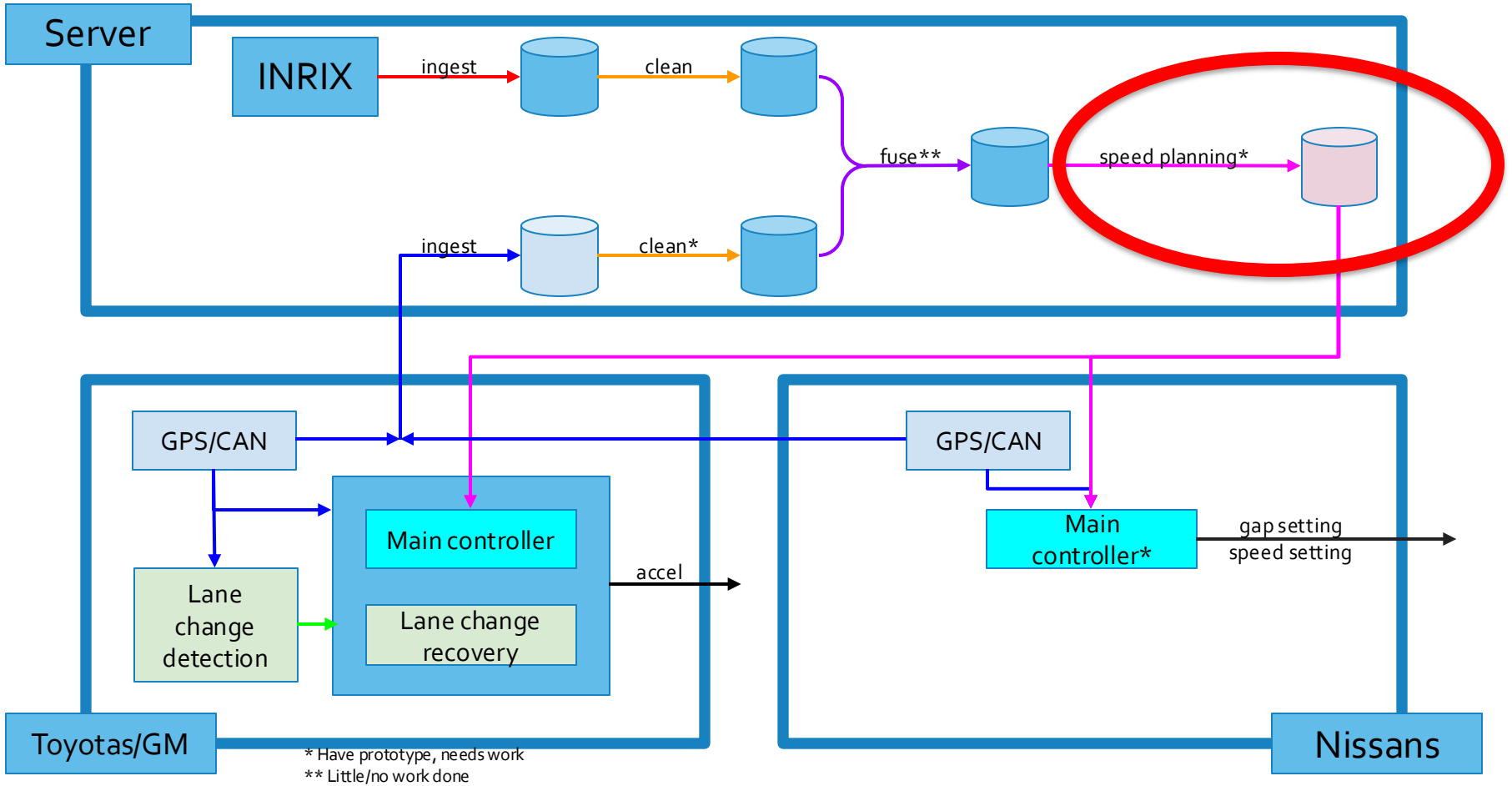
INRIX on July 28



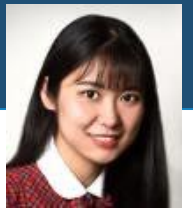
Bottleneck Identification



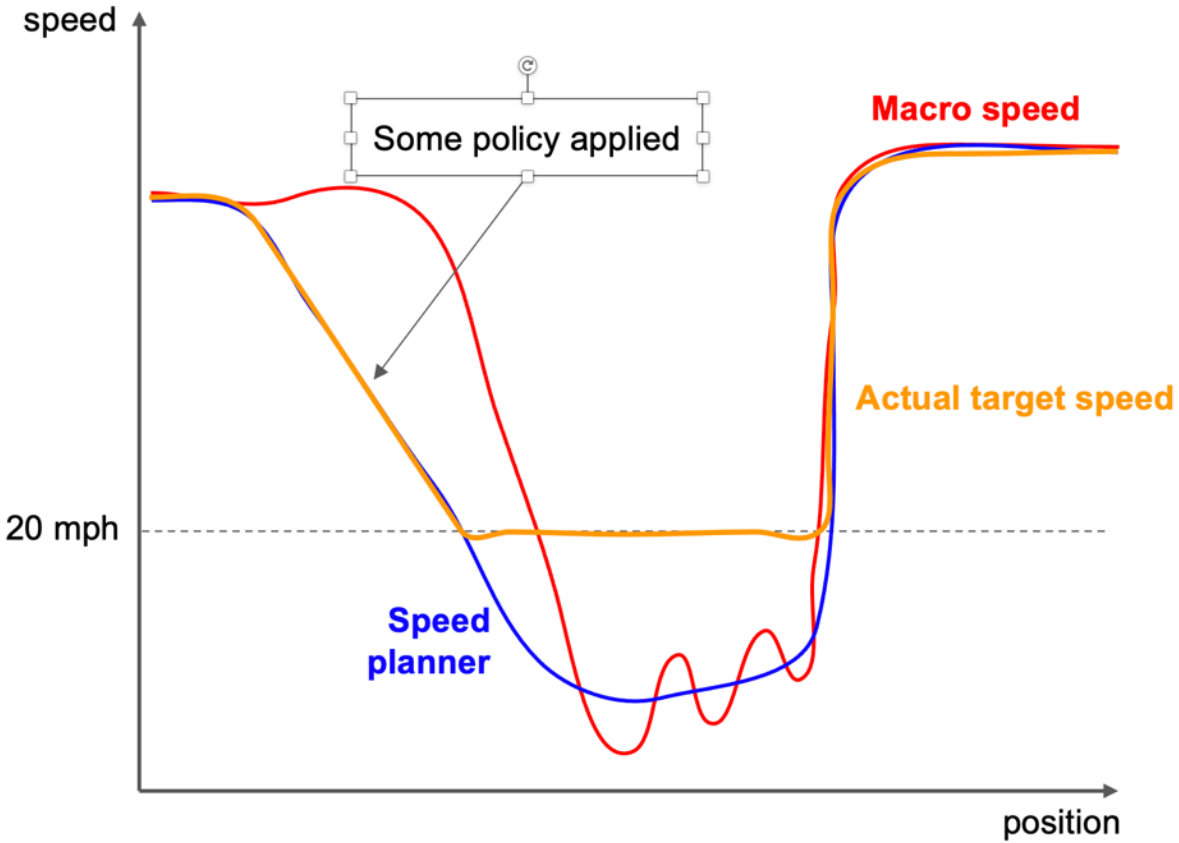
Speed planner



Speed planner



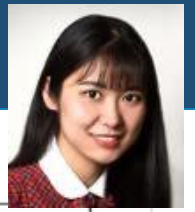
Z. Fu, UC Berkeley



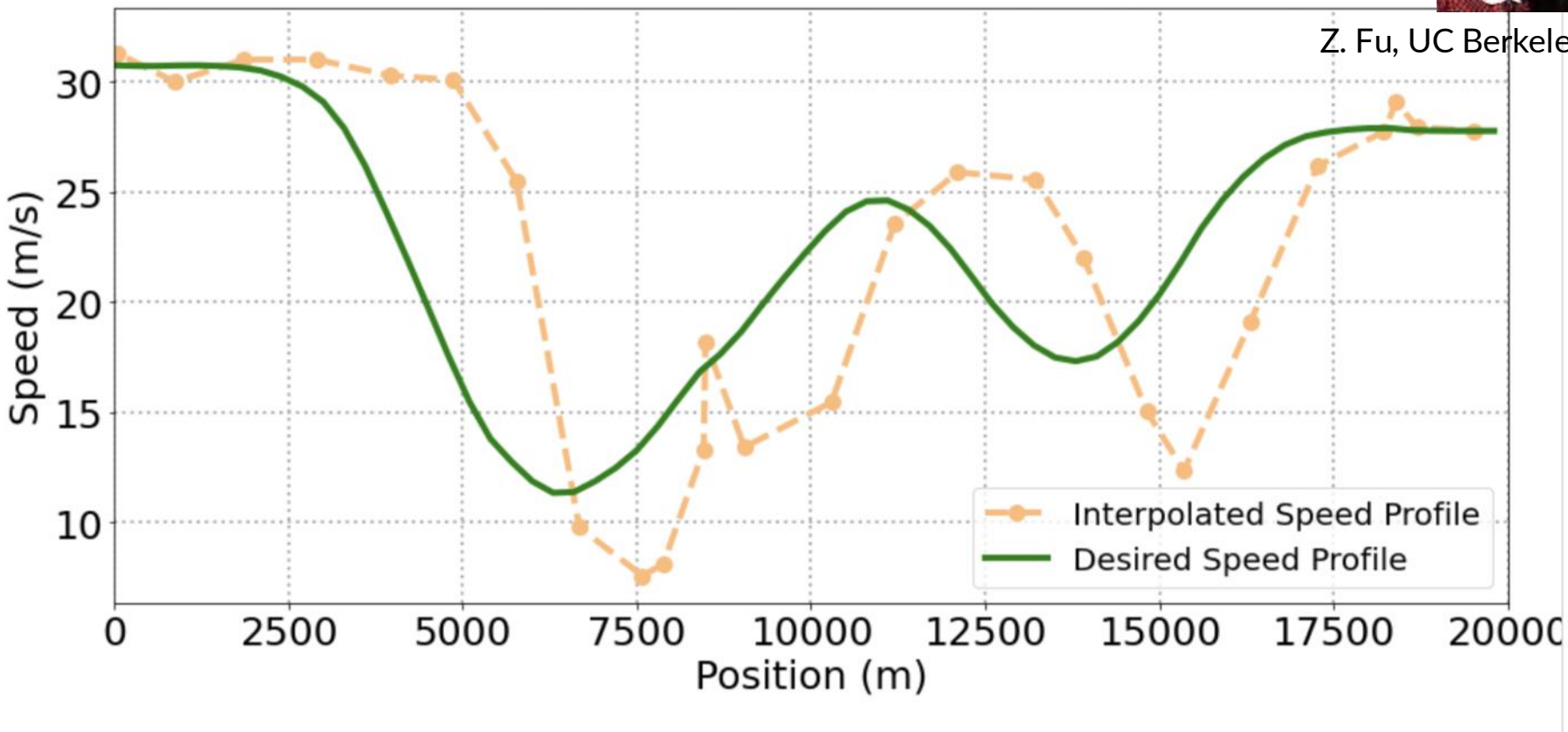
Estimate of the **macro speed** (INRIX)

Generate a **speed planner** speed profile and broadcast to vehicles

Speed planner



Z. Fu, UC Berkeley



Installing/uninstalling 100 cars



Fleet management



I24 MOTION: video-generated trajectories



Prof. D. Work, Vanderbilt

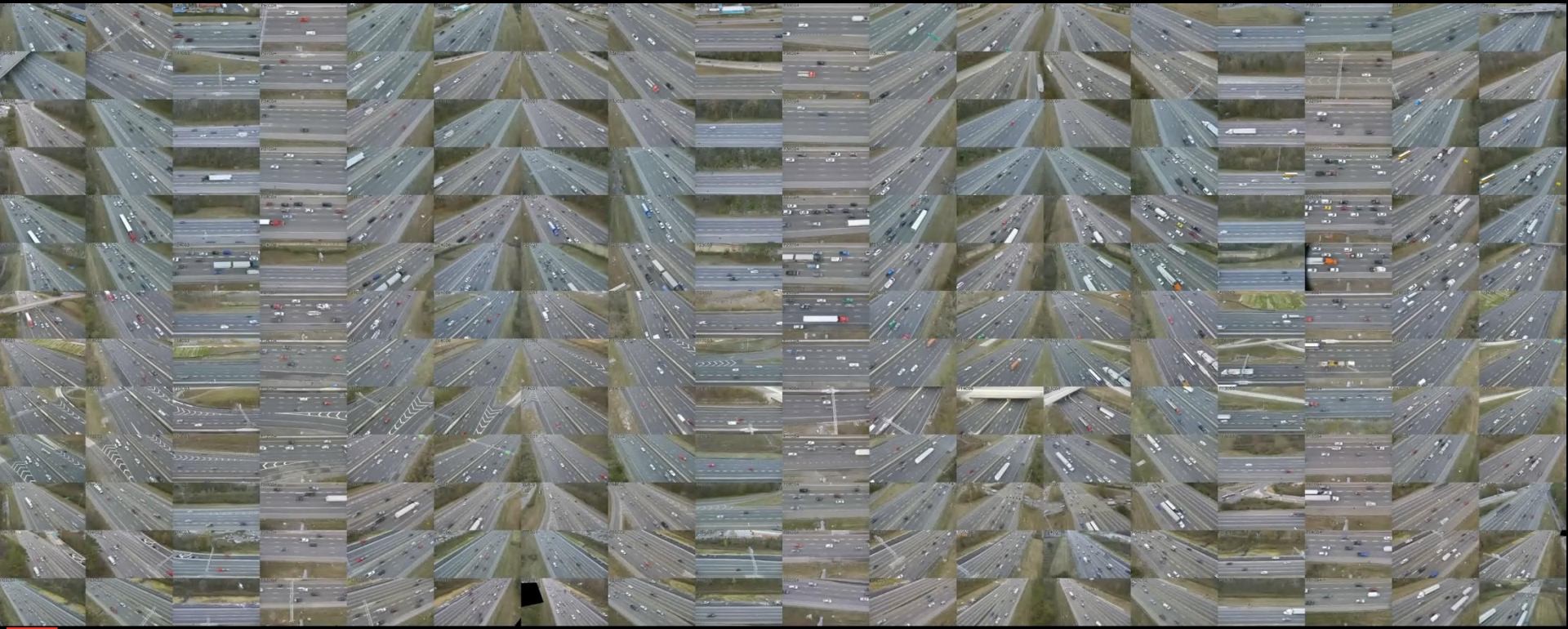


I-24 MOTION

New Instrument for Freeway Traffic Science



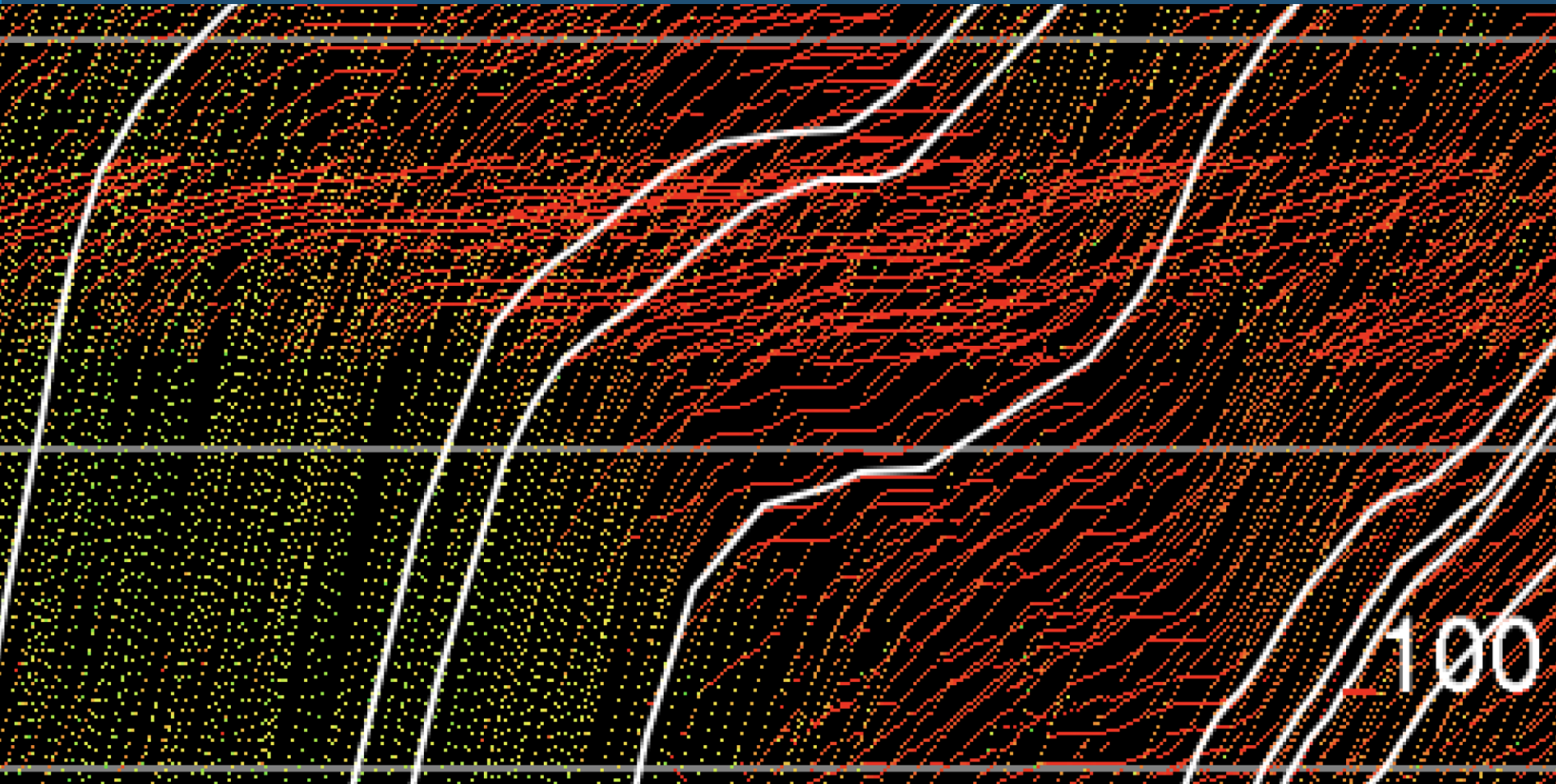
Prof. D. Work, Vanderbilt



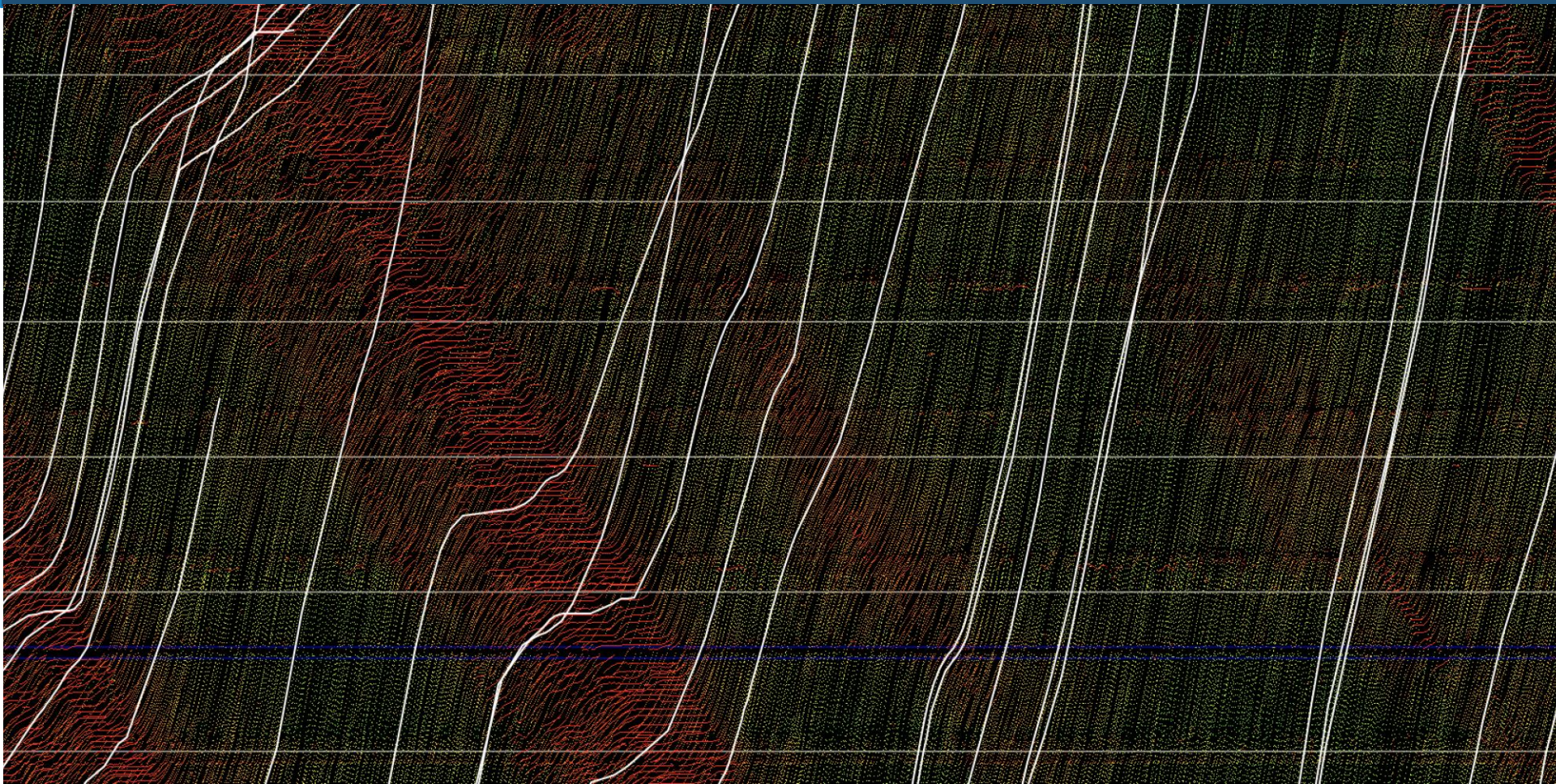
TN

4 miles of cameras observing all vehicles

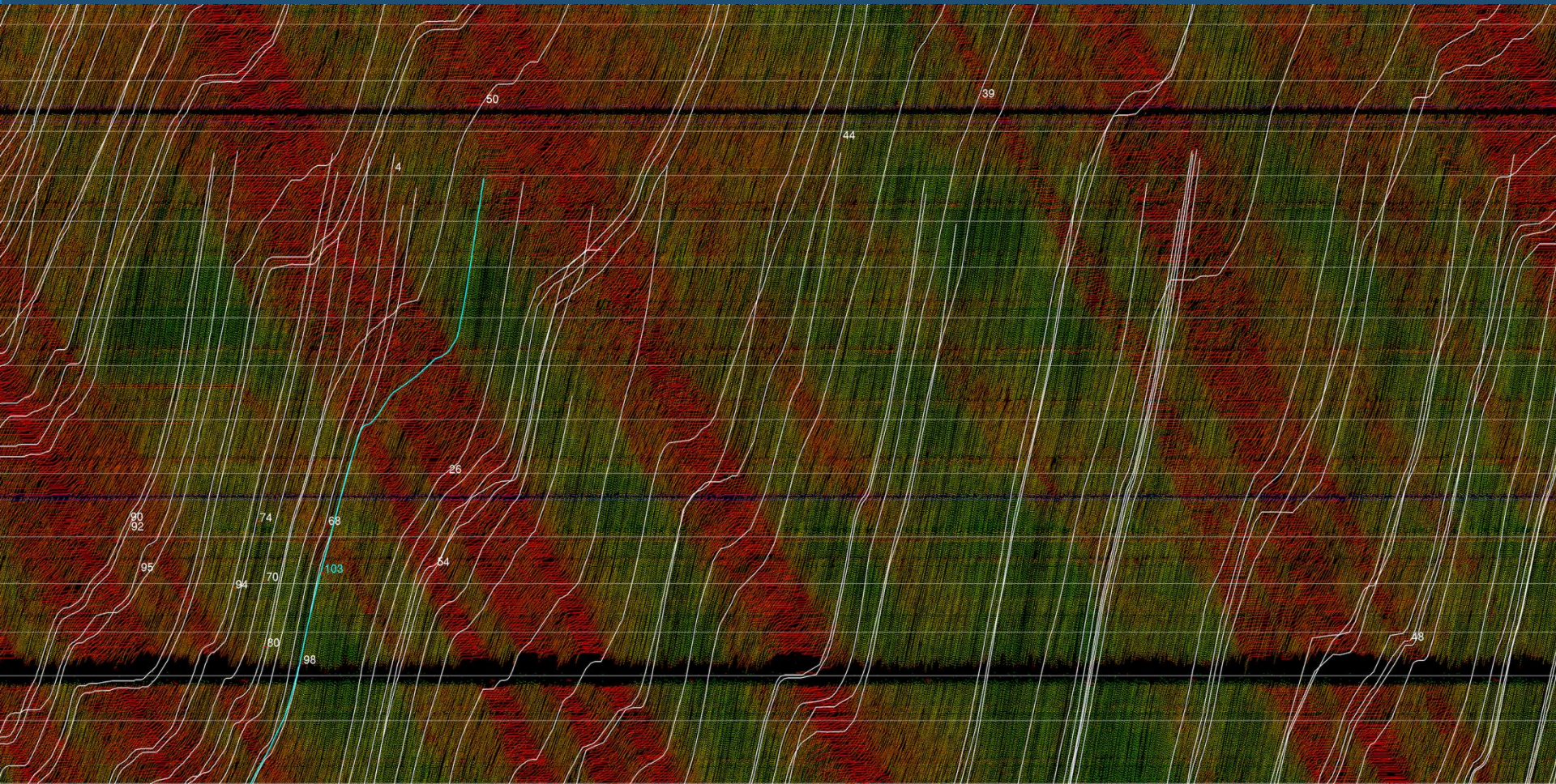
One day of operations in one picture



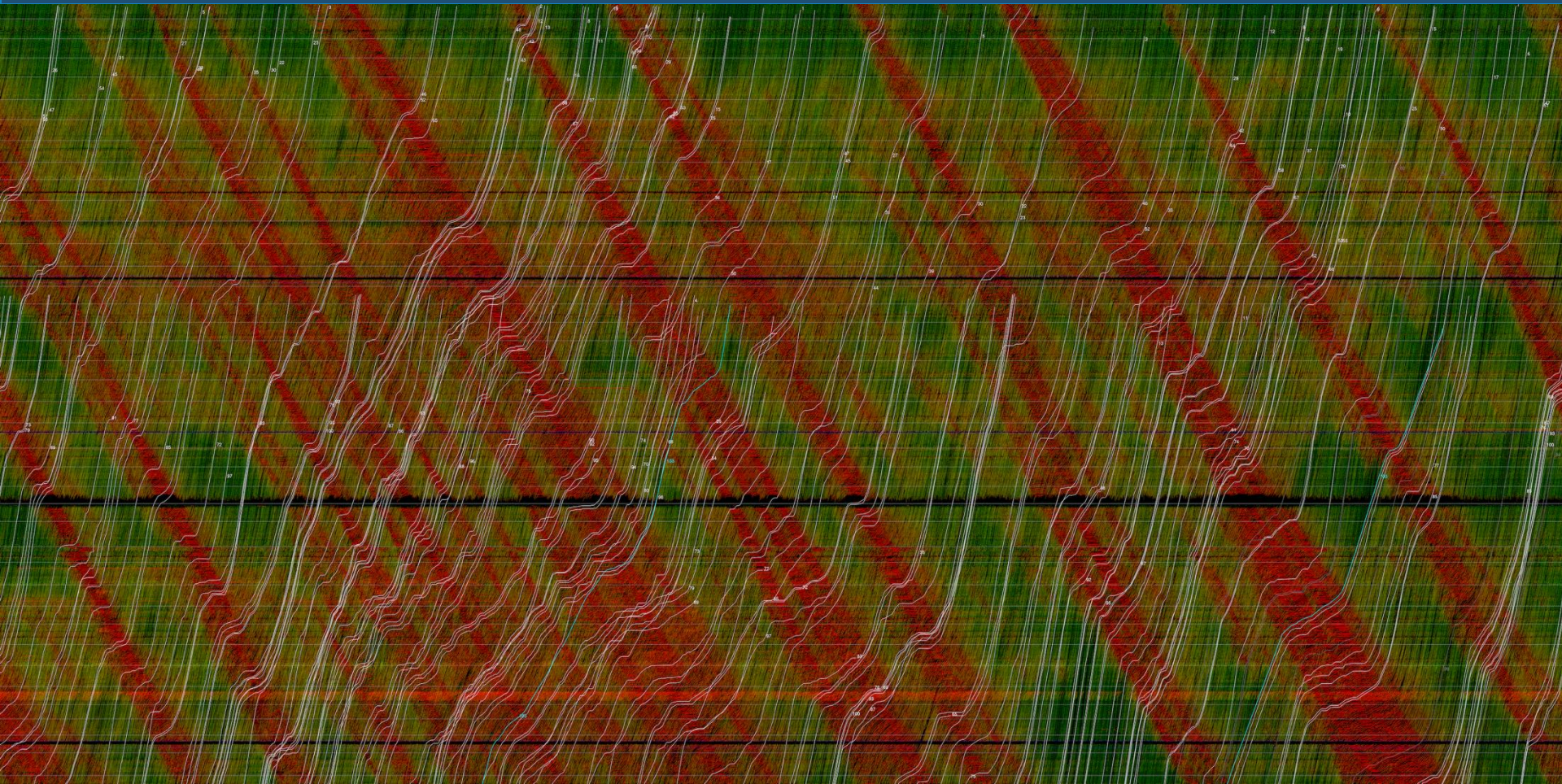
One day of operations in one picture



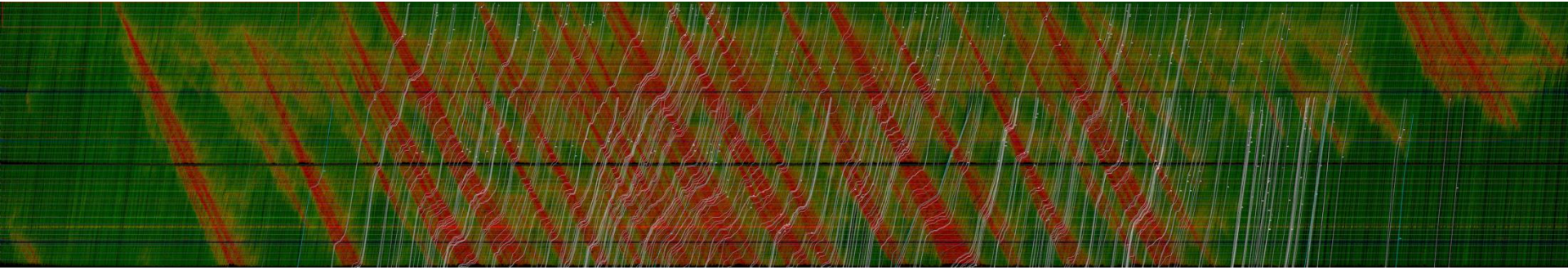
One day of operations in one picture



One day of operations in one picture



One day of operations in one picture



Transportation Science, Aug. 1961… first time space diagram

R-W GROUP DEVELOPS FREEWAY STUDY SYSTEM

NEW STUDY METHODS GATHER DATA FOR TRAFFIC CONTROL

What is a crowded freeway really like? Despite the accumulated opinion of millions of miserable motorists, three R-W engineers are painstakingly putting together dispassionate analyses of freeway jams across the nation.

The three engineers comprise the technical staff of the R-W Surface Traffic Control Section of the Information Systems Department. They are Dr. Daniel Gerlough, Section Head, Dr. Adolf May, and James Bladden.

R-W entered the traffic control field when Gerlough and May joined the company in the summer of 1959, both from distinguished traffic research institutes. Bladden joined the group this year, coming from the city of Baltimore where he participated in the installation of the nation's most advanced street traffic signal system.

Although the group is small it is working in a field where even modest improvements can have immense implications. The relief of freeway congestion means minutes of travel time saved multiplied by thousands of motorists, cupfuls of fuel multiplied by thousands of engines; and decent sun tint in thousands of homes.

But before such a standalone miracle is brought about there remains considerable data gathering to be done to put freeway congestion into the quiet, quantitative terms of the engineer. At present the Section is engaged in three studies of sections of freeways in California, Pennsylvania and Illinois. The California and Pennsylvania studies call for collection and interpretation of data and recommendations based on the findings. The Illinois study, of the Congress St. Expressway in Chicago, goes a step further: Following the study now in progress, a pilot system for the automatic surveillance of traffic density will be installed on a section of the road.



The current objective of the work being carried on at R-W is to find a way to make freeways as efficient as they were meant to be. Anyone who has happened onto a freeway at 5 a.m. on a Wednesday morning in the slack tourist season knows that freeways can indeed be the shortest route between two points. They admirably achieve their purpose of getting more vehicles further in a shorter time, up to a point known at critical density. And public enthusiasm for freeway travel pushes traffic beyond that point at certain times and at certain places on every freeway. The effect is dismal — husbands grumble, wives worry and children whine while the grey grows stiff and cold. It is not until the traffic flow has dropped to a level far below what the freeway was designed to carry that the congestion can begin to ease.

That's the social aspect of the problem and its ramifications lead to public clamor for more freeways, fewer drivers and any number of other economically difficult or socially unacceptable solutions — most of which are solutions to an imperfectly analyzed problem.

It's the analysis of the problem that the R-W engineers are concerned with currently. There are two criteria of freeway effectiveness: "user service" and "volume". The ideal of user service is to make it possible for every driver to pursue his own course along the highway unaffected by other traffic. Most efficient volume is the highest number of vehicle passing a point in a given time. These two objectives of highway engineering tend to get in each other's way. Greatest user service occurs when density (vehicles per mile per lane) is fairly low and when speeds are high and variable. Greatest volume is observed when density is just below the critical value and speeds are modest and relatively uniform.

Both objectives are served, however, by keeping traffic density below the critical point. It has been observed that every freeway studied has a place, or places, where the critical density occurs first. This point, engineers with the opportunity to observe a point, or few points, where the buildup that means trouble occurs. Through further analysis they can locate many of the factors that contribute to the buildup (such as where all the cars are coming from). The final step, of course, would be to link together the observation point with a control system which could do something about the traffic when the buildup approaches the critical level. Then commuters would have a chance to think up new excuses for being late to work and maybe get home on time to get in a few licks with the lawnmower before dinner.

The typical R-W freeway study has four parts: 1. Locate the critical controlling point, or points; 2. Determine the causes for the critical section; 3. Determine means for controlling or correcting the causes; 4. Estimate the benefits.

Data is being collected in three ways: aerial photography, time-lapse ground photography and by investigators deliberately driving the freeways during the rush hours in the unselfish interests of science.

The aerial photos show the traffic density at given points at accurately logged times. On one typical photo mission the p... made 34 passes over a seven-mile section of freeway between three and 6:30 p.m. The photos, each framing about one mile of road, are taken from 6000 feet through a hole in the floor of a Cessna 180. The results have set May and Gerlough dreaming about what they could do with a surplus U2. Besides giving the density of traffic on any section of freeway at about six-minute intervals, the aerial photos sometimes show just how jams start and how the flow of traffic is affected.

The data from the aerial photos is organized in a time-distance-density contour graph (figure 1), which was devised by

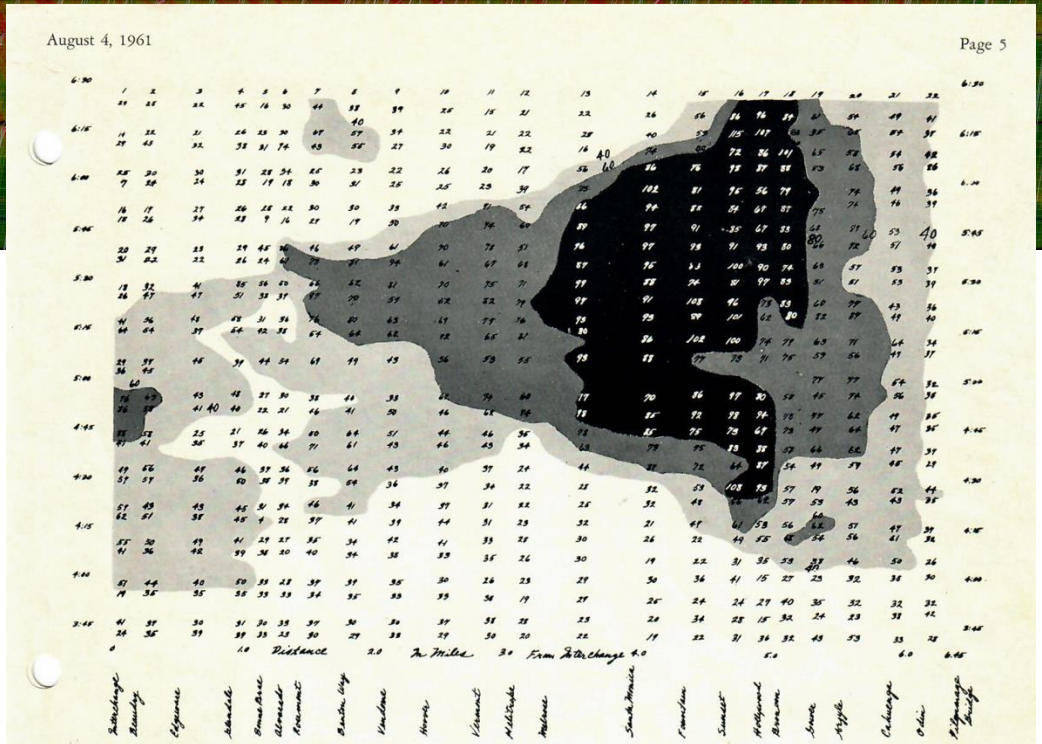
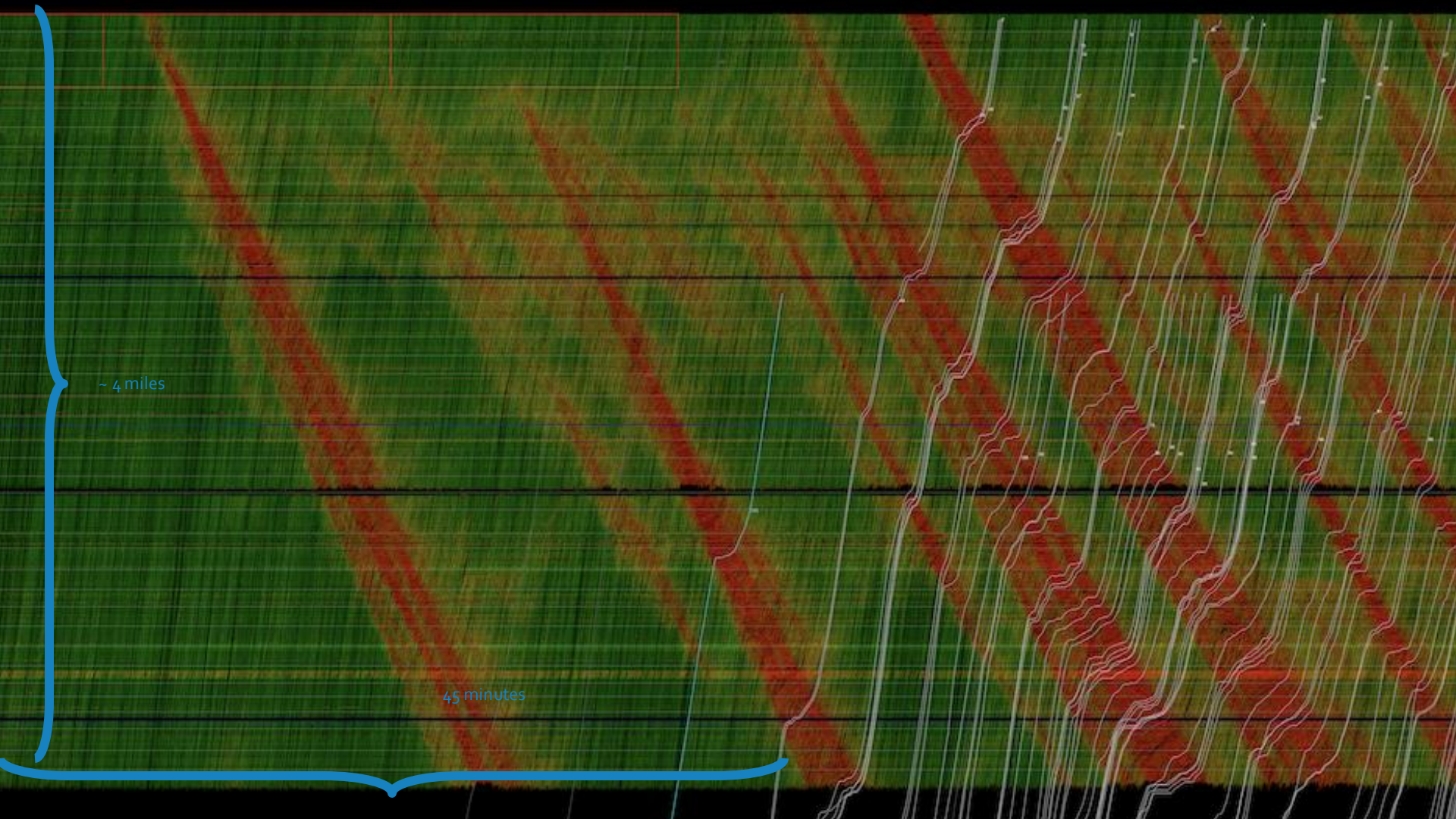
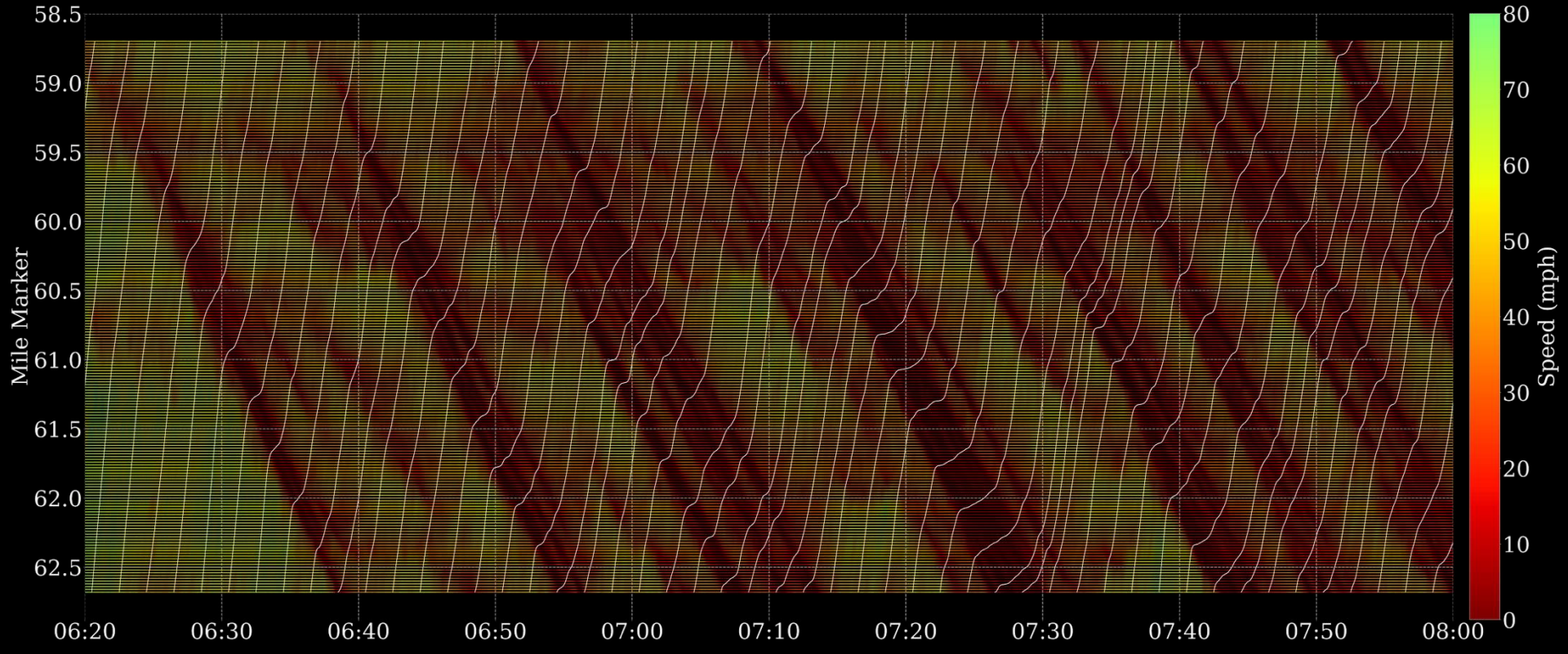


FIGURE ONE—A time-distance-density contour graph of traffic on the Hollywood Freeway at the rush hour is prepared from a simple analysis of aerial photos similar to those below. The chart records the traffic density (figures in vertical columns) at given points on the freeway for each pass of the photo plane. The contours enclose areas of the chart in which the density is higher than a given minimum. The light gray area is moderate density, higher than 40 cars per lane per mile; the darker gray is above the critical point, 60 cars per lane per mile; and the black is a miserable jam. This type chart is a major new tool in analysing freeway traffic conditions developed by R-W's Surface Traffic Control Section.



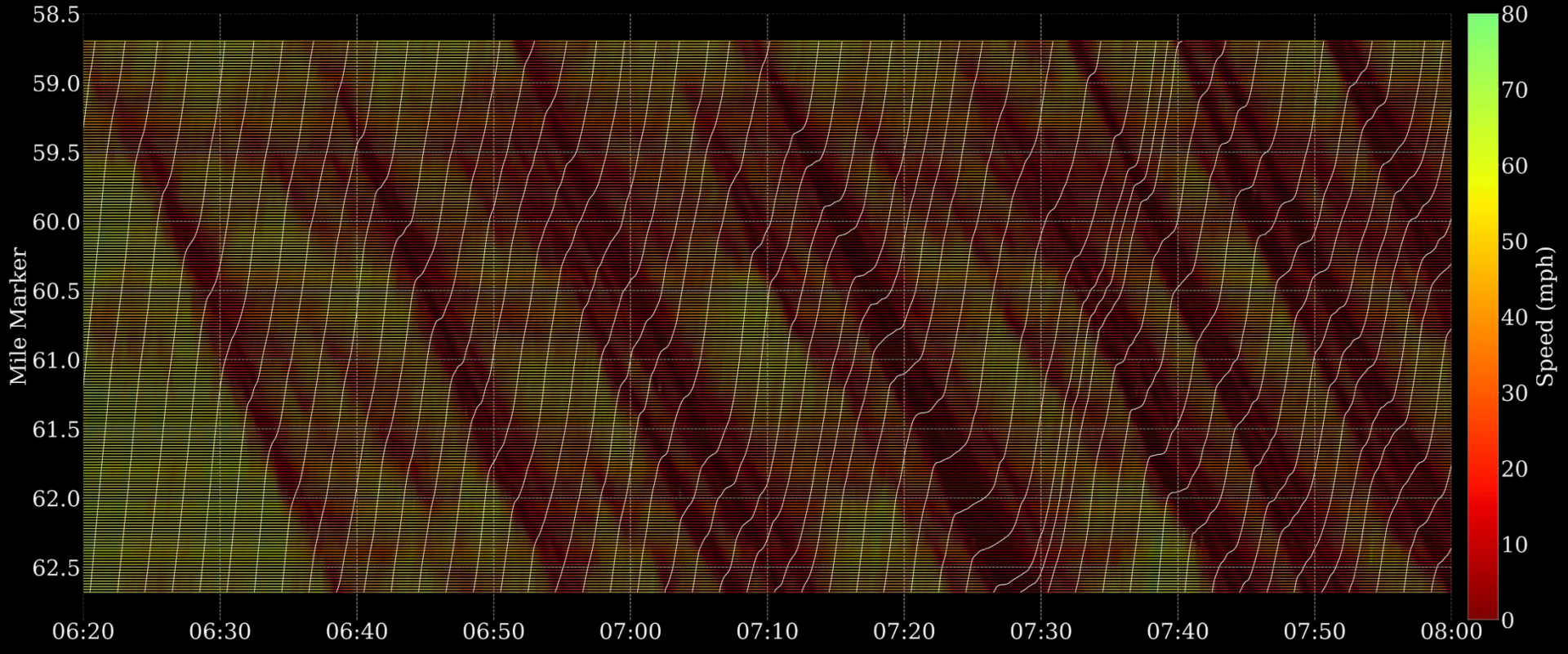

~ 4 miles
45 minutes

Lane 1 (HOV lane)



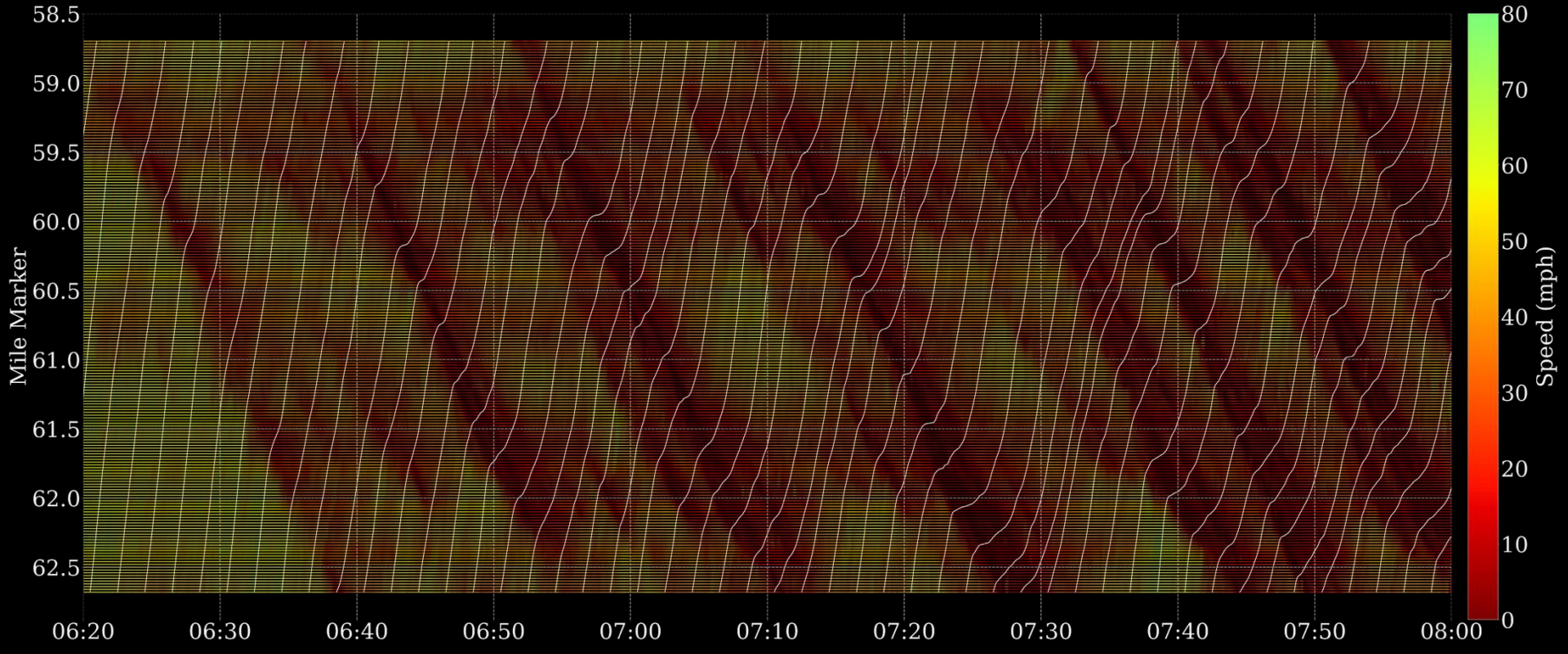
Ji, J., Wang, Y., Gloudemans, D., Zachár, G., Barbour, W., & Work, D. B. (2023). Virtual trajectories for I-24 MOTION: data and tools.

Lane 2



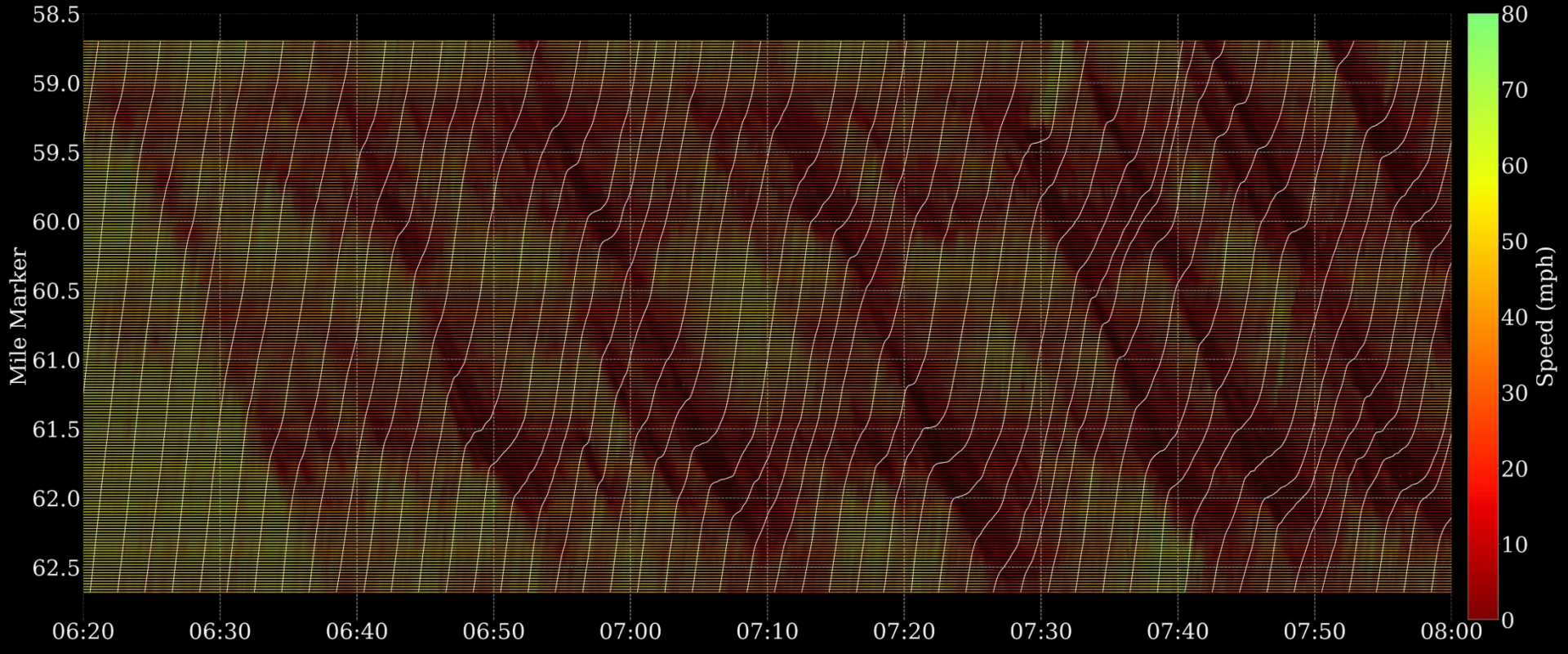
Ji, J., Wang, Y., Gloudemans, D., Zachár, G., Barbour, W., & Work, D. B. (2023). Virtual trajectories for I-24 MOTION: data and tools.

Lane 3



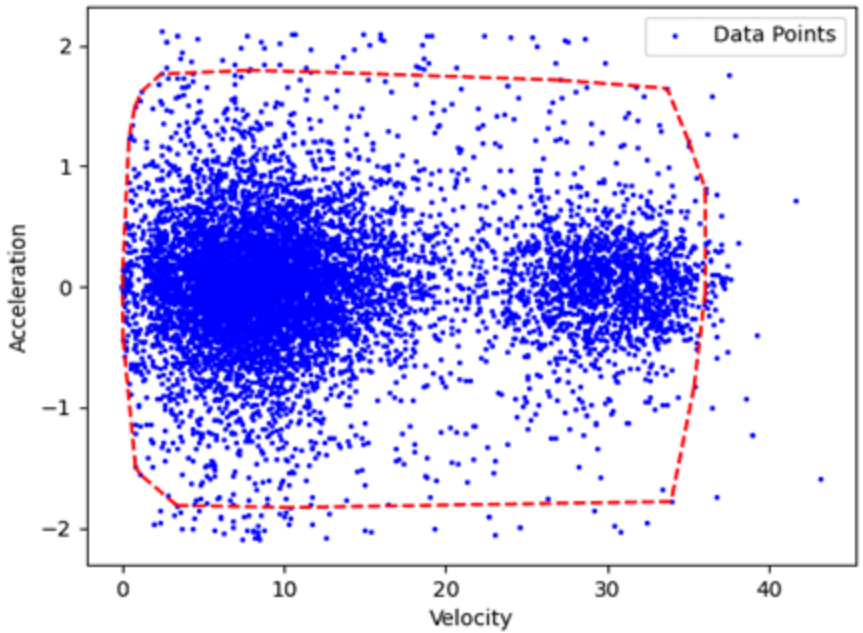
Ji, J., Wang, Y., Gloudemans, D., Zachár, G., Barbour, W., & Work, D. B. (2023). Virtual trajectories for I-24 MOTION: data and tools.

Lane 4



Ji, J., Wang, Y., Gloudemans, D., Zachár, G., Barbour, W., & Work, D. B. (2023). Virtual trajectories for I-24 MOTION: data and tools.

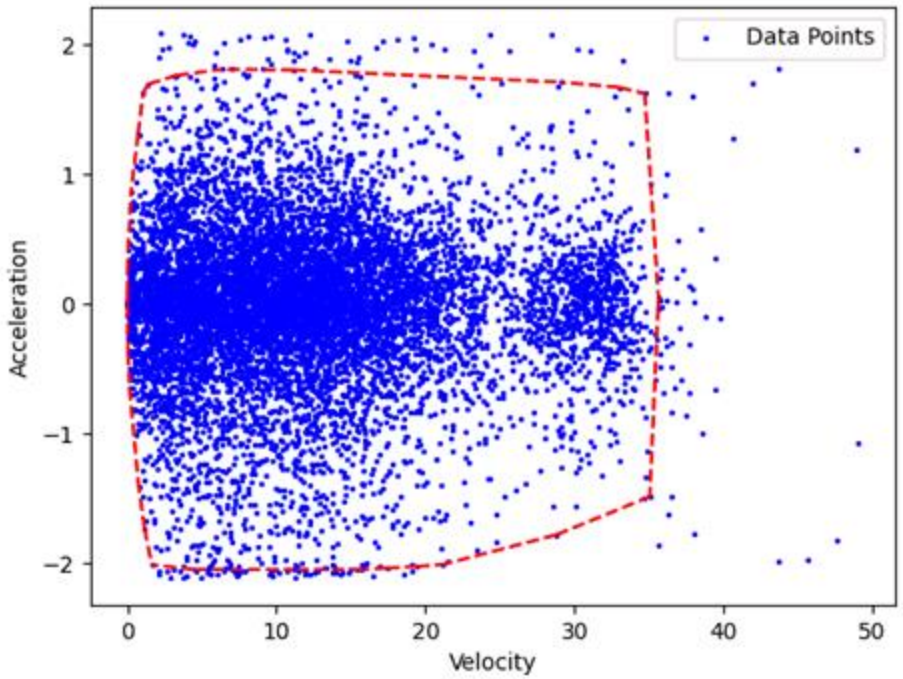
Scatterplot Convex Hull: Tuesday, Nov 15, 7-9 am CST



A scatterplot of velocity-acceleration pairs. We consider data on Tuesday, November 15th, 2022, from 7:00 am to 9:00 am CST, which we sample at a 1:1000 rate.

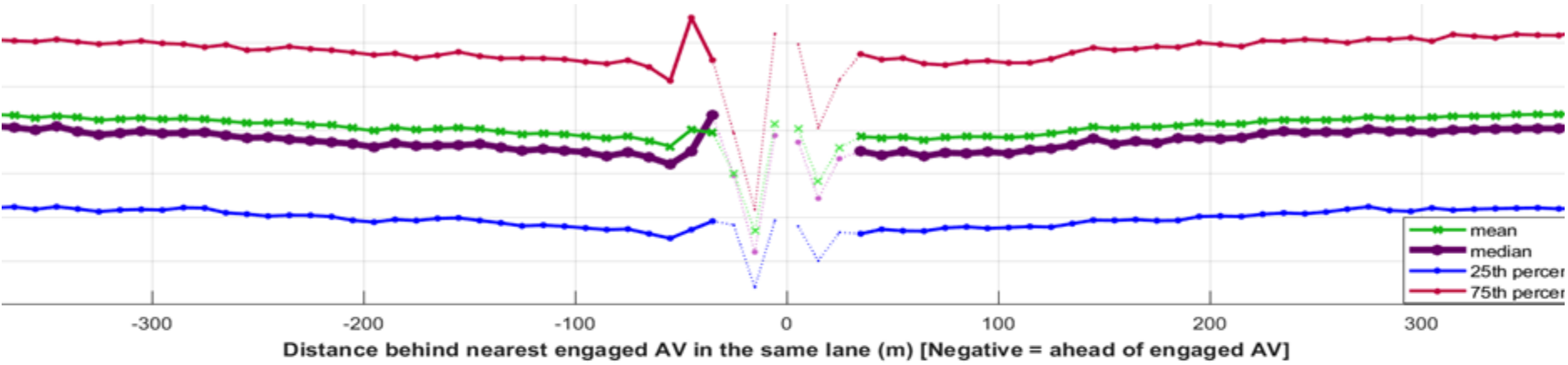
Comparing to Friday, we note that the convex hull bounding 99% of the data points is larger in the absence of AVs, and that the data appear in two distinct clusters, indicating stop-and-go waves

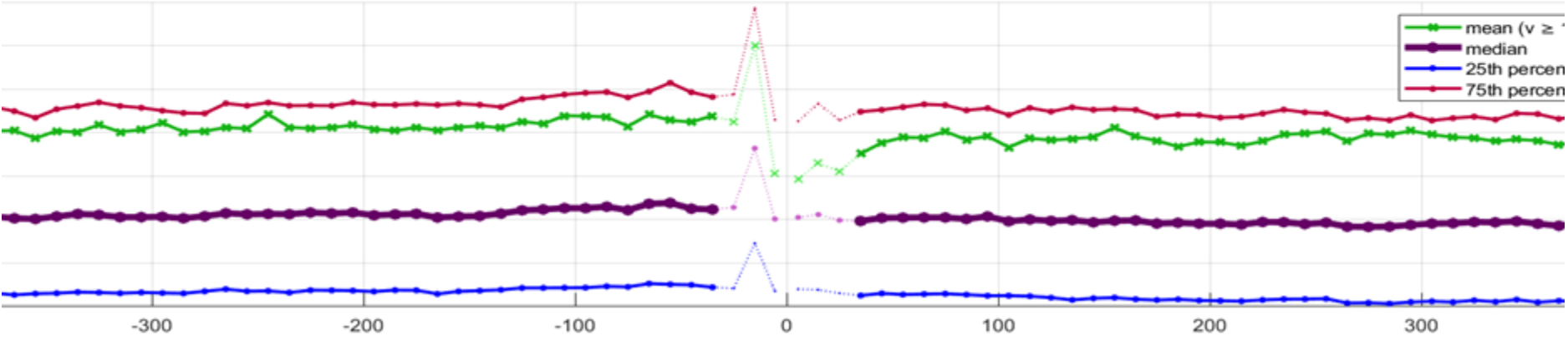
Scatterplot Convex Hull: Friday, Nov 18, 7-9 am CST

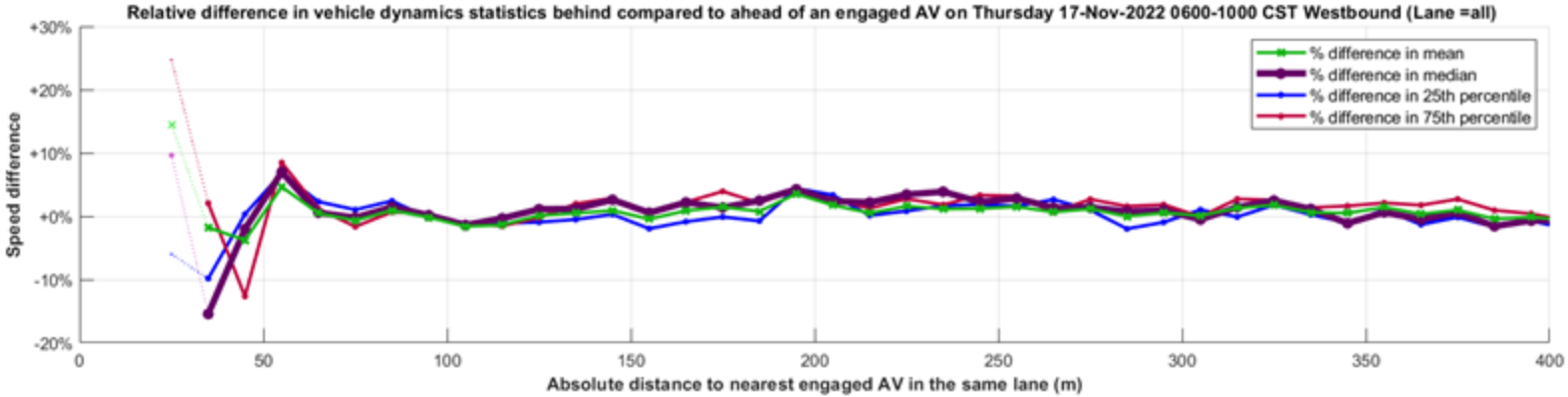


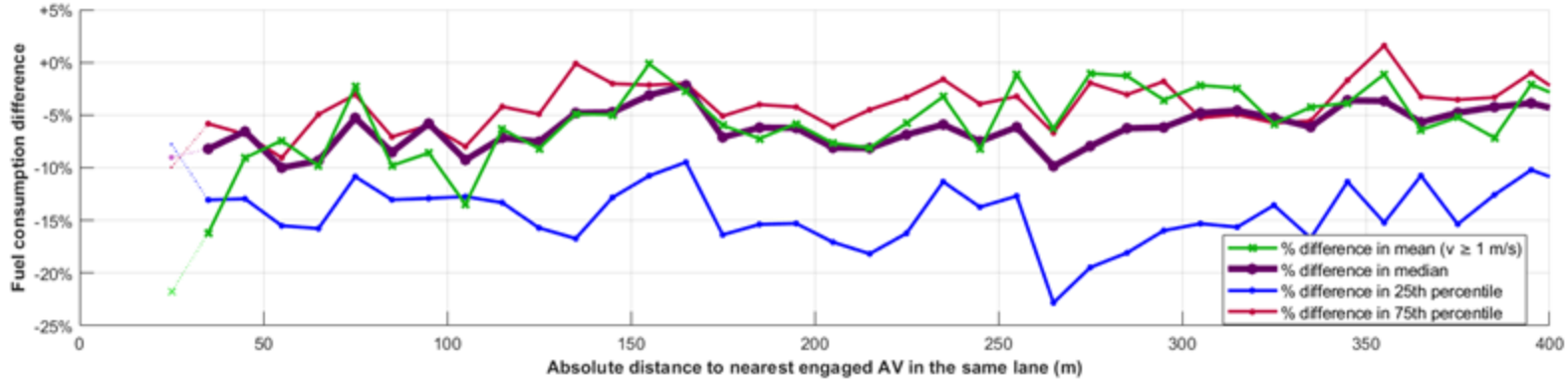
A scatterplot of velocity-acceleration pairs. We consider data on Friday, November 18th, 2022, from 7:00 am to 9:00 am CST, which we sample at a 1:1000 rate.

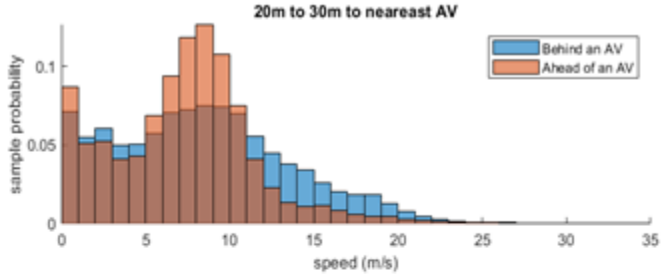
Comparing to Tuesday, we note that the convex hull bounding 99% of the data points is smaller with AVs, and that the data appears to be more spread out, indicating smoothing.

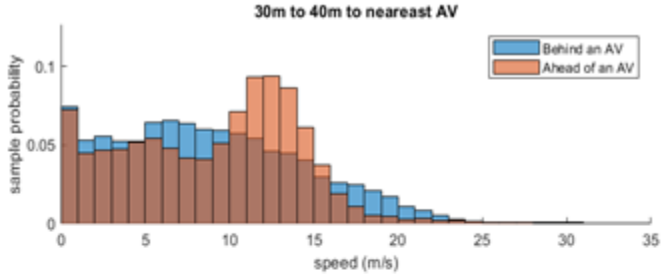

mean
median
25th percen
75th percen
-300
-200
-100
0
100
200
300
Distance behind nearest engaged AV in the same lane (m) [Negative = ahead of engaged AV]

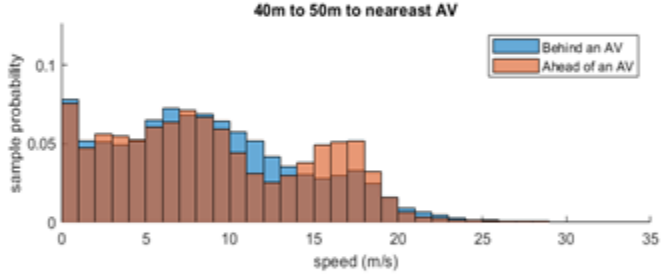

mean (v ≥
median
25th percen
75th percen
-300
-200
-100
0
100
200
300

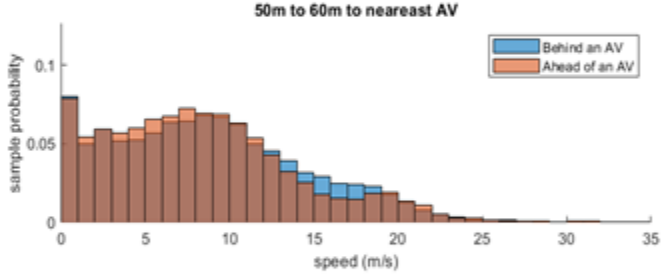

Relative difference in vehicle dynamics statistics behind compared to ahead of an engaged AV on Thursday 17-Nov-2022 0600-1000 CST Westbound (Lane =all)
Speed difference
+30%
+20%
+10%
+0%
-10%
-20%
0
50
100
150
200
250
300
350
400
Absolute distance to nearest engaged AV in the same lane (m)
% difference in mean
% difference in median
% difference in 25th percentile
% difference in 75th percentile

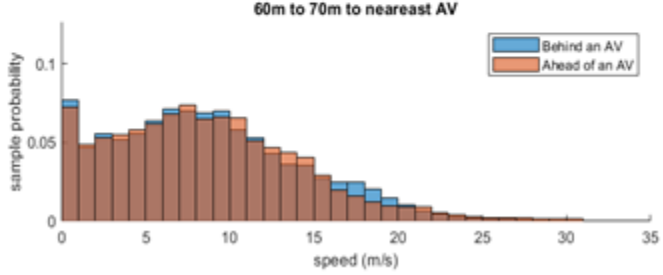

Fuel consumption difference
+5%
+0%
-5%
-10%
-15%
-20%
-25%
0
50
100
150
200
250
300
350
400
Absolute distance to nearest engaged AV in the same lane (m)
% difference in mean (v ≥ 1 m/s)
% difference in median
% difference in 25th percentile
% difference in 75th percentile

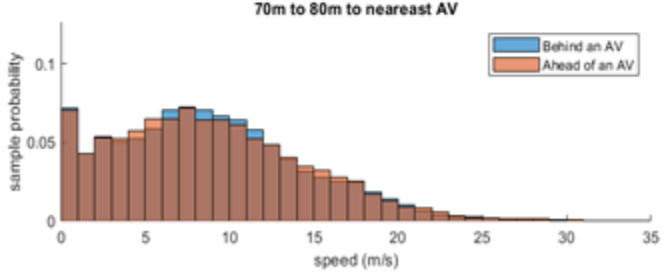

20m to 30m to neareast AV
Behind an AV
Ahead of an AV
sample probability
speed (m/s)


30m to 40m to neareast AV
Behind an AV
Ahead of an AV
sample probability
speed (m/s)

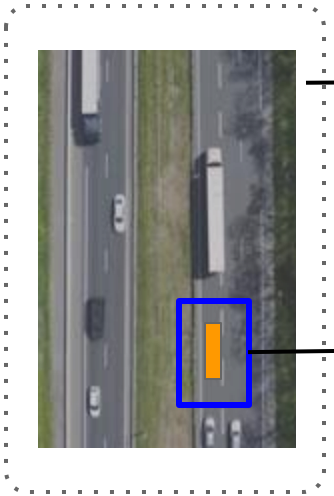

40m to 50m to neareast AV
Behind an AV
Ahead of an AV
sample probability
speed (m/s)


50m to 60m to neareast AV
Behind an AV
Ahead of an AV
sample probability
speed (m/s)


60m to 70m to neareast AV
Behind an AV
Ahead of an AV
sample probability
speed (m/s)


70m to 80m to neareast AV
Behind an AV
Ahead of an AV
sample probability
speed (m/s)

Verbalizing the action of the controller



Extract scene, describe environment textually.

Environment Context

Text description of agent (AV) state

Agent State

Description of policy, reward model, and action taken.

$\pi_\theta(\cdot)$

Contextual data is concatenation of environment context, AV state, and policy description.

Task Instructions

Contextual Data

Few Shot Examples

K few-shot examples inserted into prompt

Few Shot Database

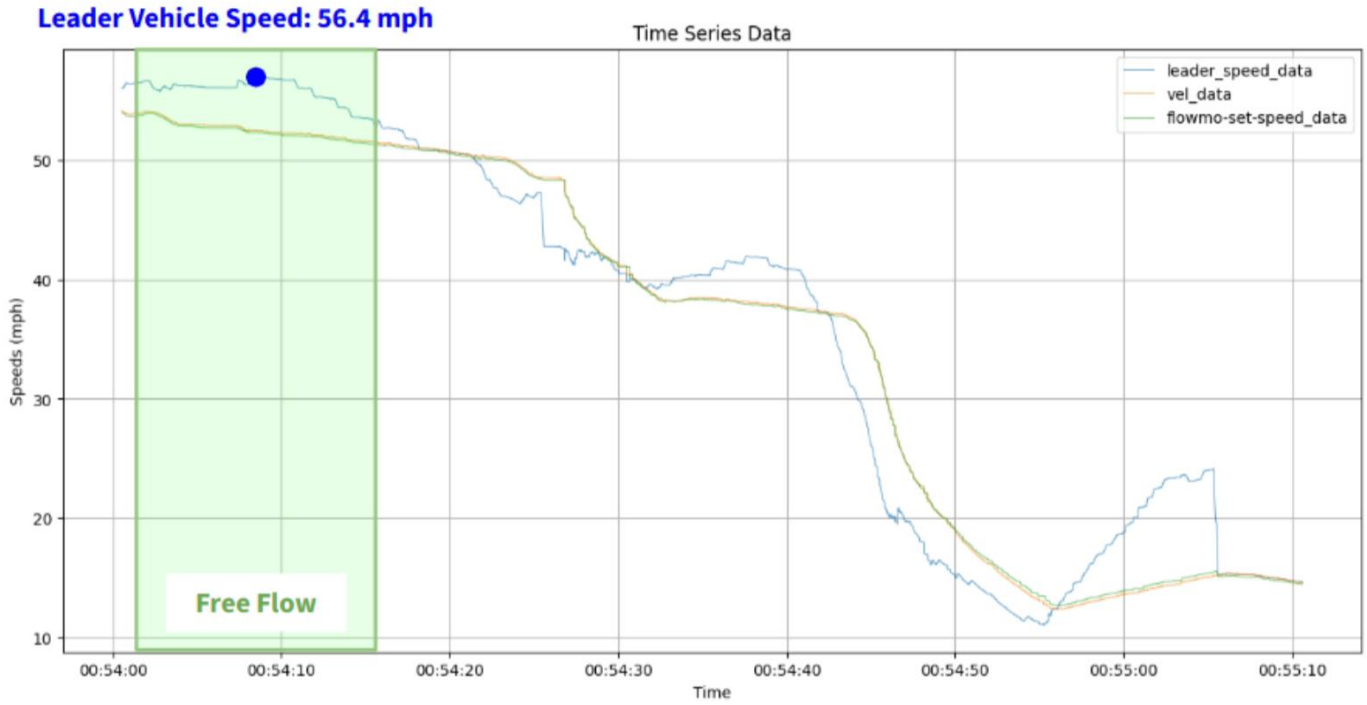
User Query: "Why are you slowing down?"

Prompt sent to LLM

Out-of-box LLM

Response: ⚠ The vehicle in front of us is slowing down, likely due to congestion. I am decelerating at 2 m/s^2 preemptively. This will help smooth the deceleration of vehicles behind me, preventing sudden stop-and-go behavior.

Verbalizing the action of the controller



MVT



Questions

