# STAResNet

## A Network in Spacetime Algebra to solve Maxwell's PDEs

Alberto Pepe[1], Sven Buchholz[2], Joan Lasenby[1]

[1]University of Cambridge, Signal Processing & Communications Lab, Engineering Department, Cambridge, UK

[2]Technische Hochschule Brandenburg, Fachbereich Informatik und Medien, Brandenburg, Germany

**Applied Geometric Algebras in Computer Science and Engineering (AGACSE)**
Amsterdam, 27-29 August 2024

# Outline

# Table of Contents

# Motivation

**Geometric Algebra (GA) Networks** have been gaining significant momentum in the past two years[1].

[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

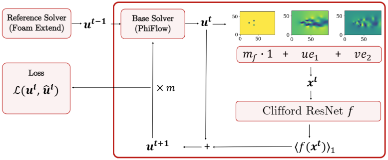
# Motivation

**Geometric Algebra (GA) Networks** have been gaining significant momentum in the past two years[1]. We have employed them to...

[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

# Motivation

**Geometric Algebra (GA) Networks** have been gaining significant momentum in the past two years[1]. We have employed them to...

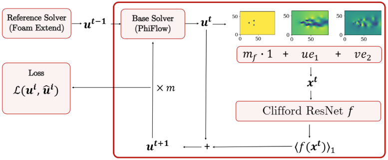



(a) ...solve PDEs

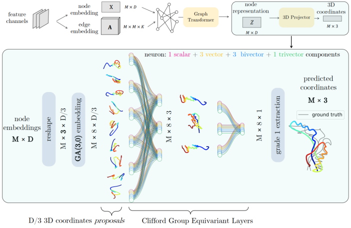
[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

# Motivation

**Geometric Algebra (GA) Networks** have been gaining significant momentum in the past two years[1]. We have employed them to...




(a) ...solve PDEs

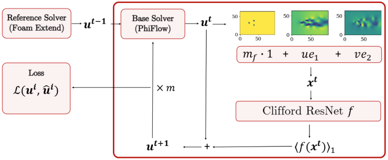



(b) ...predict protein structures

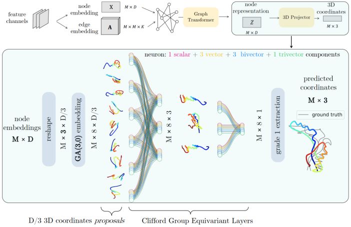
[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

# Motivation

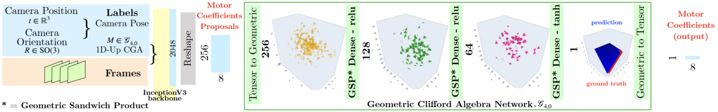
**Geometric Algebra (GA) Networks** have been gaining significant momentum in the past two years[1]. We have employed them to...




(a) ...solve PDEs




(b) ...predict protein structures




(c) ...estimate camera poses

[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

# Motivation

A key step in GA Networks is the **embedding** in a given algebra

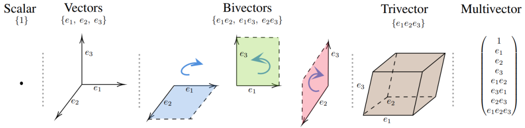



Figure 2: Elements in G(3,0,0)[1]

Examples:

[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

# Motivation

A key step in GA Networks is the **embedding** in a given algebra

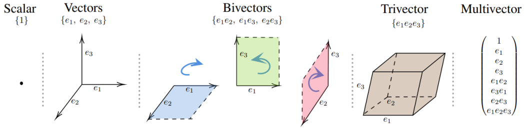



Figure 2: Elements in G(3,0,0)[1]

Examples:

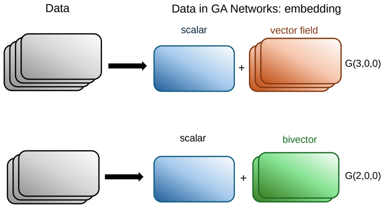



Figure 3: Two examples of GA embedding

[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

# Goal

**Question**

How does the choice of the algebra impact learning in Geometric Algebra Networks?

**We address it by studying Maxwell's equations and solving them in Spacetime Algebra (STA).**

# Goal

**Question**

How does the choice of the algebra impact learning in Geometric Algebra Networks?

**We address it by studying Maxwell's equations and solving them in Spacetime Algebra (STA).**

Contributions:

# Goal

**Question**

How does the choice of the algebra impact learning in Geometric Algebra Networks?

**We address it by studying Maxwell's equations and solving them in Spacetime Algebra (STA).**

Contributions:

- Showed that STA simplifies the solution Maxwell's PDEs *also* if learnt

# Goal

**Question**

How does the choice of the algebra impact learning in Geometric Algebra Networks?

**We address it by studying Maxwell's equations and solving them in Spacetime Algebra (STA).**

Contributions:

- Showed that STA simplifies the solution Maxwell's PDEs *also* if learnt
- Up to 2.6× lower MSE with 6× fewer trainable parameters compared to Clifford ResNet

# Goal

**Question**

How does the choice of the algebra impact learning in Geometric Algebra Networks?

**We address it by studying Maxwell's equations and solving them in Spacetime Algebra (STA).**

Contributions:

- Showed that STA simplifies the solution Maxwell's PDEs *also* if learnt
- Up to 2.6× lower MSE with 6× fewer trainable parameters compared to Clifford ResNet
- First implementation of a network entirely in STA

# Table of Contents

# Problem Definition: Describing Maxwell's PDEs

## Maxwell's Equations in Differential Form

$$\nabla \cdot \mathbf{E} = \rho \qquad \nabla \cdot \mathbf{B} = 0 \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J}$$

# Problem Definition: Describing Maxwell's PDEs

## Maxwell's Equations in Differential Form

$$\nabla \cdot \mathbf{E} = \rho \qquad \nabla \cdot \mathbf{B} = 0 \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J}$$

## Maxwell's Equations in GA - $G_{3,0,0}$

$$\left(\frac{\partial}{\partial t} + i\nabla\right) F = \mathbf{J} - i\rho, \text{ with}$$
$$F = \mathbf{E} + i\mathbf{B} = E_1 e_1 + E_2 e_2 + E_3 e_3 + B_1 e_{23} + B_2 e_{13} + B_3 e_{12}$$

# Problem Definition: Describing Maxwell's PDEs

## Maxwell's Equations in Differential Form

$$\nabla \cdot \mathbf{E} = \rho \qquad \nabla \cdot \mathbf{B} = 0 \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J}$$

## Maxwell's Equations in GA - $G_{3,0,0}$

$$\left(\frac{\partial}{\partial t} + i\nabla\right) F = \mathbf{J} - i\rho, \text{ with}$$
$$F = \mathbf{E} + i\mathbf{B} = E_1 e_1 + E_2 e_2 + E_3 e_3 + B_1 e_{23} + B_2 e_{13} + B_3 e_{12}$$

## Maxwell's Equations in STA - $G_{1,3,0}$

$$\nabla \mathbf{F} = J, \text{ with}$$
$$\mathbf{F} = \mathbf{E} + I\mathbf{B} = E_1 \gamma_{10} + E_2 \gamma_{20} + E_3 \gamma_{30} + B_1 \gamma_{13} + B_2 \gamma_{13} + B_3 \gamma_{12},$$
$$J = (\rho - \mathbf{J})\gamma_0$$
$$\nabla = \gamma^i \frac{\partial}{\partial x_i}$$

# Problem Definition: Describing Maxwell's PDEs

## Maxwell's Equations in STA

$$\nabla \mathbf{F} = J$$

# Problem Definition: Describing Maxwell's PDEs

## Maxwell's Equations in STA

$$\nabla \mathbf{F} = J$$

- Single, covariant Maxwell's ~~equations~~ equation

# Problem Definition: Describing Maxwell's PDEs

## Maxwell's Equations in STA

$$\nabla \mathbf{F} = J$$

- Single, covariant Maxwell's ~~equations~~ equation
- From Prof. Lasenby's presentation, *A new language for Physics* at GAME 2020:

# Problem Definition: Describing Maxwell's PDEs

**Maxwell's Equations in STA**

$$\nabla \mathbf{F} = J$$

- Single, covariant Maxwell's ~~equations~~ equation
- From Prof. Lasenby's presentation, *A new language for Physics* at GAME 2020:

  *" The advantage here is not merely notational [...] the geometric product with the vector derivative is invertible [...] where the separate divergence and curl operators are not"*

# Problem Definition: Describing Maxwell's PDEs

## Maxwell's Equations in STA

$$\nabla \mathbf{F} = J$$

- Single, covariant Maxwell's ~~equations~~ equation
- From Prof. Lasenby's presentation, *A new language for Physics* at GAME 2020:

  *" The advantage here is not merely notational [...] the geometric product with the vector derivative is invertible [...] where the separate divergence and curl operators are not"*
  *"This led to the development of a new method for calculating EM response of conductors to incoming plane waves [...] it was possible to change the illumination in real time"*

# Problem Definition: Architecture

**GA - $G_{3,0,0}$**

$$\left(\frac{\partial}{\partial t} + i\nabla\right) F = \mathbf{J} - i\rho$$

**STA - $G_{1,3,0}$**

$$\nabla \mathbf{F} = J$$

[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

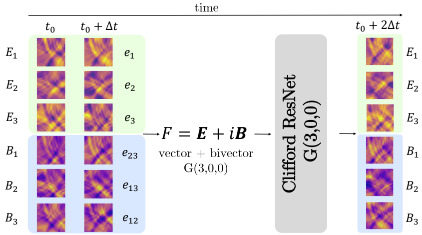
# Problem Definition: Architecture

## GA - $G_{3,0,0}$

$$\left(\frac{\partial}{\partial t} + i\nabla\right) F = \mathbf{J} - i\rho$$

## STA - $G_{1,3,0}$

$$\nabla \mathbf{F} = J$$




(a) Clifford ResNet (GA) [1]

[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

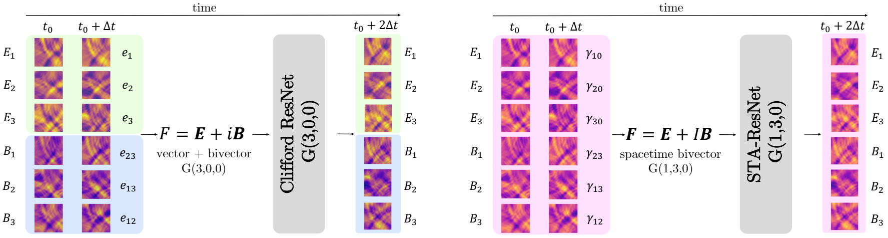
# Problem Definition: Architecture

**GA - $G_{3,0,0}$**

$$\left(\frac{\partial}{\partial t} + i\nabla\right) F = \mathbf{J} - i\rho$$

**STA - $G_{1,3,0}$**

$$\nabla \mathbf{F} = J$$




(a) Clifford ResNet (GA) [1]  (b) STAResNet (STA) [ours]

Figure 4: The two approaches to the solution of Maxwell's PDEs

[1] Clifford Neural Layers for PDE Modeling, Brandstetter et al., ICLR 2023

# Table of Contents

# Experiments

To validate the claim that STA is a better suited space to work in when dealing with Maxwell's PDEs...

# Experiments

To validate the claim that STA is a better suited space to work in when dealing with Maxwell's PDEs...

We look at:

- 2D EM fields
- 3D EM fields

Generated through a FDTD solver.

# Experiments

To validate the claim that STA is a better suited space to work in when dealing with Maxwell's PDEs...

We look at:

- 2D EM fields
- 3D EM fields

Generated through a FDTD solver. We study the impact of:

- Varying sampling period of the dataset $\Delta t$

# Experiments

To validate the claim that STA is a better suited space to work in when dealing with Maxwell's PDEs...

We look at:

- 2D EM fields
- 3D EM fields

Generated through a FDTD solver. We study the impact of:

- Varying sampling period of the dataset $\Delta t$
- Presence of obstables in the domain

# Experiments

To validate the claim that STA is a better suited space to work in when dealing with Maxwell's PDEs...

We look at:

- 2D EM fields
- 3D EM fields

Generated through a FDTD solver. We study the impact of:

- Varying sampling period of the dataset $\Delta t$
- Presence of obstables in the domain
- Number of trainable parameters

# Experiments

To validate the claim that STA is a better suited space to work in when dealing with Maxwell's PDEs...

We look at:

- 2D EM fields
- 3D EM fields

Generated through a FDTD solver. We study the impact of:

- Varying sampling period of the dataset $\Delta t$
- Presence of obstables in the domain
- Number of trainable parameters
- Rollout error

# Metrics

Three metrics to assess the quality of the PDEs solution

# Metrics

Three metrics to assess the quality of the PDEs solution

- Mean squared error

$$\mathcal{L} = \frac{1}{LMN} \sum_{j+x,y,z} \sum_{l=0}^{L} \sum_{m=0}^{M} \sum_{n=0}^{N} (E_{jlmn,i+2\Delta t} - \hat{E}_{jlmn,i+2\Delta t})^2 + (B_{jlmn,i+2\Delta t} - \hat{B}_{jlmn,i+2\Delta t})^2$$

# Metrics

Three metrics to assess the quality of the PDEs solution

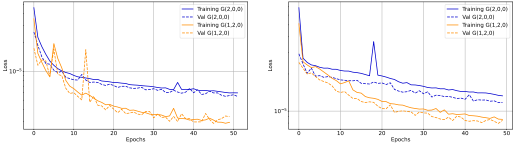
- Mean squared error

$$\mathcal{L} = \frac{1}{LMN} \sum_{j+x,y,z} \sum_{l=0}^{L} \sum_{m=0}^{M} \sum_{n=0}^{N} (E_{jlmn,i+2\Delta t} - \hat{E}_{jlmn,i+2\Delta t})^2 + (B_{jlmn,i+2\Delta t} - \hat{B}_{jlmn,i+2\Delta t})^2$$

- Structural Similarity Index Measure

$$\text{SSIM}(\mathbf{F}^2, \hat{\mathbf{F}^2}) = \frac{(2\mu_{\mathbf{F}^2}\mu_{\hat{\mathbf{F}^2}} + C_1)(2\sigma_{\mathbf{F}^2\hat{\mathbf{F}^2}} + C_2)}{(\mu_{\mathbf{F}^2}^2 + \mu_{\hat{\mathbf{F}^2}}^2 + C_1)(\sigma_{\mathbf{F}^2}^2 + \sigma_{\hat{\mathbf{F}^2}}^2 + C_2)}$$

# Experiments: 2D fields, varying $\Delta t$

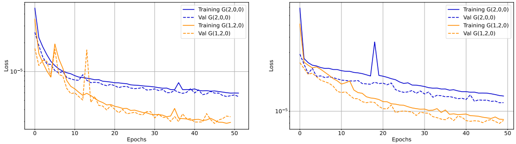



(a) $\Delta t = 25$s

(b) $\Delta t = 50$s

Figure 5: Training and validation losses versus number of epochs for 2D Maxwell's PDEs.

# Experiments: 2D fields, varying $\Delta t$




(a) $\Delta t = 25$s (b) $\Delta t = 50$s

Figure 5: Training and validation losses versus number of epochs for 2D Maxwell's PDEs.

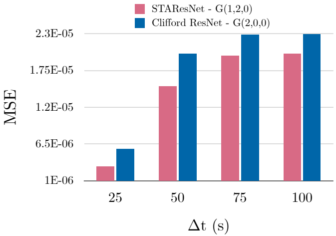



Figure 6: MSE between estimated and GT EM fields vs $\Delta t$.

# Experiments: 2D fields, varying $\Delta t$

$$\mathbf{F}^2 = (\mathbf{E} + I\mathbf{B})^2 = (E_1\gamma_{10} + E_2\gamma_{20} + E_3\gamma_{30} + B_1\gamma_{13} + B_2\gamma_{13} + B_3\gamma_{12})^2$$

# Experiments: 2D fields, varying $\Delta t$

$$\mathbf{F}^2 = (\mathbf{E} + I\mathbf{B})^2 = (E_1\gamma_{10} + E_2\gamma_{20} + E_3\gamma_{30} + B_1\gamma_{13} + B_2\gamma_{13} + B_3\gamma_{12})^2$$

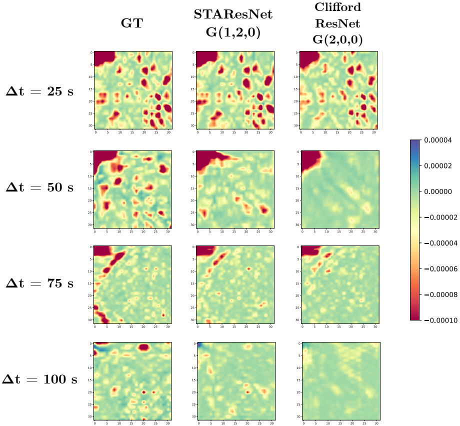



Figure 7: Squared magnitude of the Faraday bivector $\mathbf{F}^2$ for varying $\Delta t$.

# Experiments: impact of obstacles

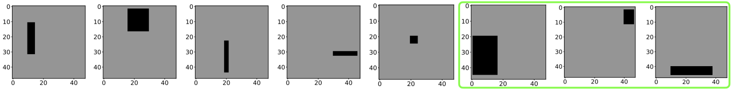



Figure 8: The 5 different obstacle configurations. The 3 unseen geometries are highlighted.

# Experiments: impact of obstacles

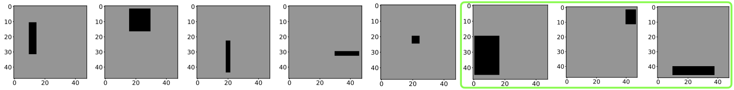
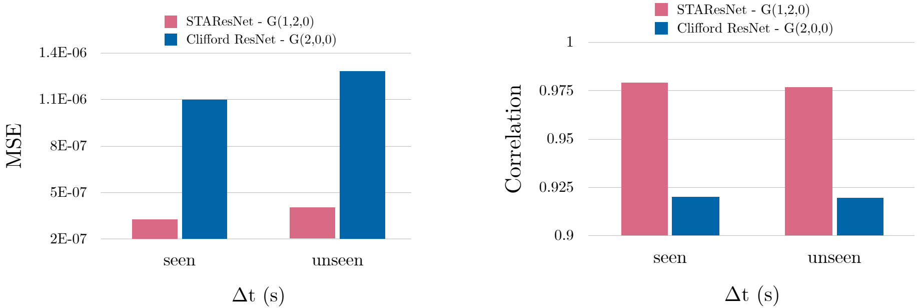


Figure 8: The 5 different obstacle configurations. The 3 unseen geometries are highlighted.




(a) MSE (↓). (b) Correlation (↑).

Figure 9: (a) MSE and (b) correlation between estimated and GT fields with obstacles.

# Experiments: impact of obstacles

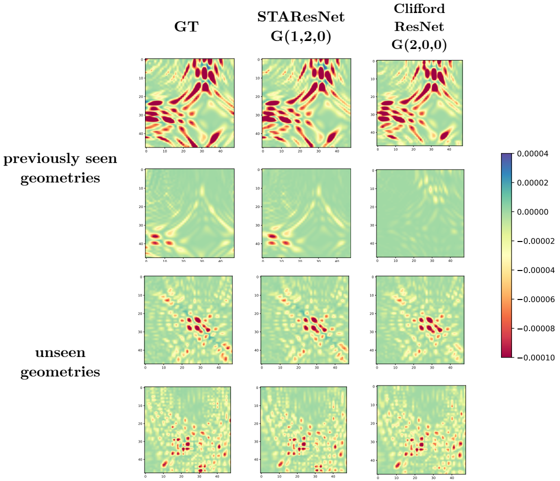



Figure 10: Squared magnitude of the Faraday bivector $\mathbf{F}^2$ over the test set with seen and unseen obstacles configurations.

# Experiments: impact of trainable parameters

STA is a $(n+1)$D space compared to $n$D GA: for the same number of channels or for the same size of the convolutional filters, STAResNet will have a larger number of trainable parameters.

# Experiments: impact of trainable parameters

STA is a $(n+1)$D space compared to $n$D GA: for the same number of channels or for the same size of the convolutional filters, STAResNet will have a larger number of trainable parameters.

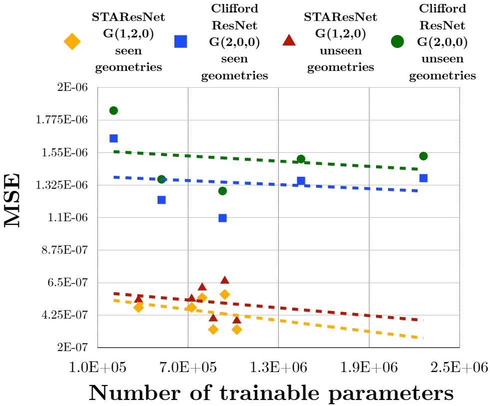


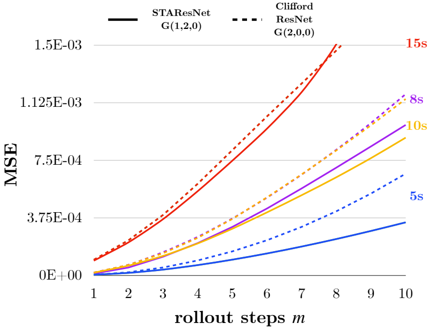

Figure 11: Test error over the estimated EM fields in the presence of seen and unseen obstacle geometries versus the number of trainable parameters.

# Experiments: rollout error

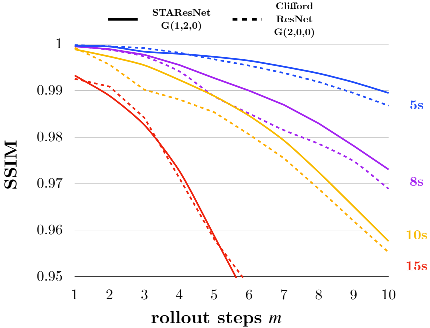
*Rollout* refers to the process of using the model's own predictions as inputs to generate future predictions.

# Experiments: rollout error

*Rollout* refers to the process of using the model's own predictions as inputs to generate future predictions.



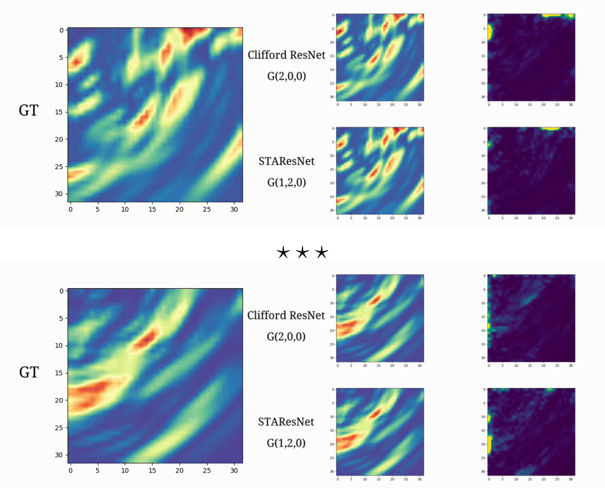

(a) MSE (↓) versus rollout steps.




(b) SSIM (↑) versus rollout steps.

Figure 12: (a) Mean squared error and (b) correlation between estimated and ground truth EM fields over test set versus rollout steps $m$ for the 2D case.

# Experiments: rollout error - visualizing $\mathbf{F}^2$



# Experiments: 3D fields

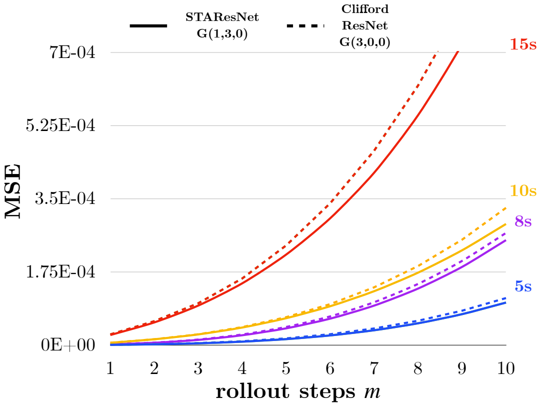
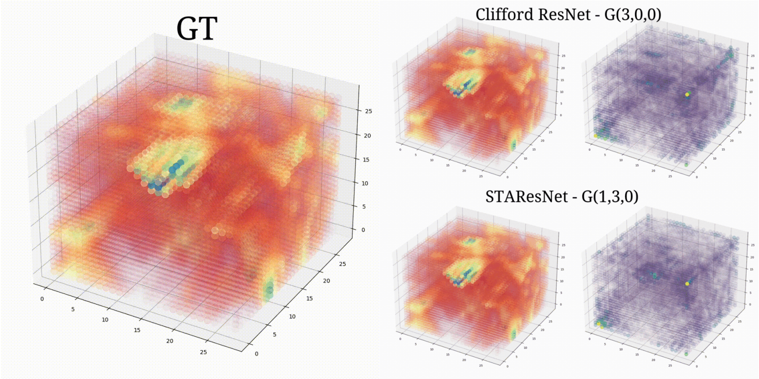
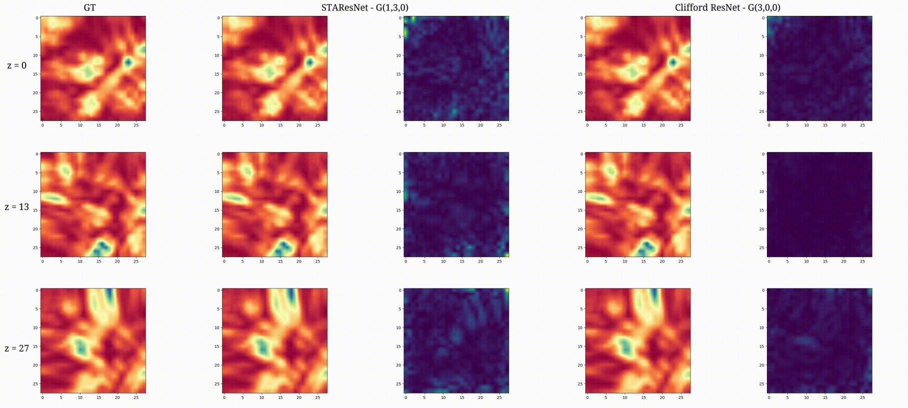



Figure 13: MSE between estimated and GT EM fields vs rollout steps $m$ for the 3D case.

# Experiments: 3D fields - visualizing $\mathbf{F}^2$

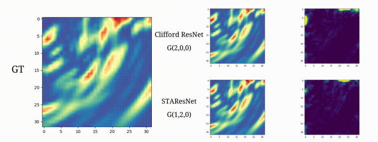


# Experiments: 3D fields - visualizing $\mathbb{F}^2$



# Table of Contents

# Conclusion



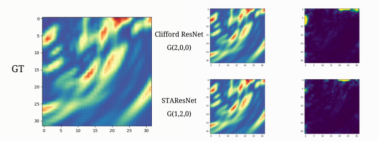

code

paper

**With STAResNet, we**:

- Shed light on the importance of the right algebra in which to embed data

# Conclusion

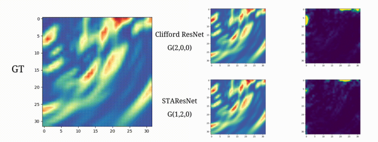



code paper

**With STAResNet, we**:

- Shed light on the importance of the right algebra in which to embed data
- Showed improvement in the PDEs solution over a vanilla GA network

# Conclusion




code

paper

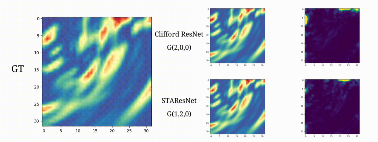
**With STAResNet, we**:

- Shed light on the importance of the right algebra in which to embed data
- Showed improvement in the PDEs solution over a vanilla GA network

**Why does it matter?**

- Many problems in ML deal with geometrical data

# Conclusion




code

paper

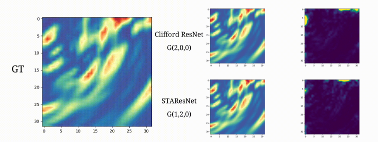
**With STAResNet, we**:

- Shed light on the importance of the right algebra in which to embed data
- Showed improvement in the PDEs solution over a vanilla GA network

**Why does it matter?**

- Many problems in ML deal with geometrical data
- Very few models can capture the geometry of data like GA networks

# Conclusion




code

paper

**With STAResNet, we**:

- Shed light on the importance of the right algebra in which to embed data
- Showed improvement in the PDEs solution over a vanilla GA network

**Why does it matter?**

- Many problems in ML deal with geometrical data
- Very few models can capture the geometry of data like GA networks
- Growing interest in industry: Huawei, Microsoft, ...

**Acknowledgements:** Joan Lasenby, Sven Buchholz, Christian Hockey, David Bowie