

# Multi Resolution Analysis (MRA) for Approximate Self-Attention

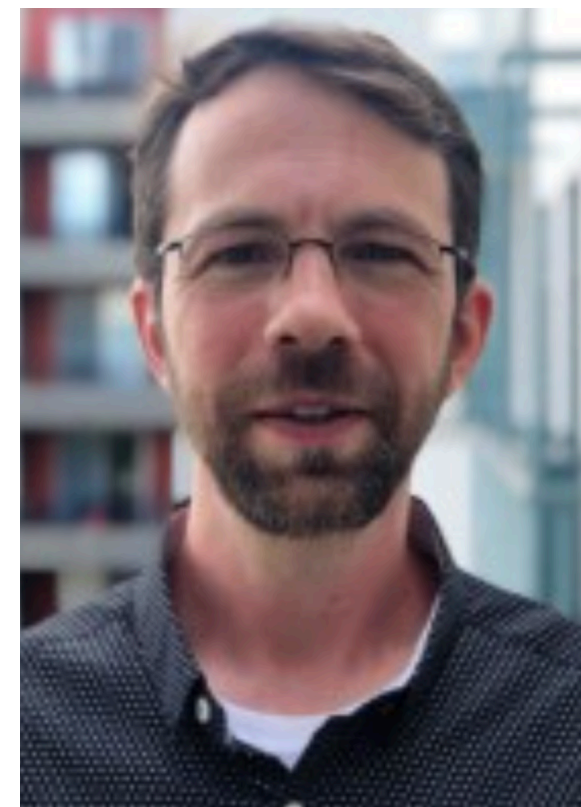
**Zhanpeng Zeng**



**Sourav Pal**



**Jeffery Kline**



**Glenn Fung**



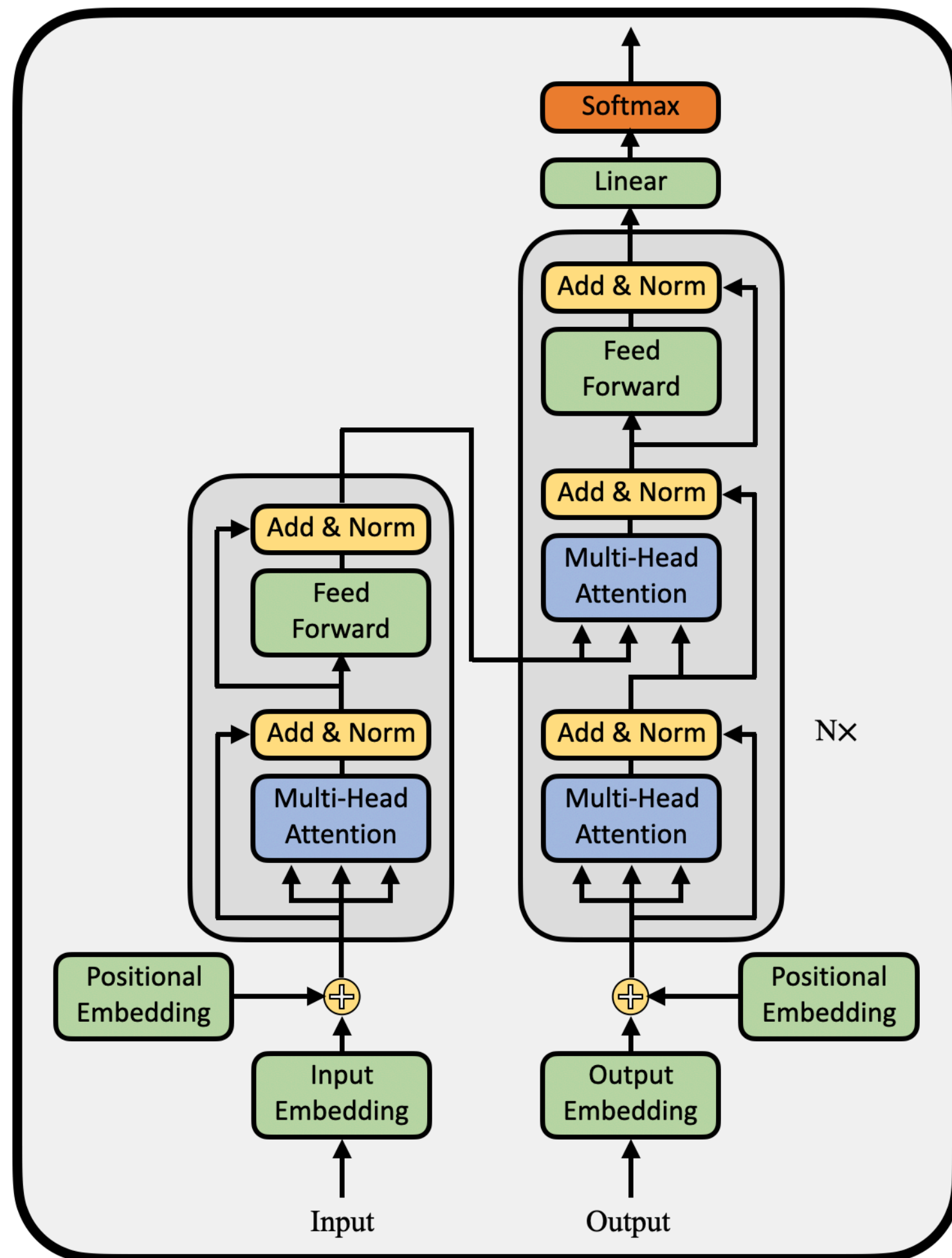
**Vikas Singh**



# **Motivation Transformer**

# Motivation

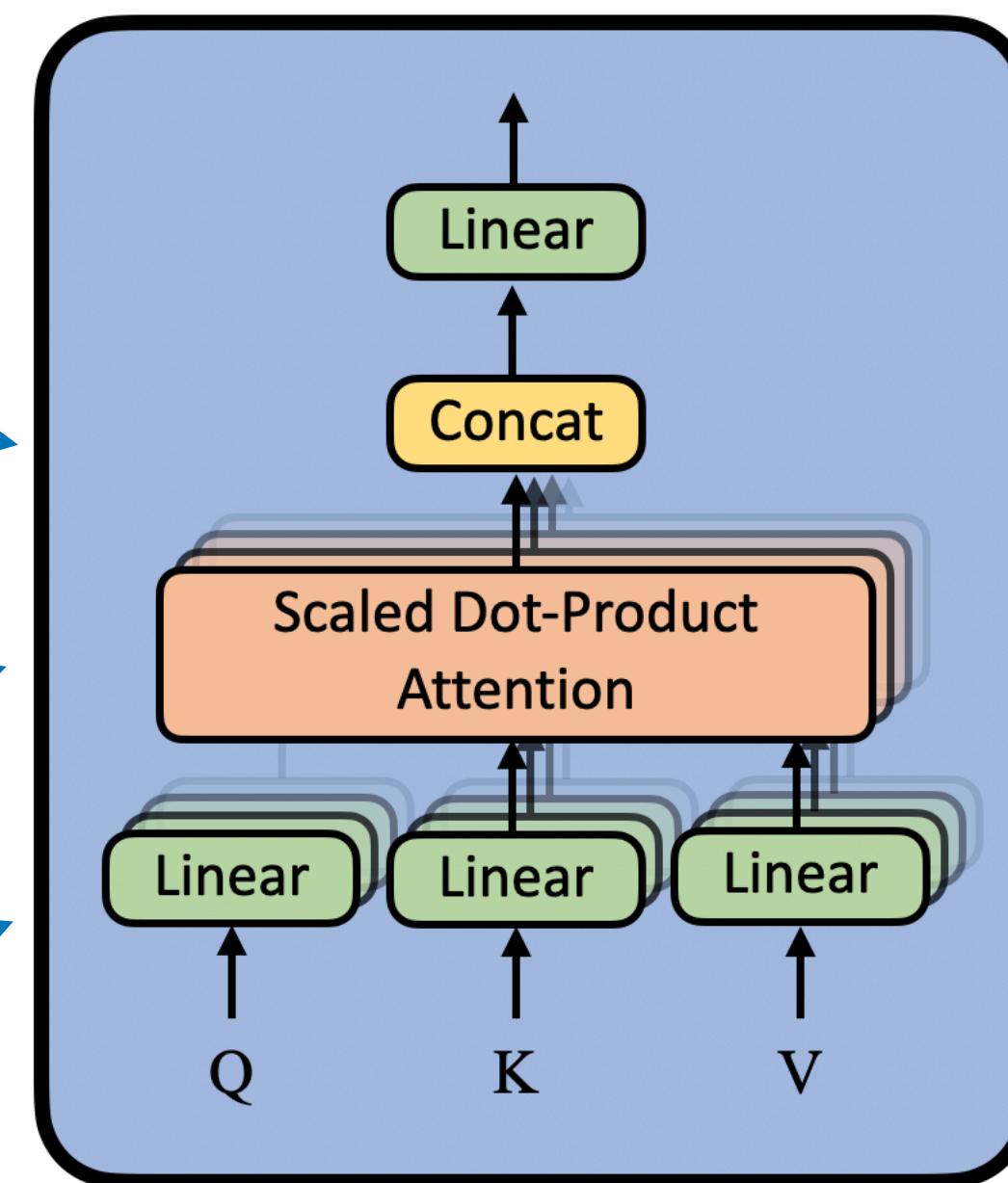
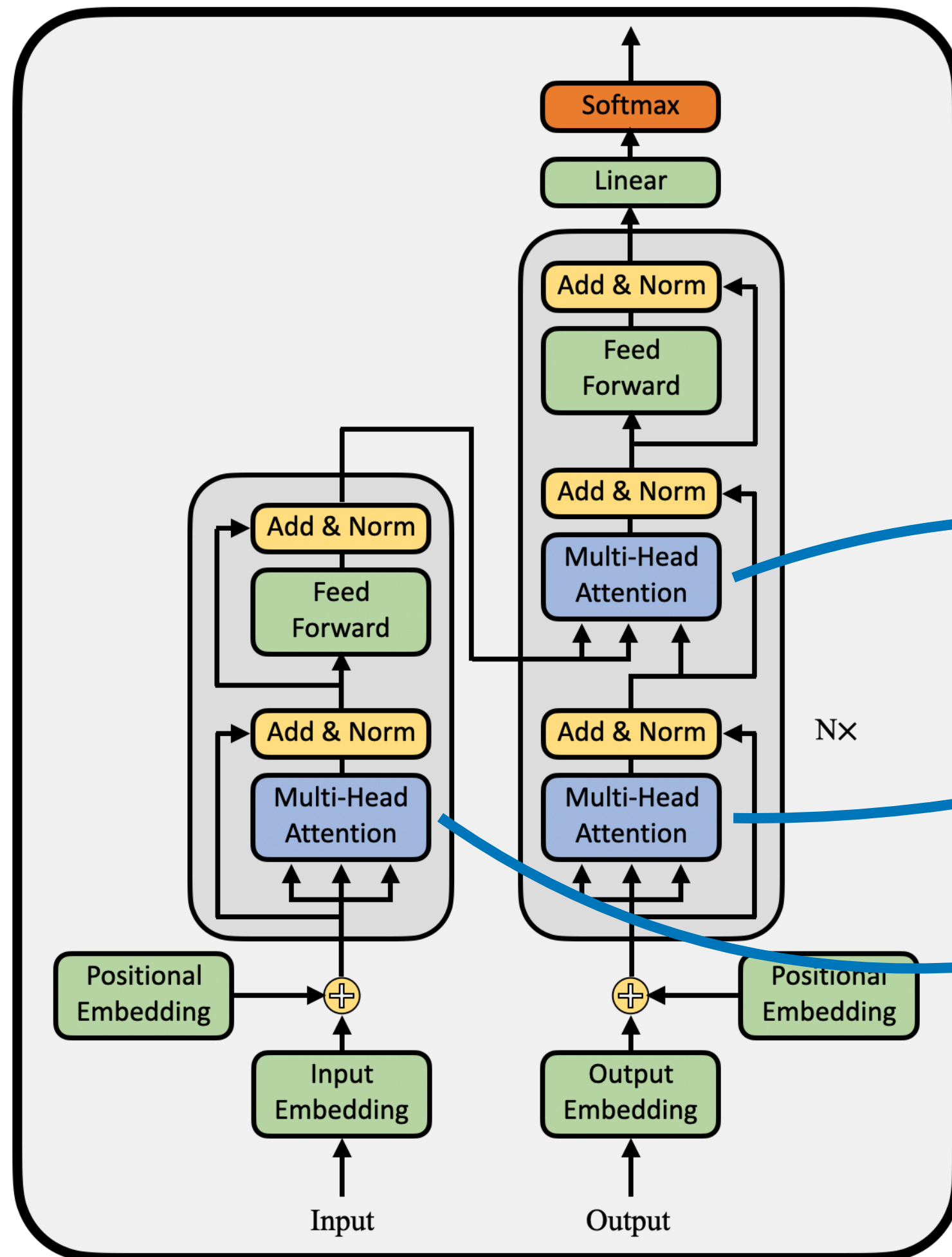
## Transformer





# Motivation

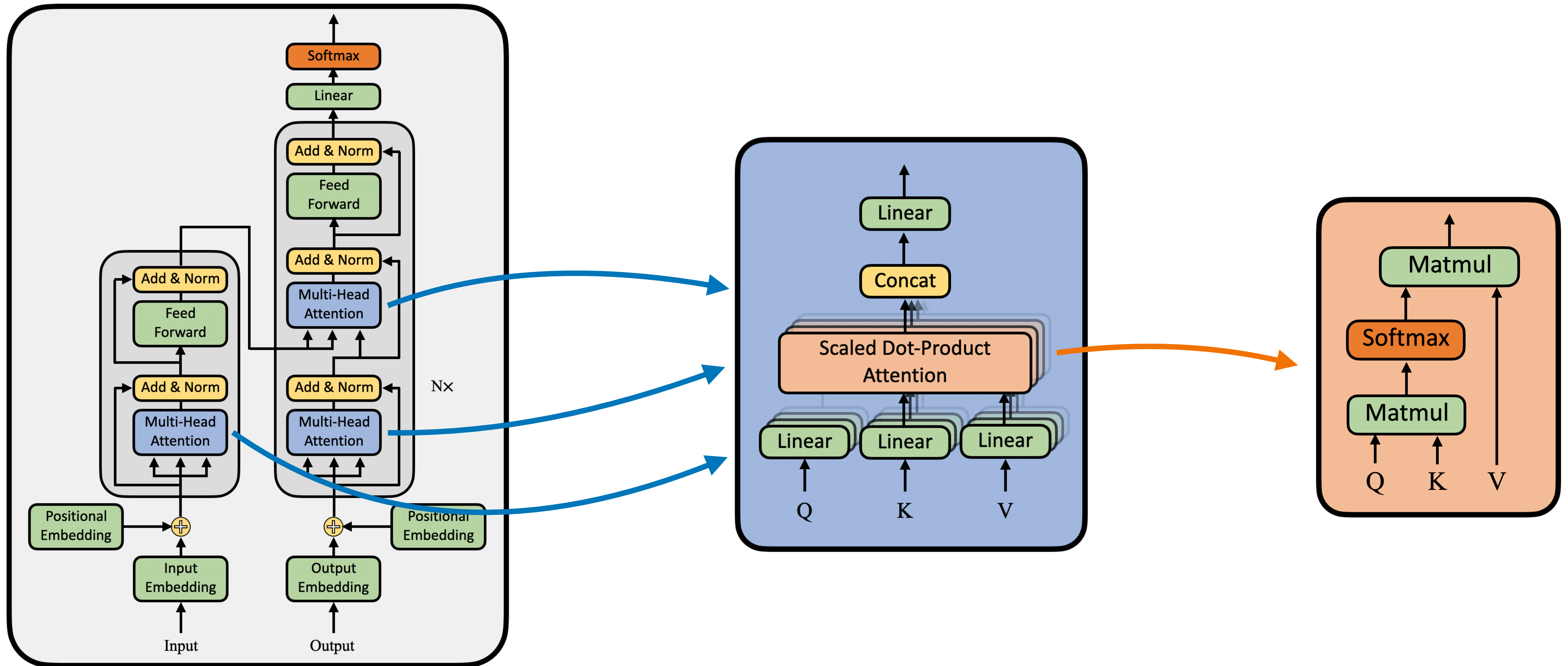
## Transformer





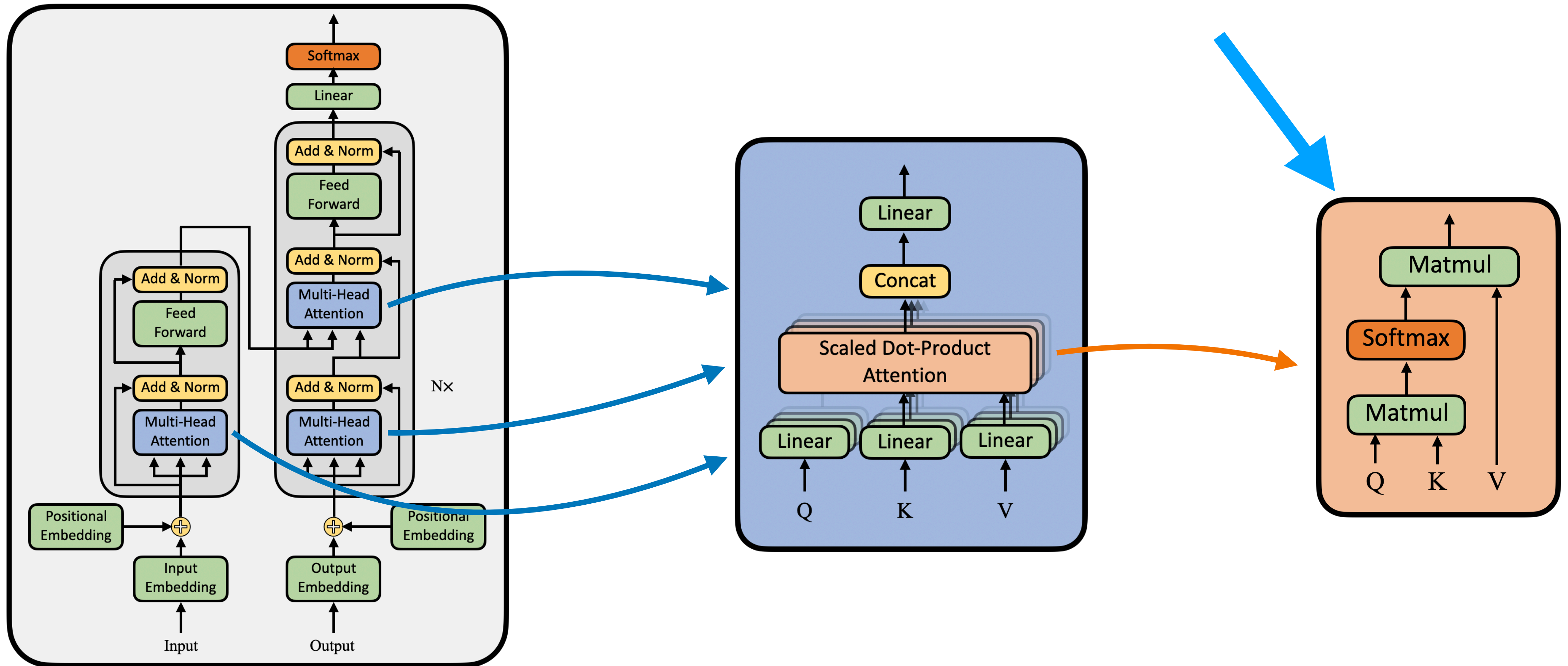
# Motivation

## Transformer



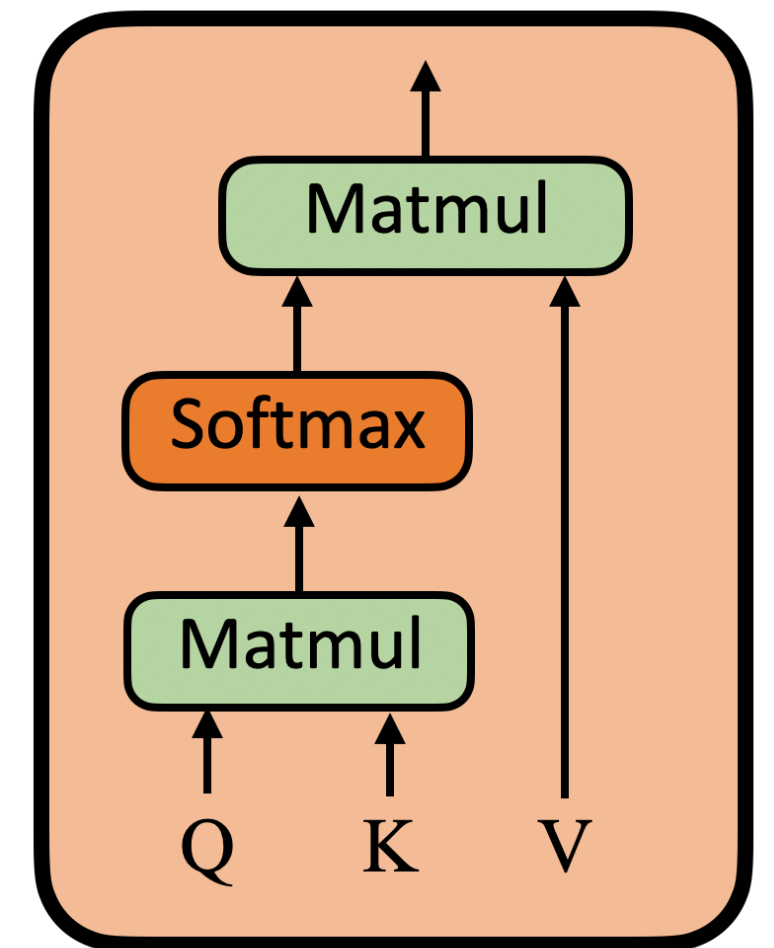
# Motivation

## Transformer



# Motivation

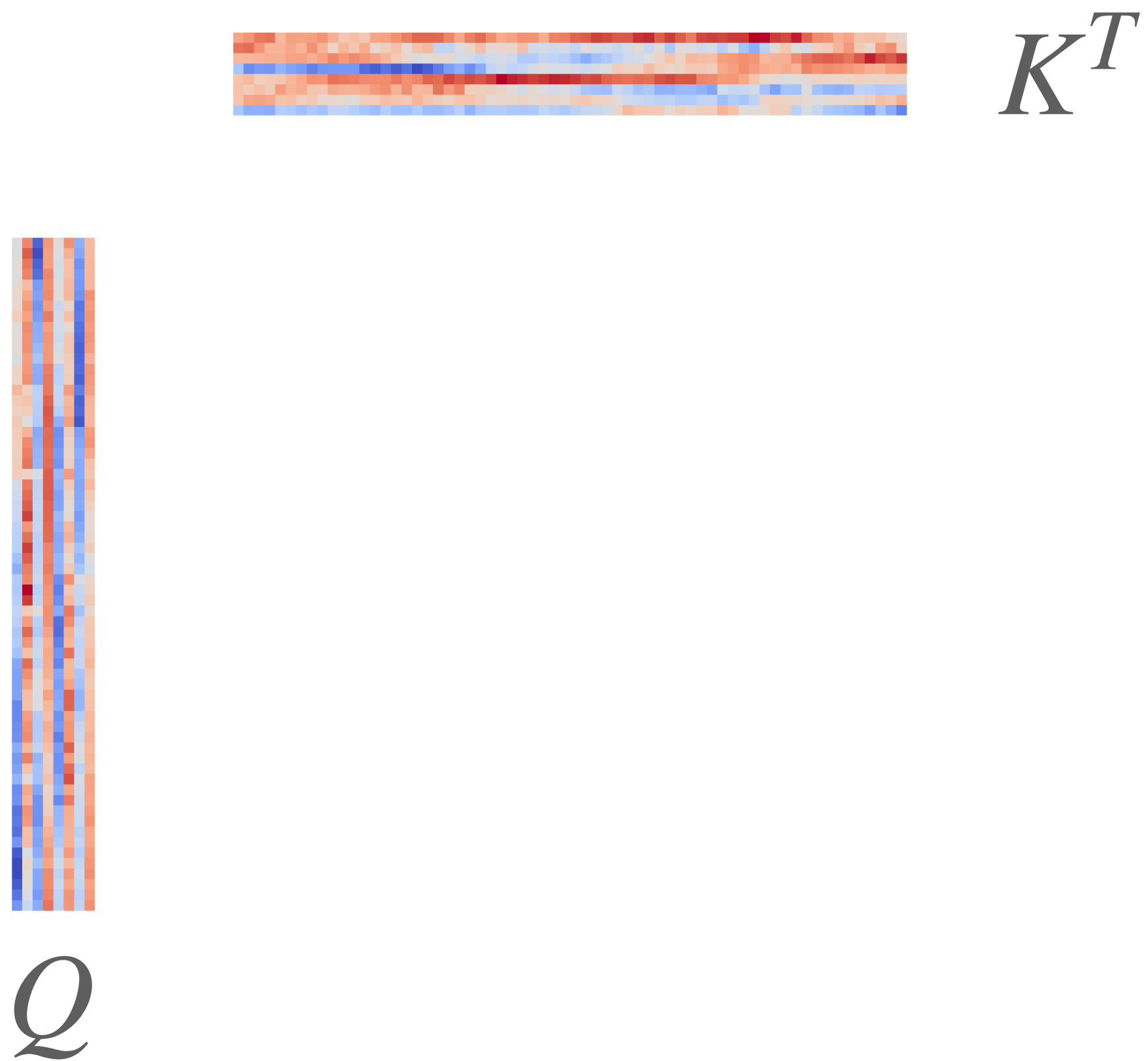
## Transformer



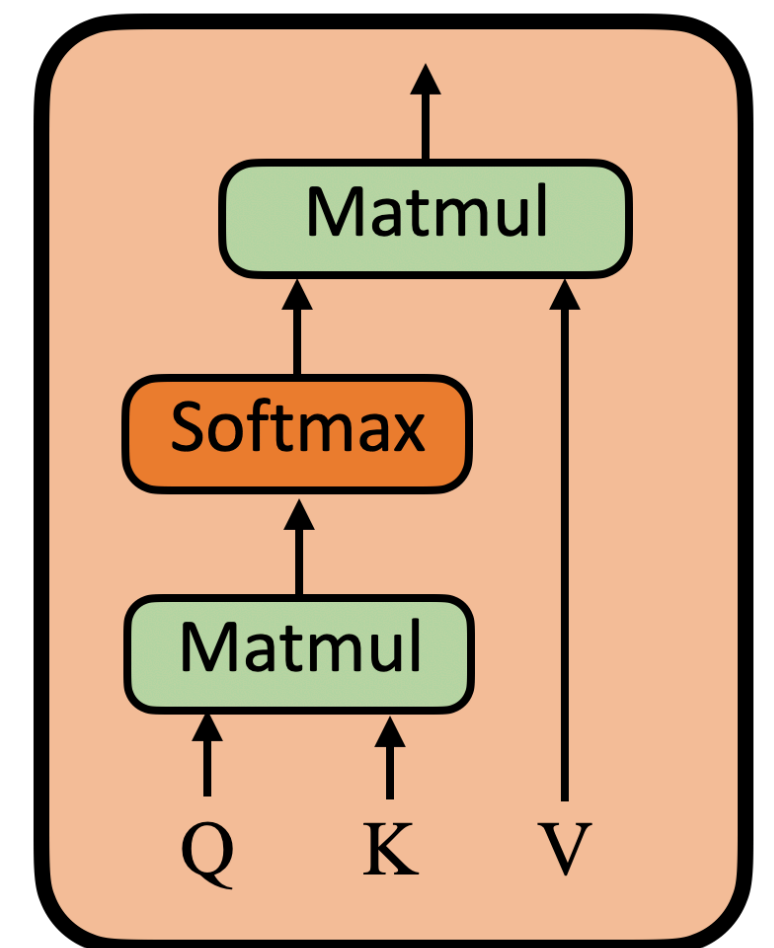


# Motivation

## Transformer

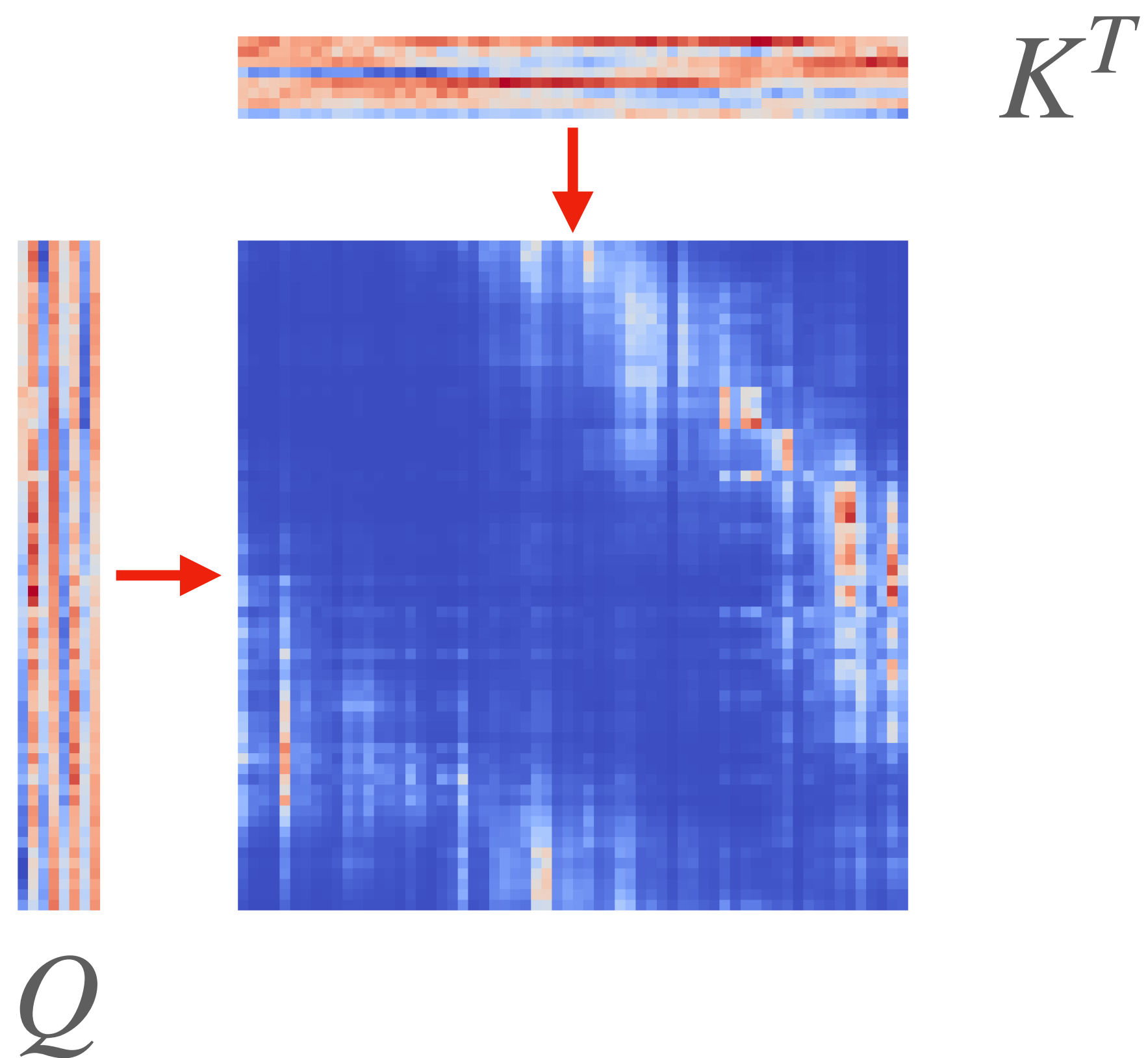


$$Q, K, V \in \mathbb{R}^{n \times d}$$



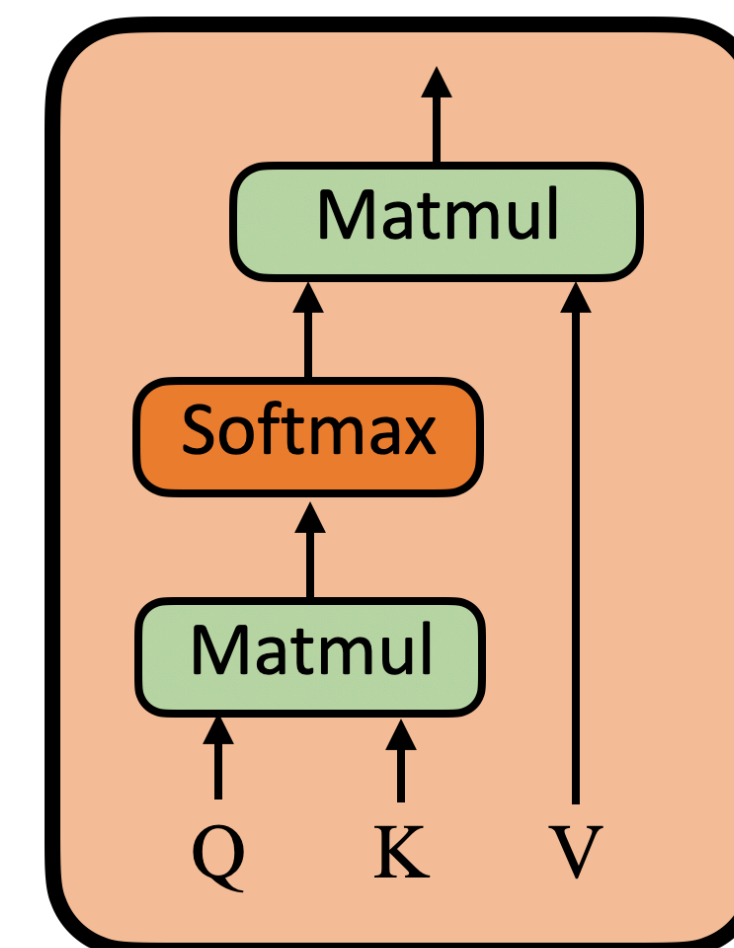
# Motivation

## Transformer



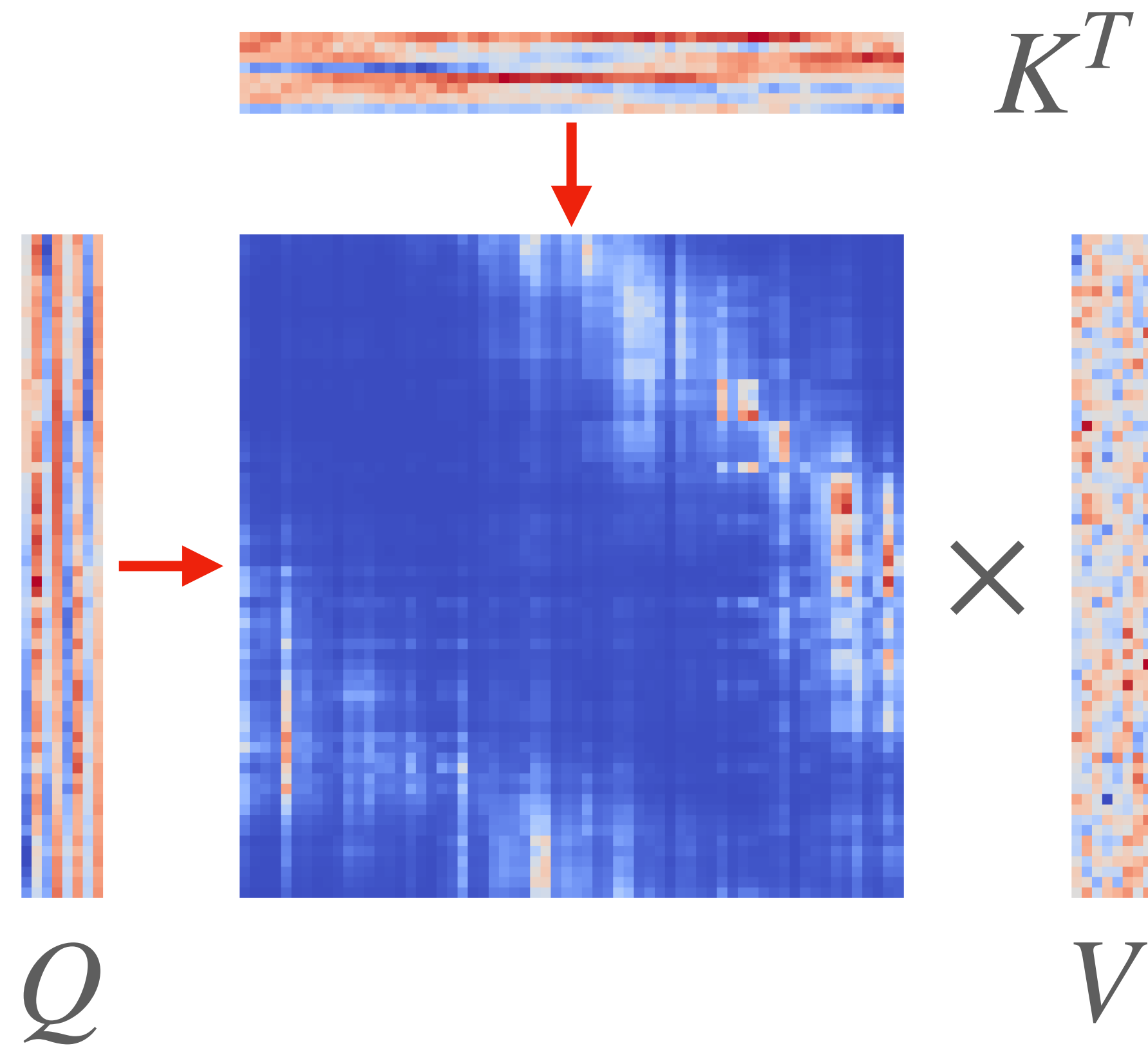
$$Q, K, V \in \mathbb{R}^{n \times d}$$

$$\exp(QK^T)$$



# Motivation

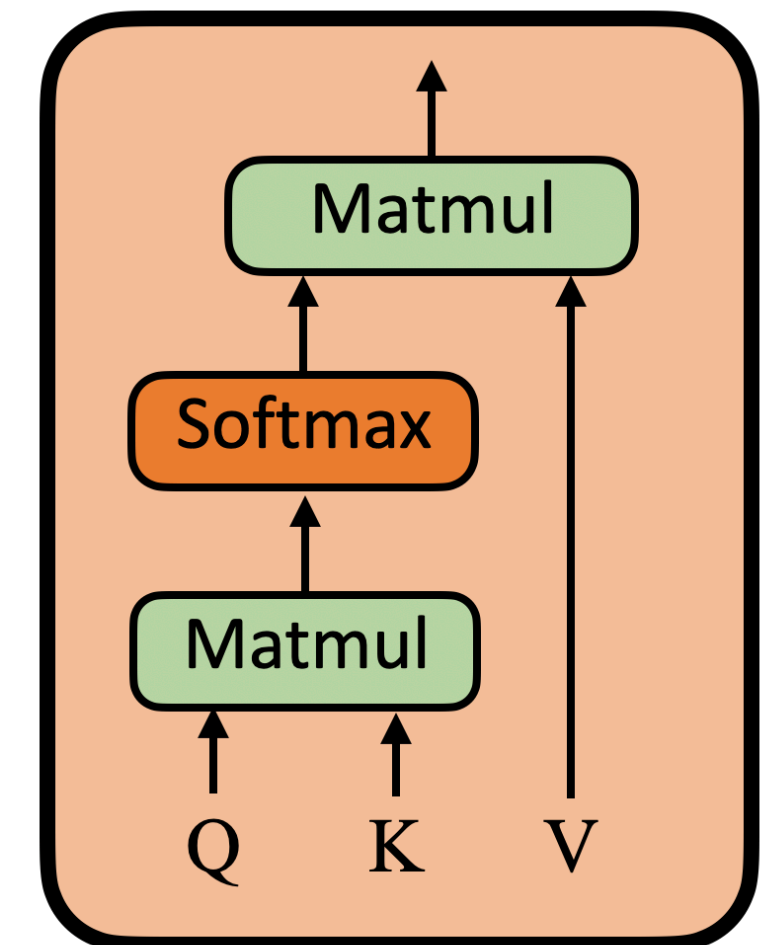
## Transformer



$$Q, K, V \in \mathbb{R}^{n \times d}$$

$$\exp(QK^T)$$

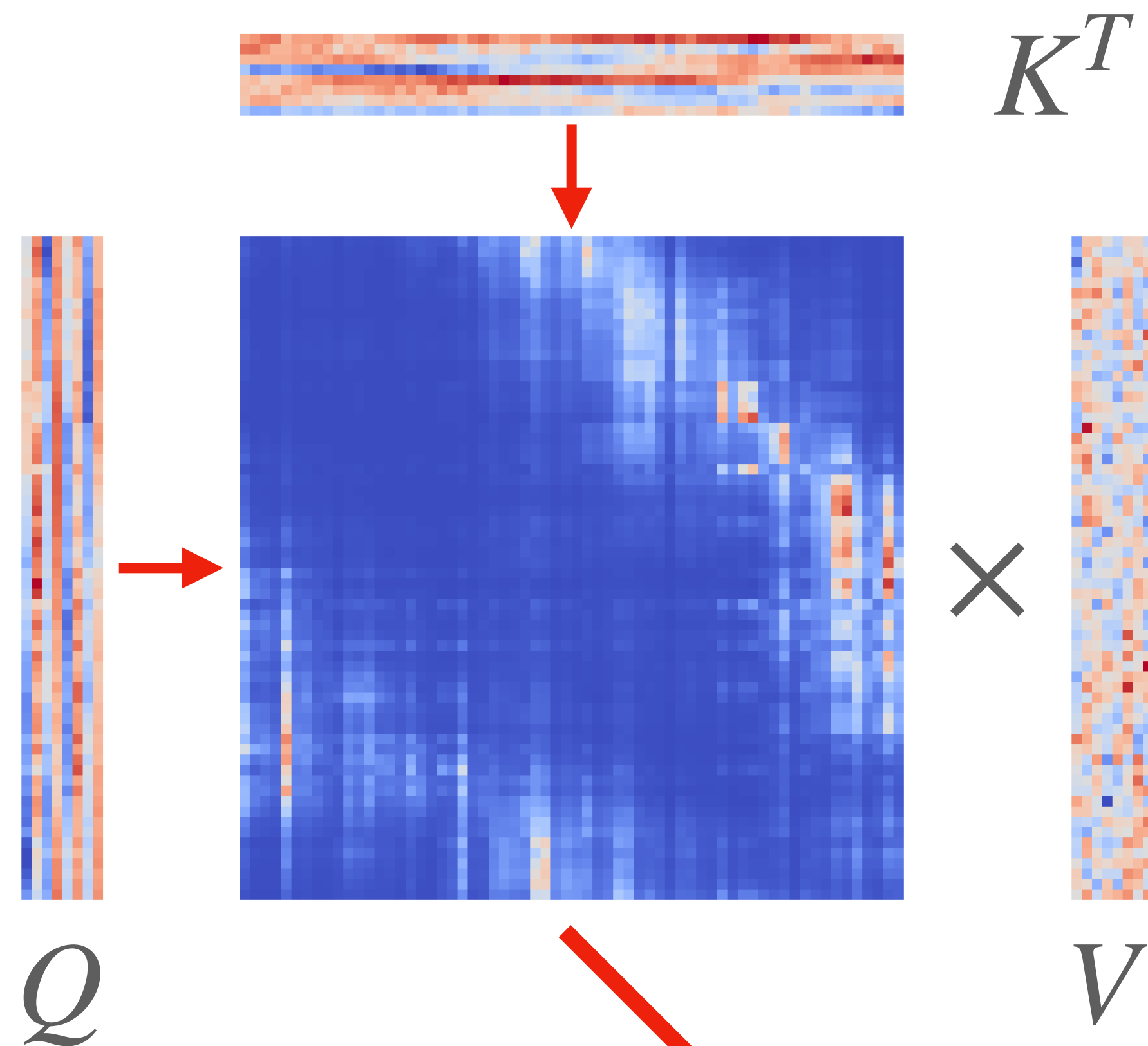
$$\exp(QK^T)V$$





# Motivation

## Transformer

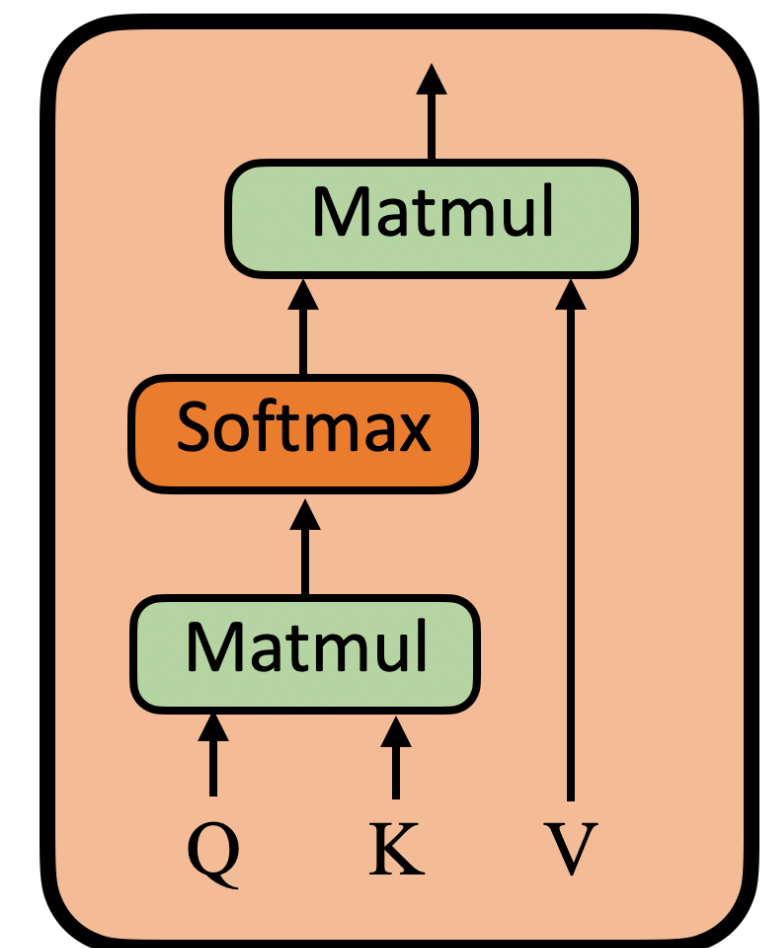


$$Q, K, V \in \mathbb{R}^{n \times d}$$

$$\exp(QK^T)$$

$$\exp(QK^T)V$$

Challenge:  $O(n^2)$  complexity



# Related Works

## REFORMER: THE EFFICIENT TRANSFORMER

**Nikita Kitaev\***  
U.C. Berkeley & Google Research  
kitaev@cs.berkeley.edu

**Łukasz Kaiser\***  
Google Research  
{lukaszkaizer, levskoy}

## Longformer: The Long-Document Transformer

## H-Transformer-1D: Fast One-Dimensional Hierarchical Attention for Sequences

**Zhenhai Zhu**  
Google Research  
zhenhai@google.com

**Radu Soricut**  
Google Research  
rsoricut@google.com

# Related Works

**REFORMER: THE EFFICIENT TRANSFORMER**

**Longformer: The Long-Document Transformer**

**Linformer: Self-Attention with Linear Complexity**

**Fast One-Dimensional Hierarchical Attention for Sequences**

Sinong Wang, Belinda Z. Li, Madian Khabsa, Han Fang, Hao Ma  
Facebook AI, Seattle, WA  
{sinongwang, belindali, hanfang, mkhabasa, haom}@fb.com

**Free Transformer with Linear Complexity**

**Nyströmformer: A Nyström-based Algorithm for Approximating Self-Attention**

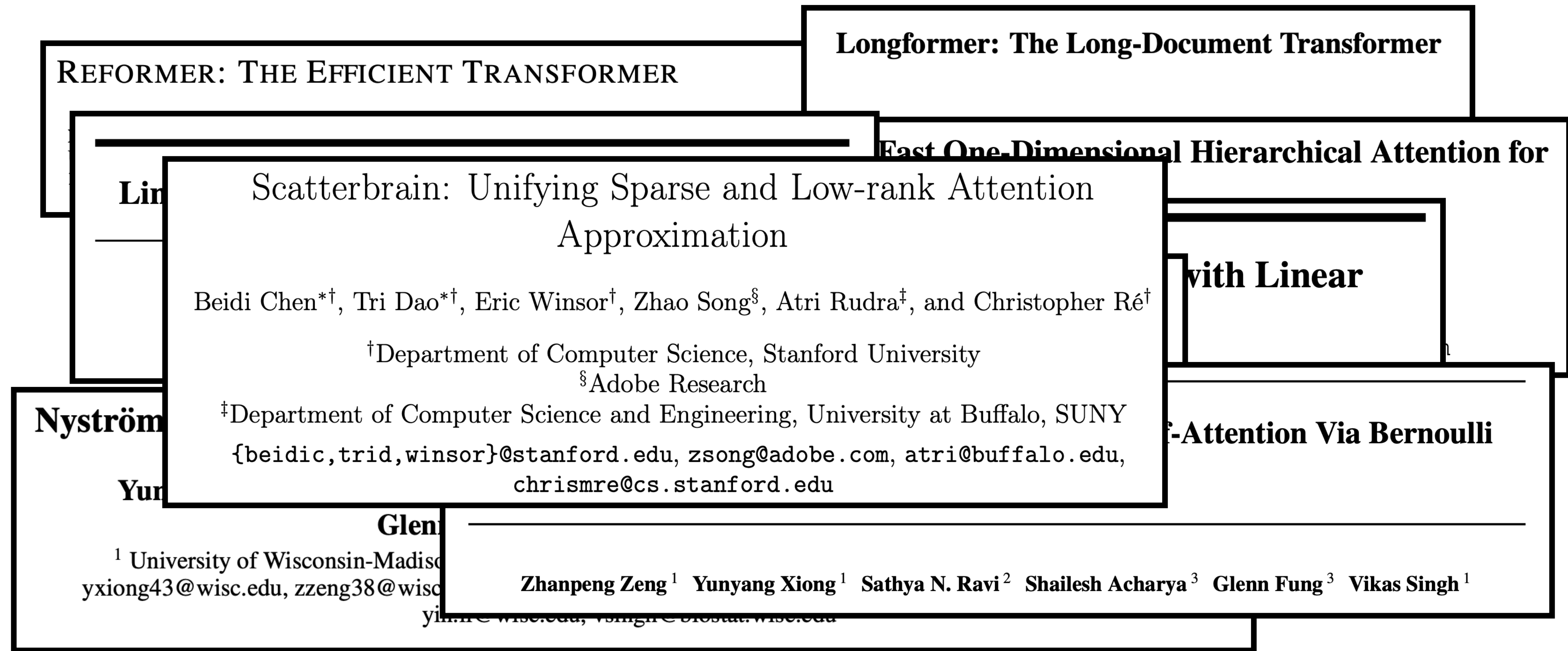
**Yunyang Xiong<sup>1</sup>   Zhanpeng Zeng<sup>1</sup>   Rudrasis Chakraborty<sup>2</sup>   Mingxing Tan<sup>3</sup>  
Glenn Fung<sup>4</sup>   Yin Li<sup>1</sup>   Vikas Singh<sup>1</sup>**

<sup>1</sup> University of Wisconsin-Madison   <sup>2</sup> UC Berkeley   <sup>3</sup> Google Brain   <sup>4</sup> American Family Insurance  
yxiong43@wisc.edu, zzeng38@wisc.edu, rudra@berkeley.edu, tanmingxing@google.com, gfung@amfam.com,  
yin.li@wisc.edu, vsingh@biostat.wisc.edu

Hang Xu<sup>3</sup>  
1\*  
Ark Lab



# Related Works



# Related Works

## REFORMER: THE EFFICIENT TRANSFORMER

**Nikita Kitaev\***  
U.C. Berkeley & Google Research  
kitaev@cs.berkeley.edu

**Łukasz Kaiser\***  
Google Research  
{lukaszkaiser, levskaya}@google.com

**Anselm Levskaya**  
Google Research

## Longformer: The Long-Document Transformer

**Iz Beltagy\***

**Matthew E. Peters\***

**Arman Cohan\***

Allen Institute for Artificial Intelligence, Seattle, WA, USA  
{beltagy, matthewp, armanc}@allenai.org

## Big Bird: Transformers for Longer Sequences

**Manzil Zaheer,**  
Joshua Ainslie,  
Anirudh Ravula,

**Guru Guruganesh,**  
Chris Alberti,  
Qifan Wang,  
Li Yang,  
Google Research

**Avinava Dubey,**  
Santiago Ontanon,  
Philip Pham,  
Amr Ahmed

{manzilz, gurug, avinavadubey}@google.com

# Related Works

## REFORMER: THE EFFICIENT TRANSFORMER

**Nikita Kitaev\***  
U.C. Berkeley & Google Research  
kitaev@cs.berkeley.edu

**Łukasz Kaiser\***  
Google Research  
{lukaszkaiser, levskaya}@google.com

**Anselm Levskaya**  
Google Research

## Longformer: The Long-Document Transformer

**Iz Beltagy\***   **Matthew E. Peters\***   **Arman Cohan\***  
Allen Institute for Artificial Intelligence, Seattle, WA, USA  
{beltagy, matthewp, armanc}@allenai.org

## Big Bird: Transformers for Longer Sequences

**Manzil Zaheer,**   **Guru Guruganesh,**   **Avinava Dubey,**  
**Joshua Ainslie,**   **Chris Alberti,**   **Santiago Ontanon,**   **Philip Pham,**  
**Anirudh Ravula,**   **Qifan Wang,**   **Li Yang,**   **Amr Ahmed**  
Google Research  
{manzilz, gurug, avinavadubey}@google.com

## RETHINKING ATTENTION WITH PERFORMERS

**Krzysztof Choromanski\*<sup>1</sup>, Valerii Likhoshesterov\*<sup>2</sup>, David Dohan\*<sup>1</sup>, Xingyou Song\*<sup>1</sup>,  
Andreea Gane\*<sup>1</sup>, Tamas Sarlos\*<sup>1</sup>, Peter Hawkins\*<sup>1</sup>, Jared Davis\*<sup>3</sup>, Afroz Mohiuddin<sup>1</sup>,  
Łukasz Kaiser<sup>1</sup>, David Belanger<sup>1</sup>, Lucy Colwell<sup>1,2</sup>, Adrian Weller<sup>2,4</sup>**  
<sup>1</sup>Google <sup>2</sup>University of Cambridge <sup>3</sup>DeepMind <sup>4</sup>Alan Turing Institute

## Linformer: Self-Attention with Linear Complexity

**Sinong Wang, Belinda Z. Li, Madian Khabsa, Han Fang, Hao Ma**  
Facebook AI, Seattle, WA  
{sinongwang, belindali, hanfang, mkhabsa, haom}@fb.com

## Nyströmformer: A Nyström-based Algorithm for Approximating Self-Attention

**Yunyang Xiong<sup>1</sup>   Zhanpeng Zeng<sup>1</sup>   Rudrasis Chakraborty<sup>2</sup>   Mingxing Tan<sup>3</sup>**  
**Glenn Fung<sup>4</sup>   Yin Li<sup>1</sup>   Vikas Singh<sup>1</sup>**  
<sup>1</sup> University of Wisconsin-Madison   <sup>2</sup> UC Berkeley   <sup>3</sup> Google Brain   <sup>4</sup> American Family Insurance  
yxiong43@wisc.edu, zzeng38@wisc.edu, rudra@berkeley.edu, tanmingxing@google.com, gfung@amfam.com,  
yin.li@wisc.edu, vsingh@biostat.wisc.edu

## SOFT: Softmax-free Transformer with Linear Complexity

**Jiachen Lu<sup>1</sup>   Jinghan Yao<sup>1</sup>   Junge Zhang<sup>1</sup>   Xiatian Zhu<sup>2</sup>   Hang Xu<sup>3</sup>**  
**Weiguo Gao<sup>1</sup>   Chunjing Xu<sup>3</sup>   Tao Xiang<sup>2</sup>   Li Zhang<sup>1\*</sup>**  
<sup>1</sup>Fudan University   <sup>2</sup>University of Surrey   <sup>3</sup>Huawei Noah's Ark Lab



# Related Works

REFORMER: THE EFFICIENT TRANSFORMER

Nikita Kitaev\*

U.C. Berkeley & Google Research

kitaev@cs.berkeley.edu

Łukasz Kaiser\*

Google Research

{lukaszkaizer, levskaya}@google.com

Anselm Levskaya

Google Research

RETHINKING ATTENTION WITH PERFORMERS

Krzysztof Choromanski\*<sup>1</sup>, Valerii Likhoshesterov\*<sup>2</sup>, David Dohan\*<sup>1</sup>, Xingyou Song\*<sup>1</sup>

Andreea Gane\*<sup>1</sup>, Tamas Sarlos\*<sup>1</sup>, Peter Hawkins\*<sup>1</sup>, Jared Davis\*<sup>3</sup>, Afroz Mohiuddin<sup>1</sup>

Lukasz Kaiser<sup>1</sup>, David Belanger<sup>1</sup>, Lucy Colwell<sup>1,2</sup>, Adrian Weller<sup>2,4</sup>

<sup>1</sup>Google <sup>2</sup>University of Cambridge <sup>3</sup>DeepMind <sup>4</sup>Alan Turing Institute

Longform

Iz Beltagy

Allen Institute for AI

{beltagy}

Scatterbrain: Unifying Sparse and Low-rank Attention Approximation

Beidi Chen\*<sup>†</sup>, Tri Dao\*<sup>†</sup>, Eric Winsor<sup>†</sup>, Zhao Song<sup>§</sup>, Atri Rudra<sup>‡</sup>, and Christopher Ré<sup>†</sup>

<sup>†</sup>Department of Computer Science, Stanford University

<sup>§</sup>Adobe Research

<sup>‡</sup>Department of Computer Science and Engineering, University at Buffalo, SUNY

{beidic, trid, winsor}@stanford.edu, zsong@adobe.com, atri@buffalo.edu, chrismre@cs.stanford.edu

Big Bird

Complexity

Hao Ma

@fb.com

Reducing Self-Attention Complexity

Mingxing Tan<sup>3</sup>

Canadian Family Insurance  
Company, gfung@amfam.com,

Manzil Zaheer,

Joshua Ainslie,

Anirudh Ravula,

Guru Guruganesh,

Chris Alberti,

Qifan Wang,

Avinava Dubey,

Philip Pham,

Amr Ahmed

Google Research

{manzilz, gurug, avinavadubey}@google.com

SOFT: Softmax-free Transformer with Linear Complexity

Jiachen Lu<sup>1</sup> Jinghan Yao<sup>1</sup> Junge Zhang<sup>1</sup> Xiatian Zhu<sup>2</sup> Hang Xu<sup>3</sup>

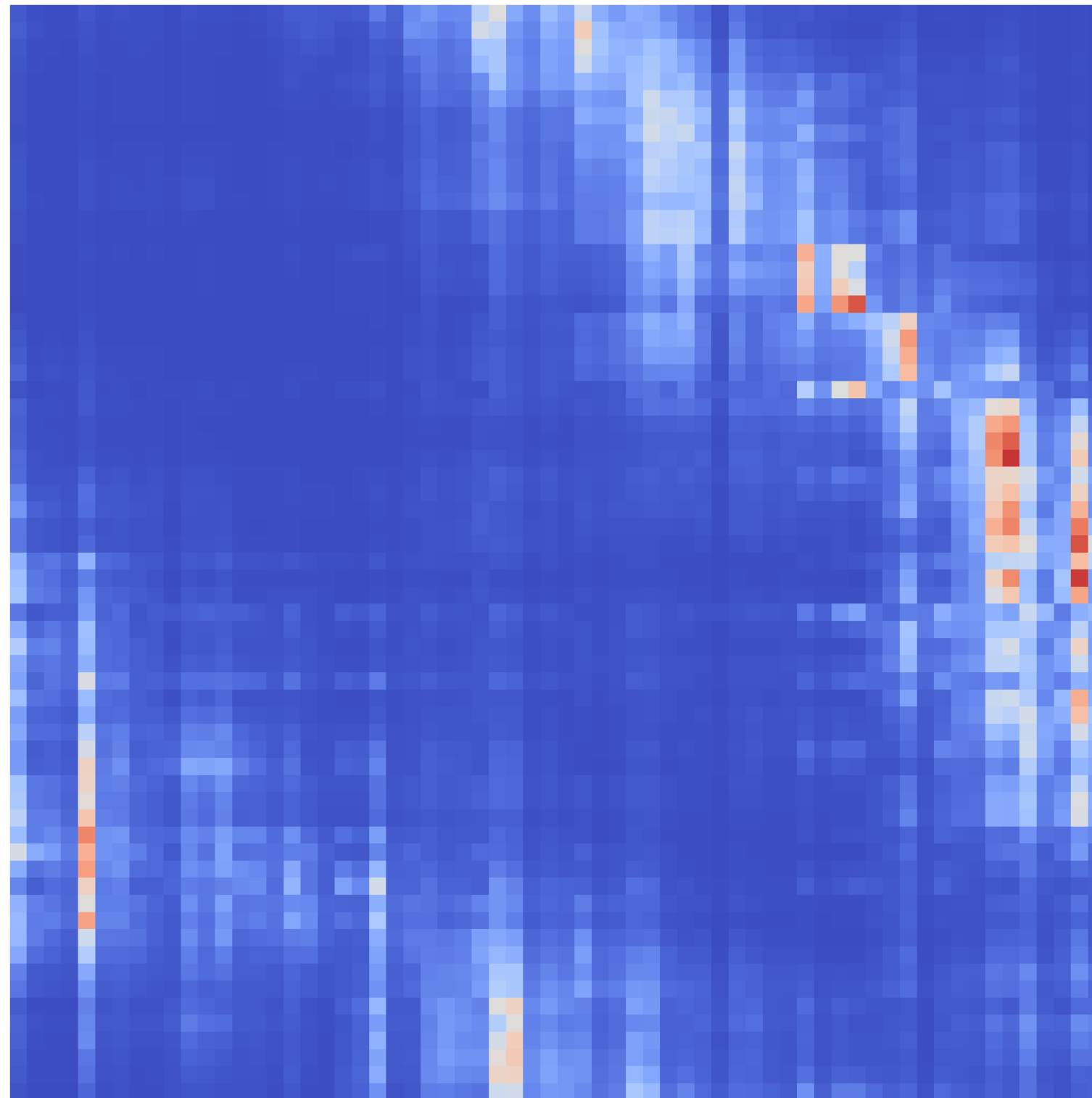
Weiguo Gao<sup>1</sup> Chunjing Xu<sup>3</sup> Tao Xiang<sup>2</sup> Li Zhang<sup>1\*</sup>

<sup>1</sup>Fudan University <sup>2</sup>University of Surrey <sup>3</sup>Huawei Noah's Ark Lab

# Multi-Resolution Approximation

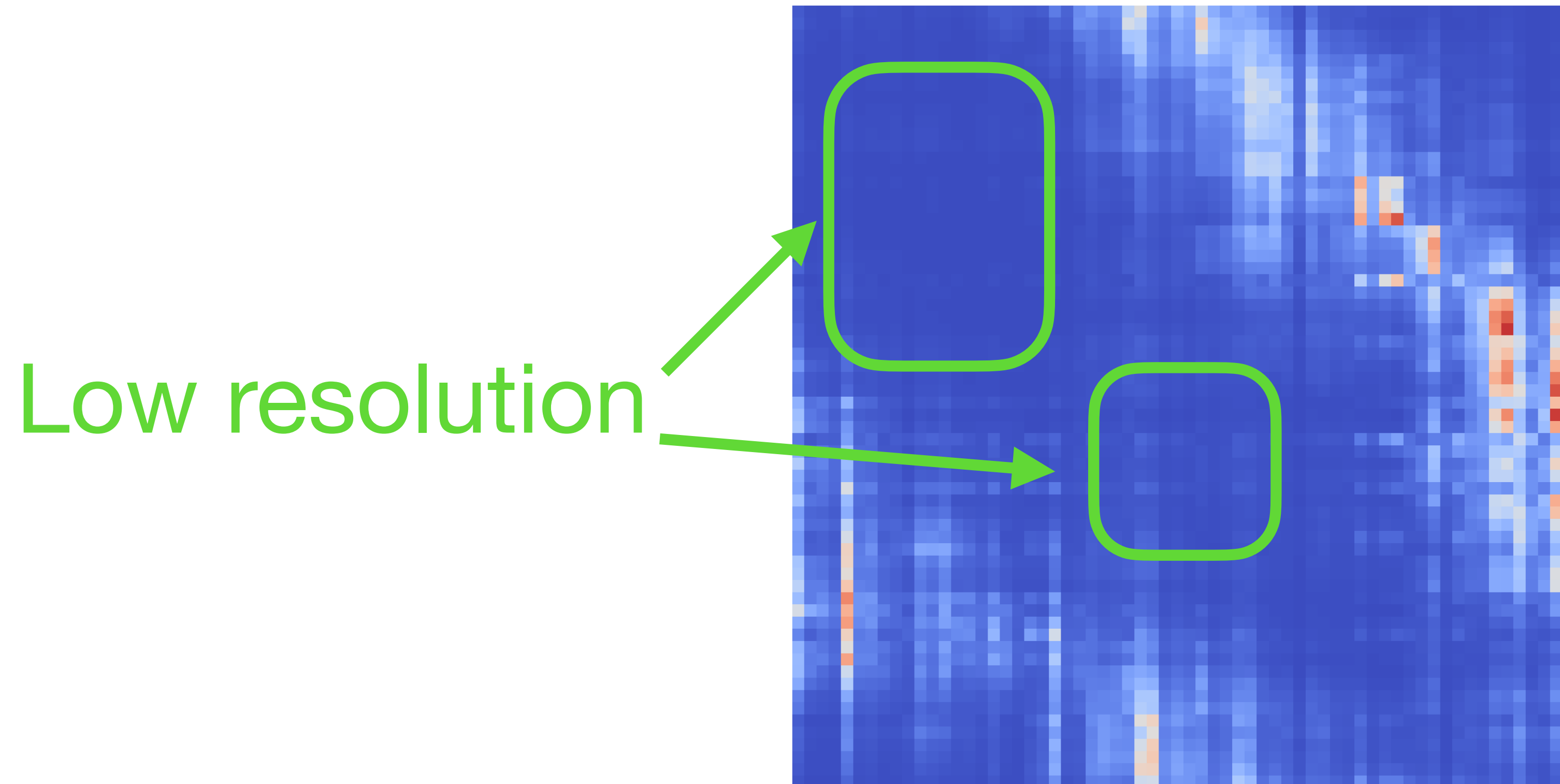
# Multi-Resolution Approximation

Why?



# Multi-Resolution Approximation

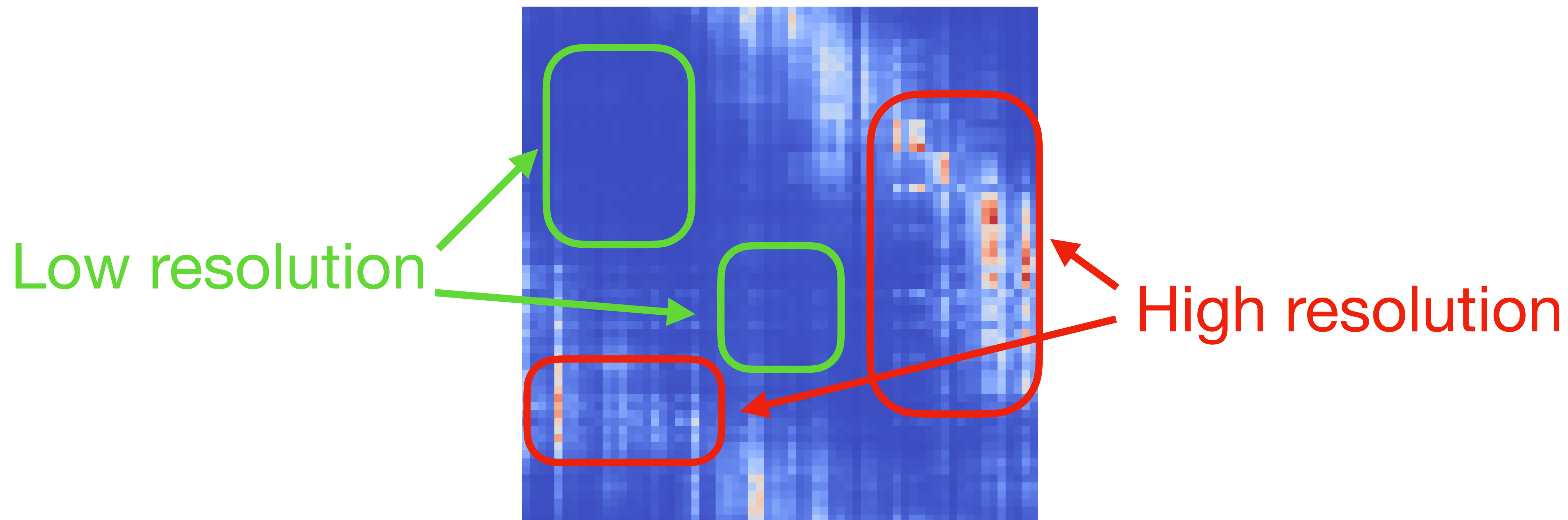
Why?





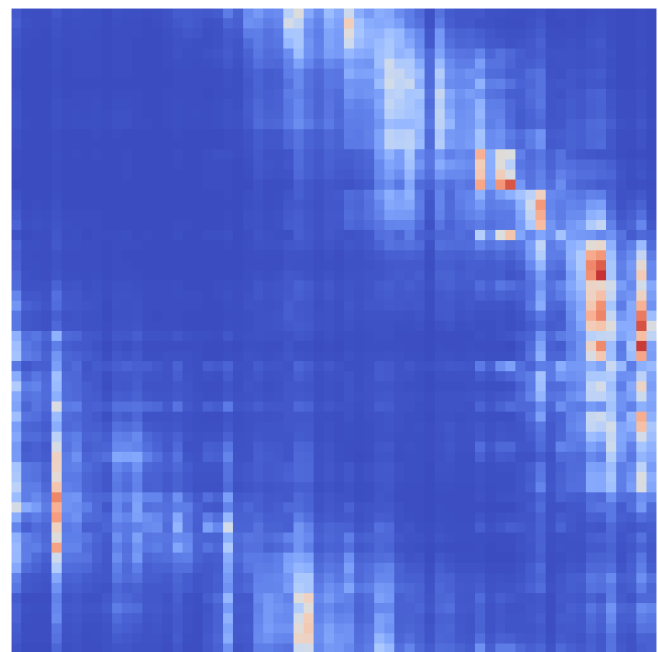
# Multi-Resolution Approximation

Why?



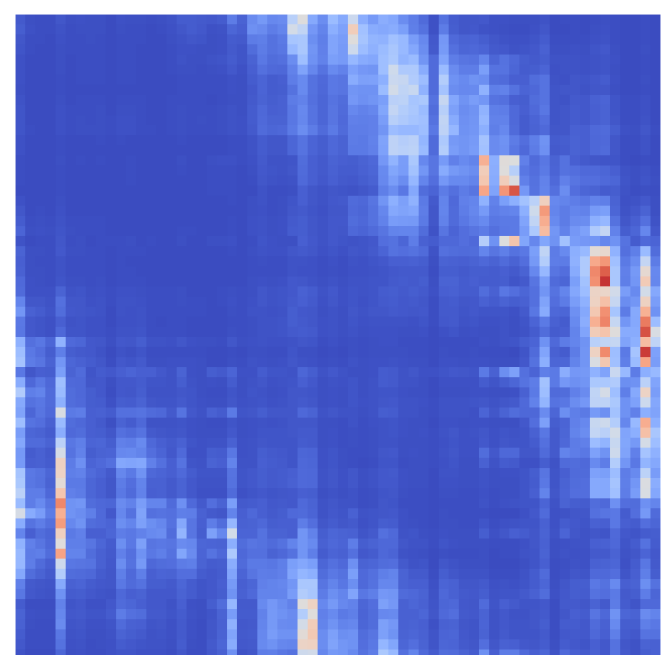
# Multi-Resolution Approximation

## Wavelets



# Multi-Resolution Approximation

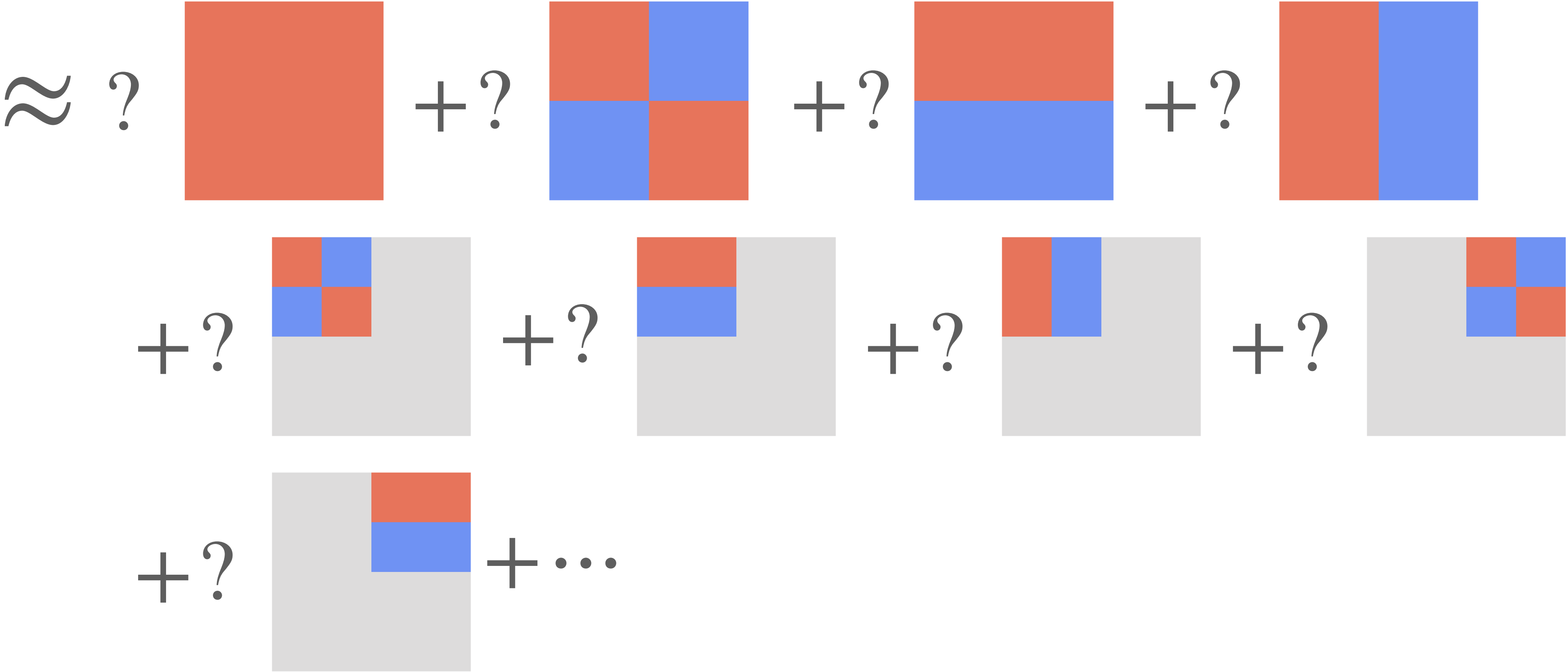
## Wavelets



$$\begin{aligned} \approx & \alpha_1 \begin{array}{|c|} \hline \text{Red} \\ \hline \end{array} + \alpha_2 \begin{array}{|c|c|} \hline \text{Red} & \text{Blue} \\ \hline \text{Blue} & \text{Red} \end{array} + \alpha_3 \begin{array}{|c|} \hline \text{Red} \\ \hline \text{Blue} \end{array} + \alpha_4 \begin{array}{|c|} \hline \text{Red} \\ \hline \text{Blue} \end{array} \\ & + \alpha_5 \begin{array}{|c|c|} \hline \text{Red} & \text{Blue} \\ \hline \text{Blue} & \text{Red} \end{array} + \alpha_6 \begin{array}{|c|c|} \hline \text{Red} & \text{Blue} \\ \hline \text{Red} & \text{Blue} \end{array} + \alpha_7 \begin{array}{|c|c|} \hline \text{Red} & \text{Blue} \\ \hline \text{Red} & \text{Blue} \end{array} + \alpha_8 \begin{array}{|c|c|} \hline \text{Red} & \text{Blue} \\ \hline \text{Red} & \text{Blue} \end{array} \\ & + \alpha_9 \begin{array}{|c|c|} \hline \text{Red} & \text{Blue} \\ \hline \text{Red} & \text{Blue} \end{array} + \dots \end{aligned}$$

# Multi-Resolution Approximation

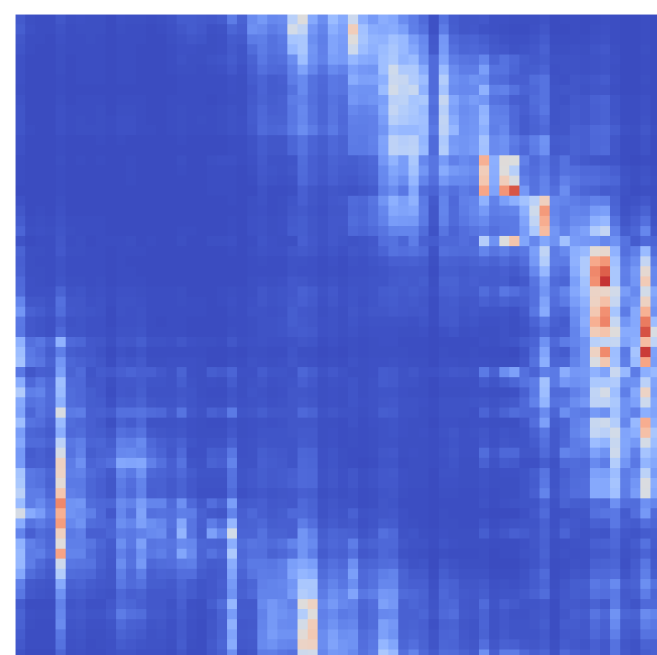
## Wavelets





# Multi-Resolution Approximation

## Wavelets



$$\begin{aligned} f(x) \approx & \mu_1 \begin{array}{|c|} \hline \text{red square in top-left} \\ \hline \end{array} + \mu_2 \begin{array}{|c|} \hline \text{red square in bottom-right} \\ \hline \end{array} + \mu_3 \begin{array}{|c|} \hline \text{red square in top-right} \\ \hline \end{array} + \mu_4 \begin{array}{|c|} \hline \text{red square in top-right} \\ \hline \end{array} \\ & + \mu_5 \begin{array}{|c|} \hline \text{red square in bottom-left} \\ \hline \end{array} + \mu_6 \begin{array}{|c|} \hline \text{red square in bottom-right} \\ \hline \end{array} + \mu_7 \begin{array}{|c|} \hline \text{red square in bottom-left} \\ \hline \end{array} + \mu_8 \begin{array}{|c|} \hline \text{red square in bottom-right} \\ \hline \end{array} \\ & + \mu_9 \begin{array}{|c|} \hline \text{red square in bottom-left} \\ \hline \end{array} + \dots \end{aligned}$$

# Multi-Resolution Approximation

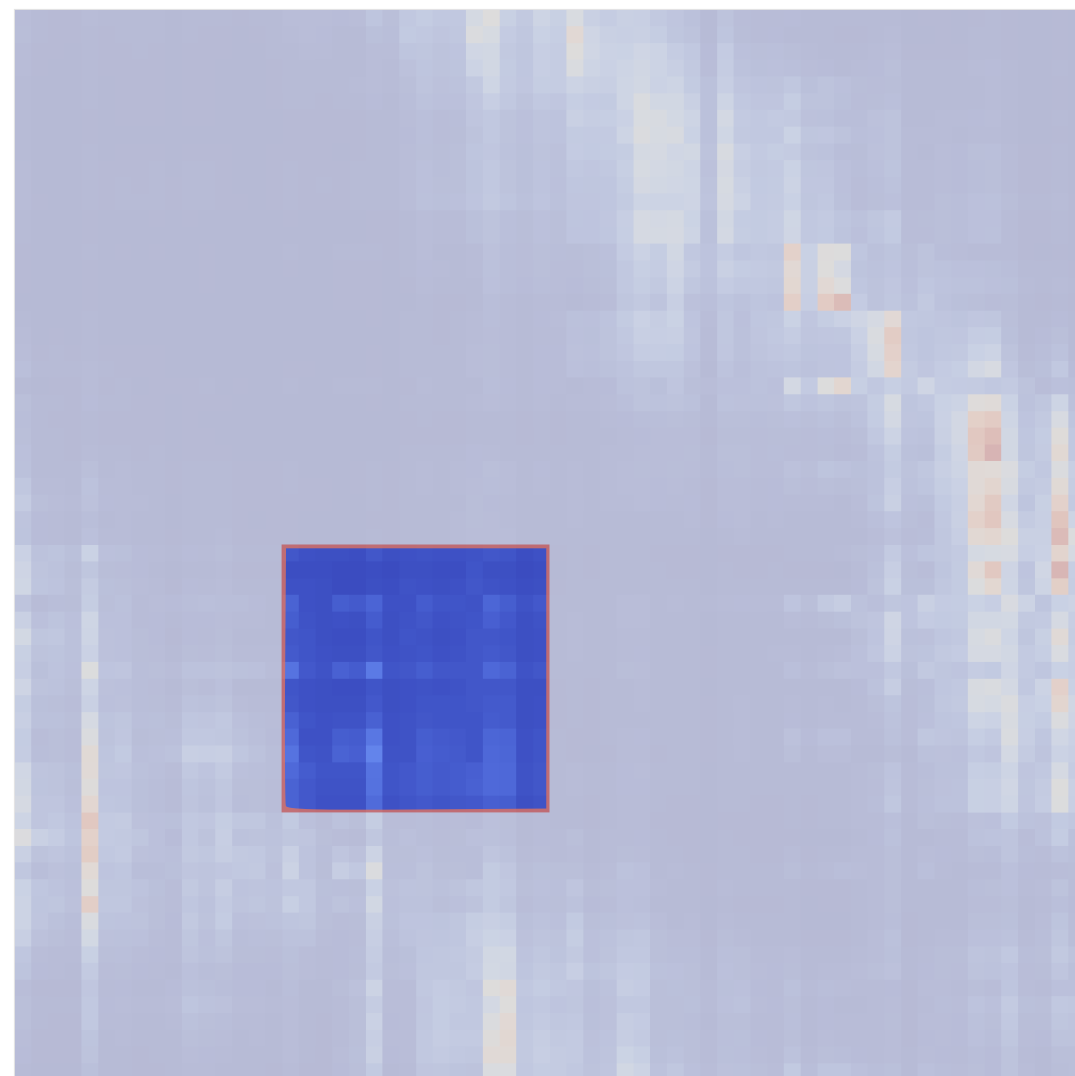
## Coefficient Approximation

$$\dots + \mu_5 \left[ \text{gray square with red square inside} \right] + \dots$$

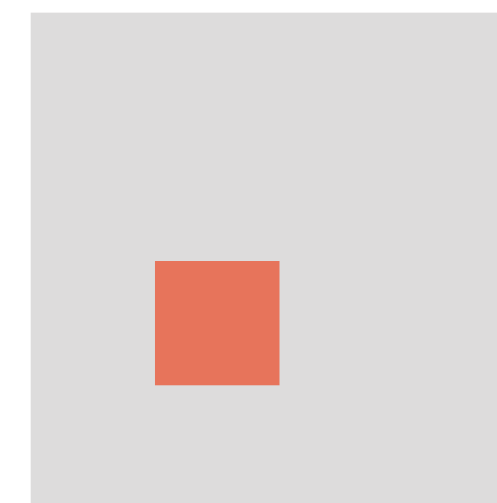
# Multi-Resolution Approximation

## Coefficient Approximation

$$\mu_5 = \text{Avg}(\quad)$$



$$\dots + \mu_5$$

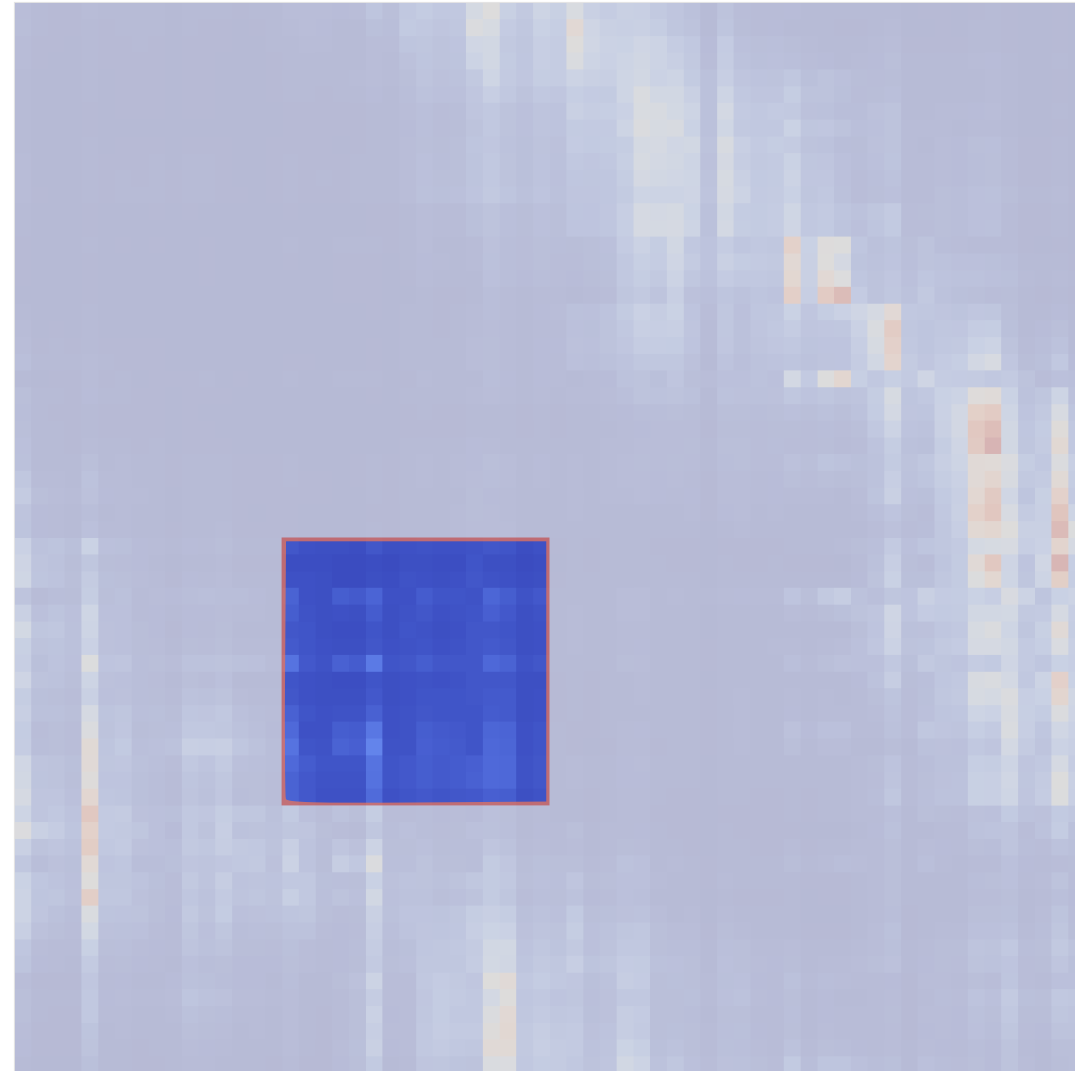


$$+ \dots$$

# Multi-Resolution Approximation

## Coefficient Approximation

$$\mu_5 = \text{Avg}(\quad)$$



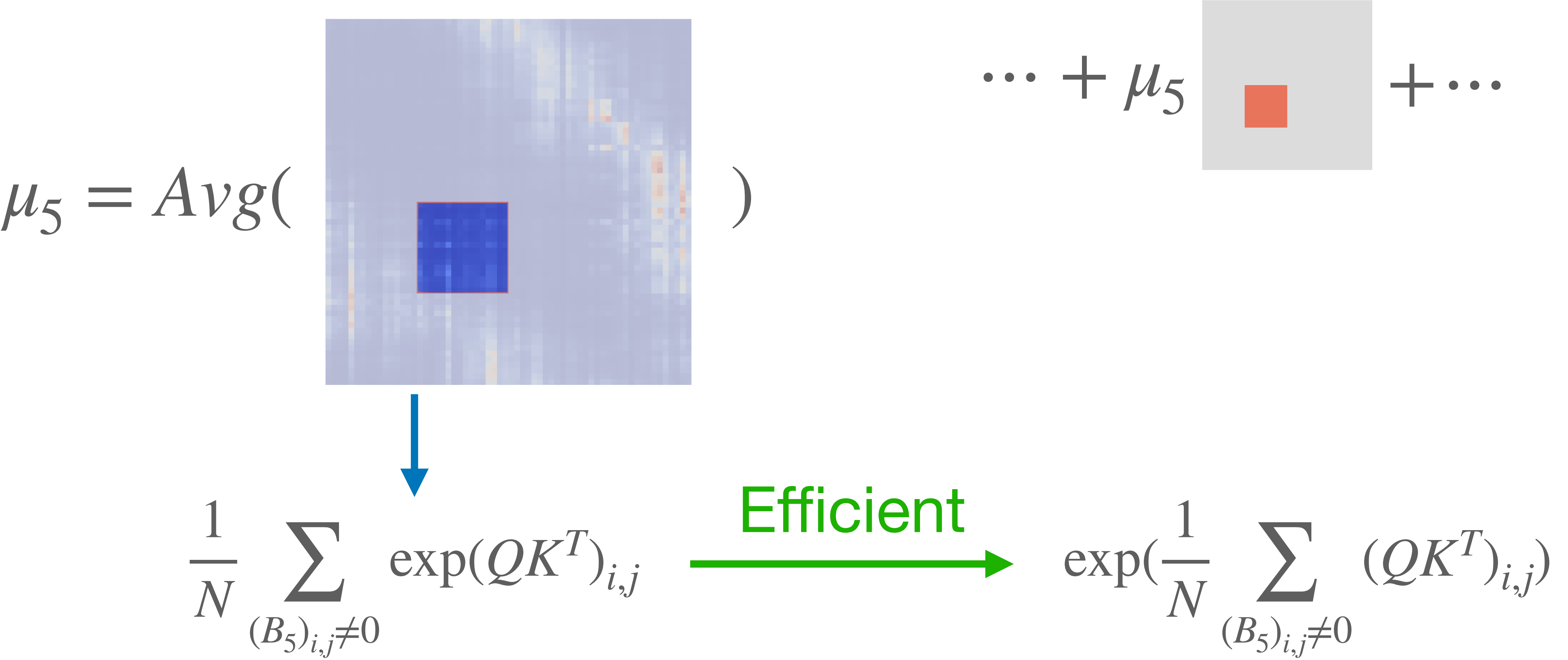
$$\frac{1}{N} \sum_{(B_5)_{i,j} \neq 0} \exp(QK^T)_{i,j}$$

$$\dots + \mu_5 \quad \text{[red square on gray background]} \quad + \dots$$



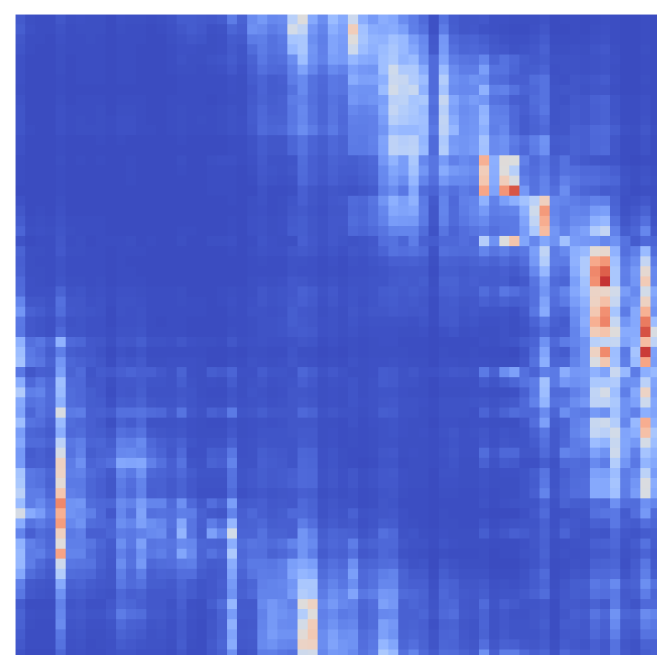
# Multi-Resolution Approximation

## Coefficient Approximation



# Multi-Resolution Approximation

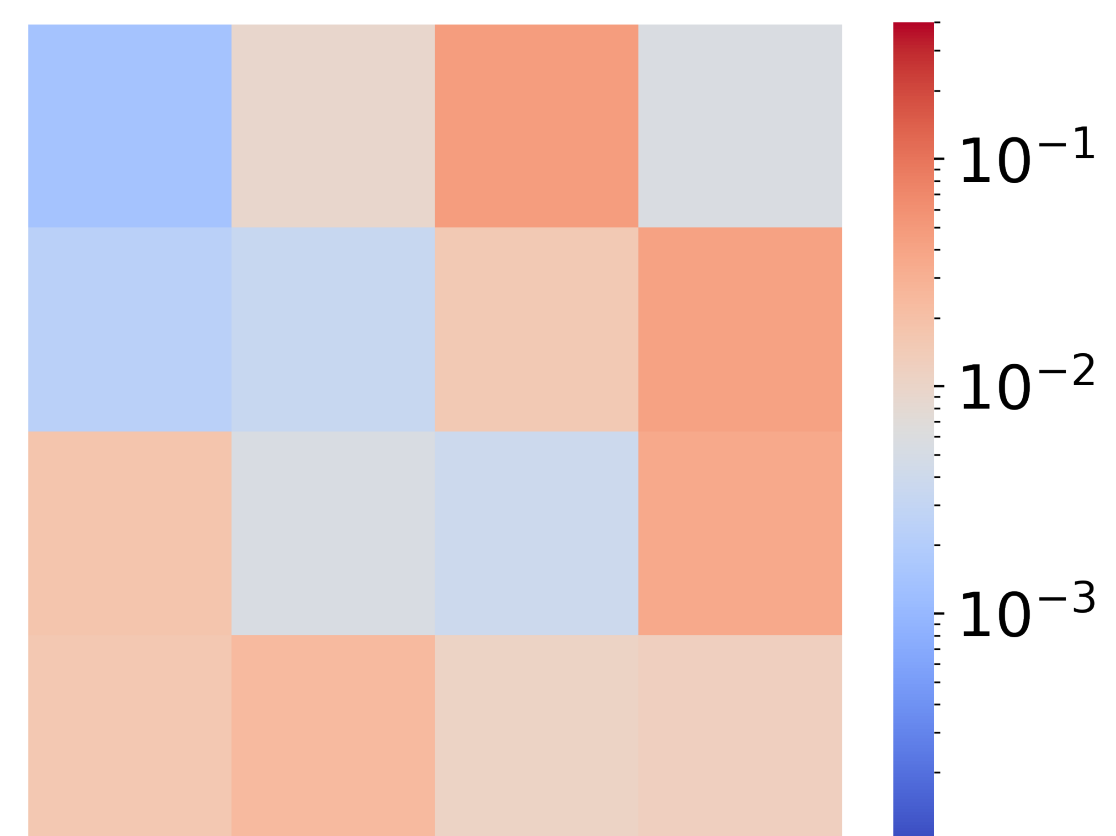
## Component Selection



$$\begin{aligned} &\approx \mu_1 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} + \mu_2 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} + \mu_3 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} + \mu_4 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} \\ &+ \mu_5 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} + \mu_6 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} + \mu_7 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} + \mu_8 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} \\ &+ \mu_9 \begin{array}{|c|} \hline \text{?} \\ \hline \end{array} + \dots \end{aligned}$$

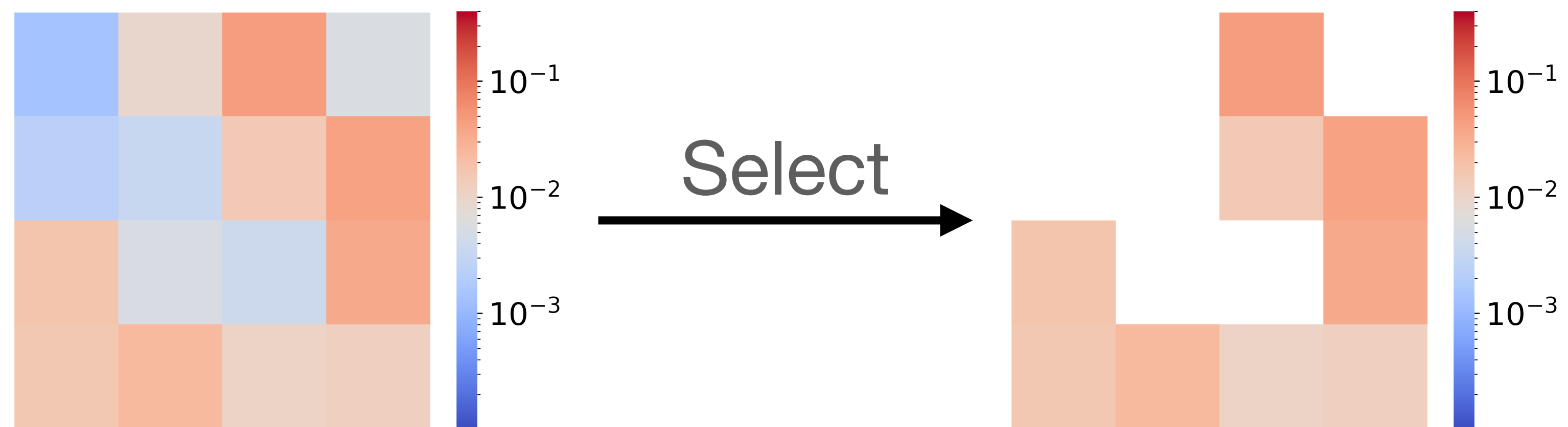
# Multi-Resolution Approximation

## Illustration



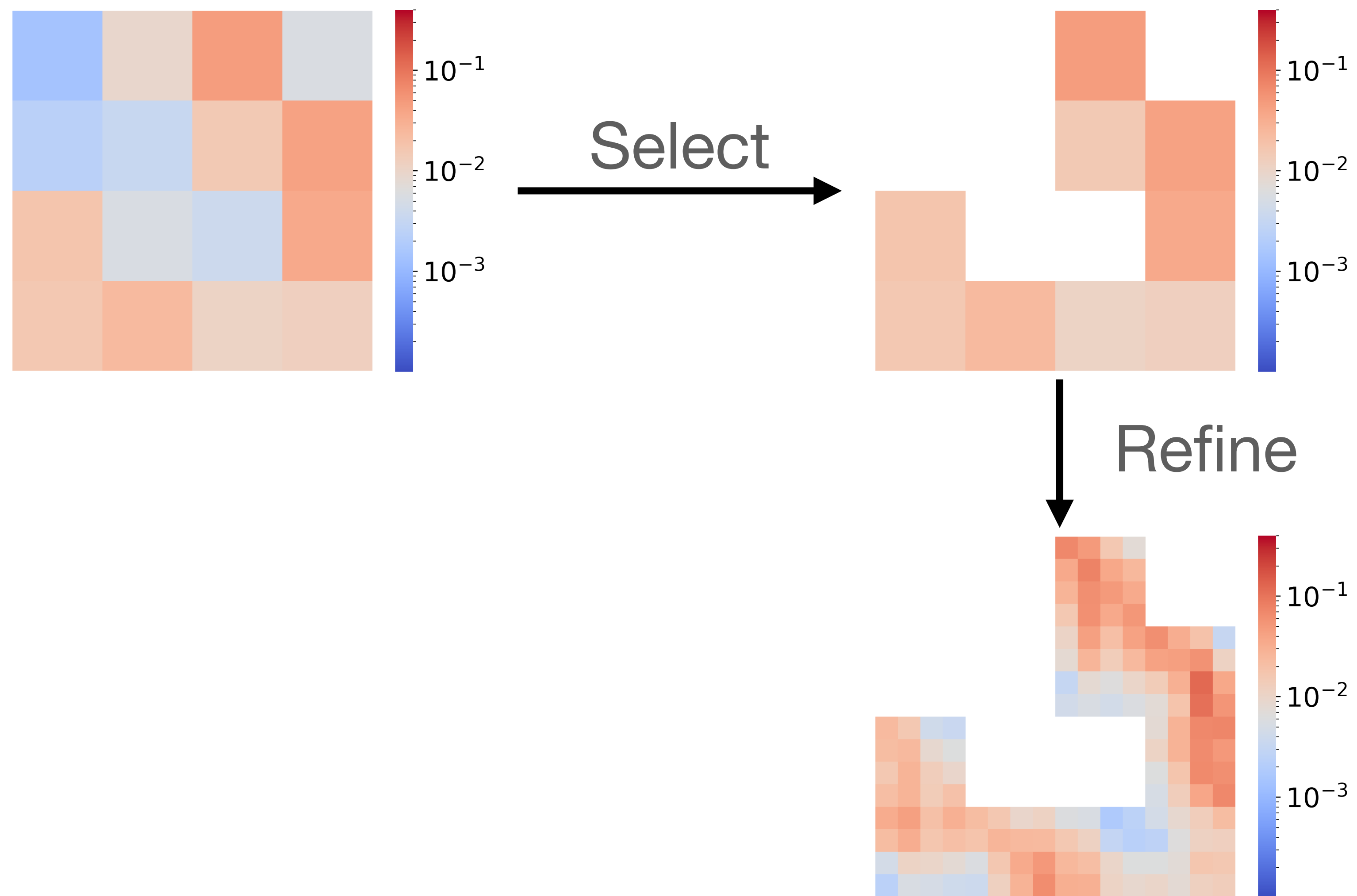
# Multi-Resolution Approximation

## Illustration



# Multi-Resolution Approximation

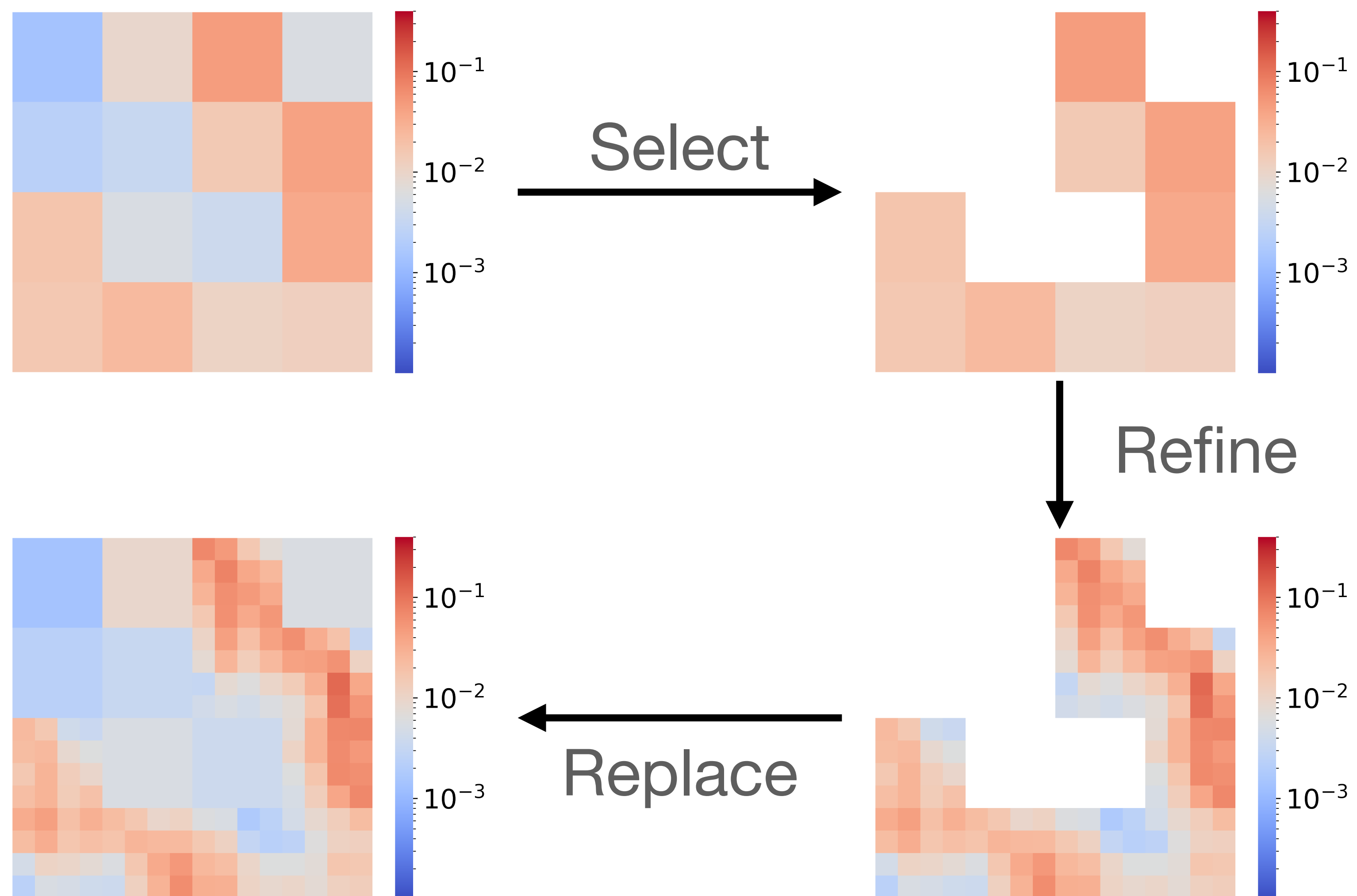
## Illustration





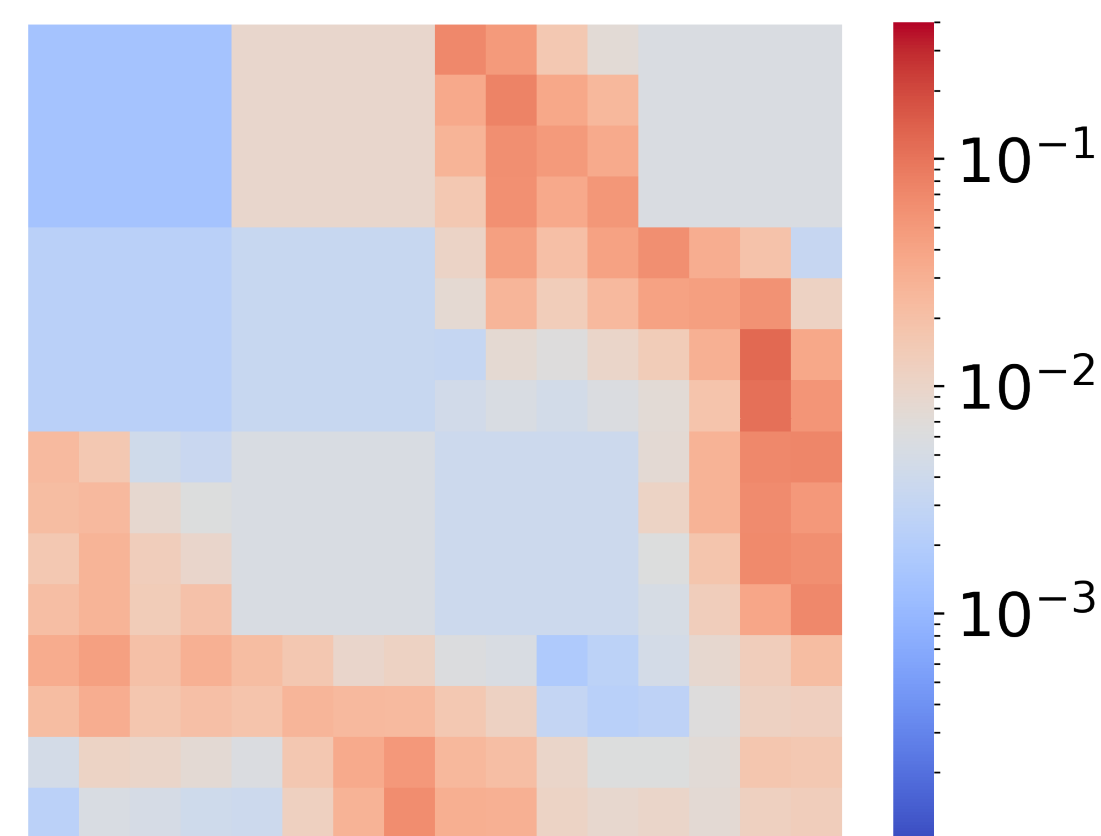
# Multi-Resolution Approximation

## Illustration



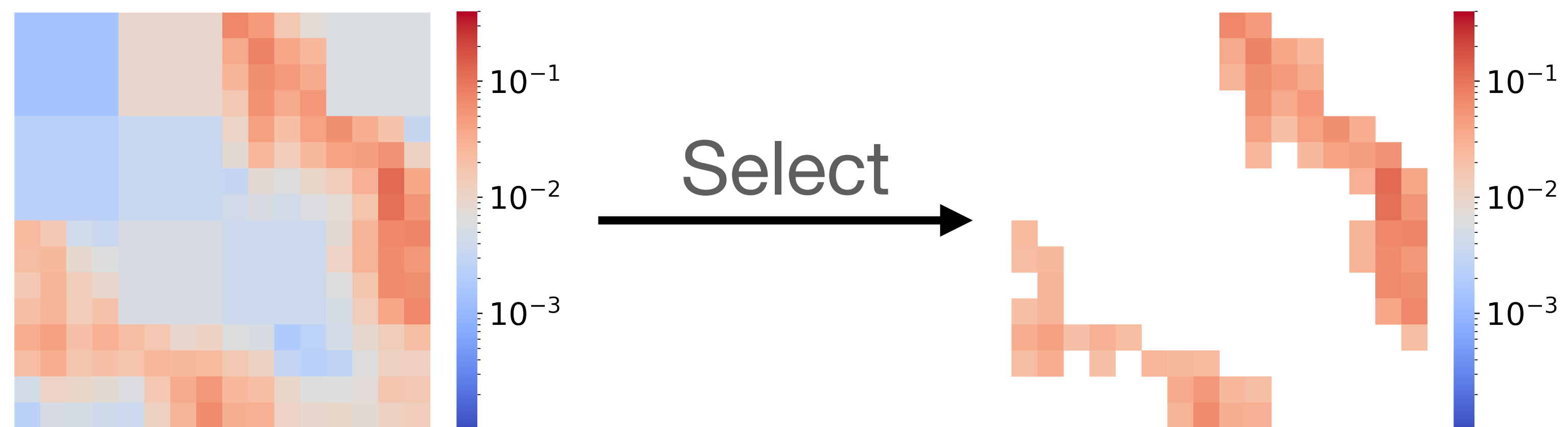
# Multi-Resolution Approximation

## Illustration



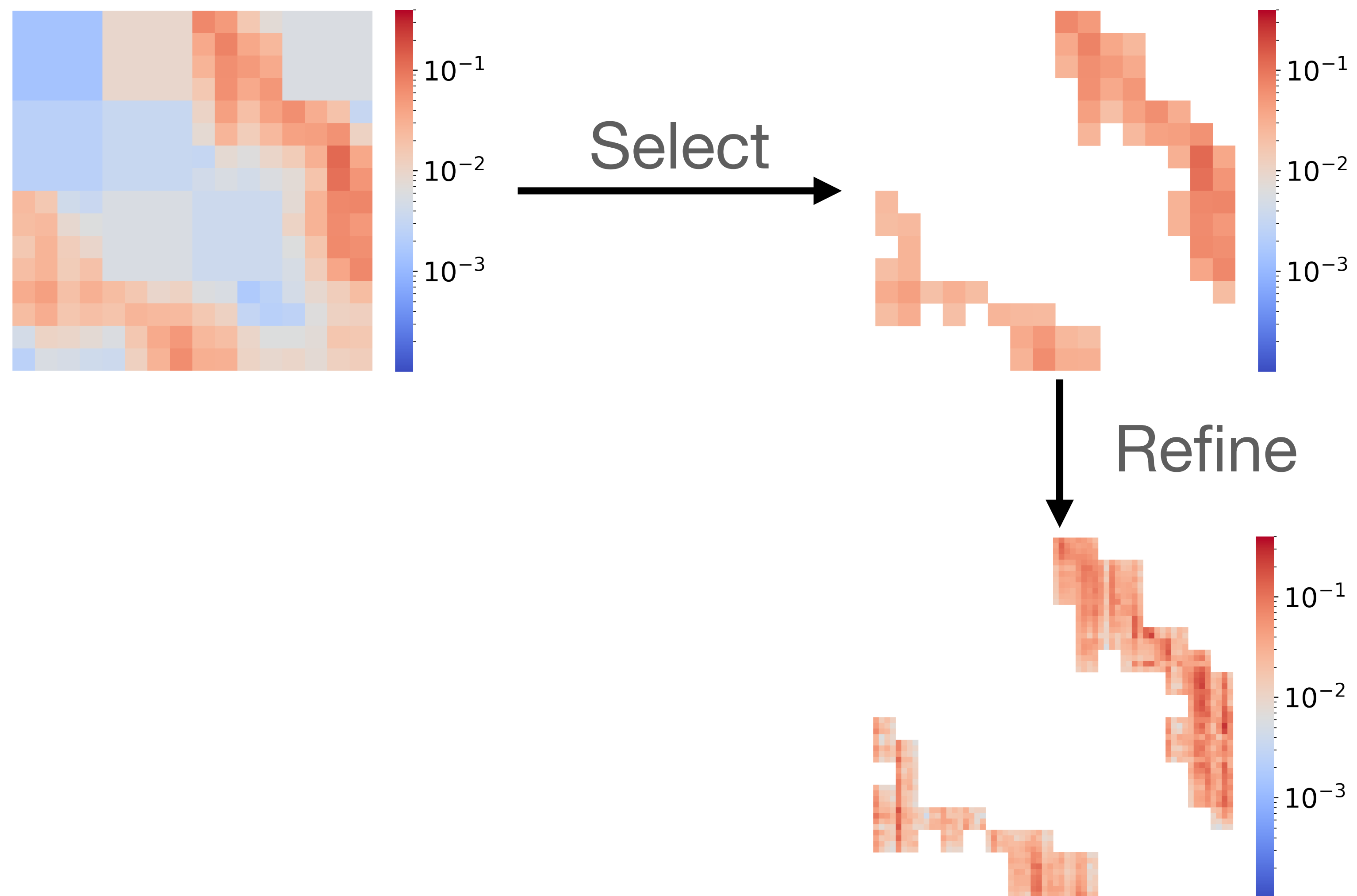
# Multi-Resolution Approximation

## Illustration



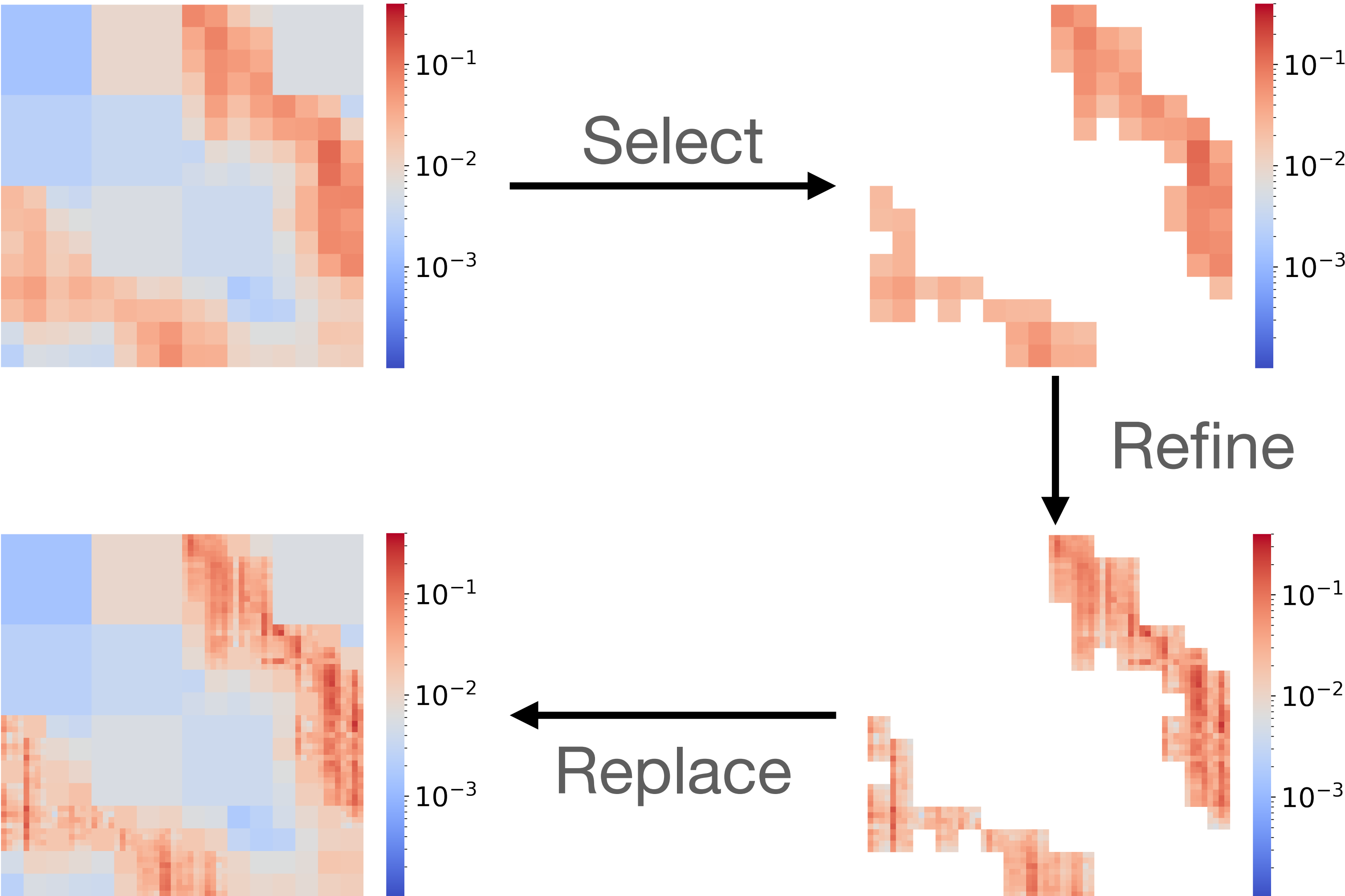
# Multi-Resolution Approximation

## Illustration



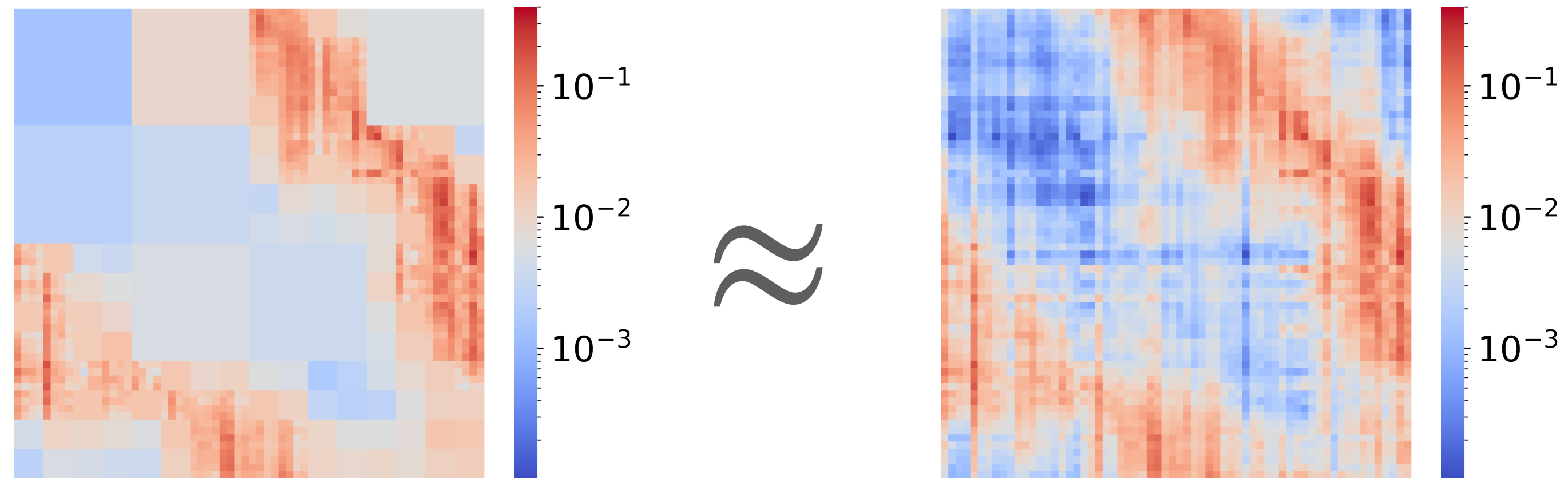
# Multi-Resolution Approximation

## Illustration



# Multi-Resolution Approximation

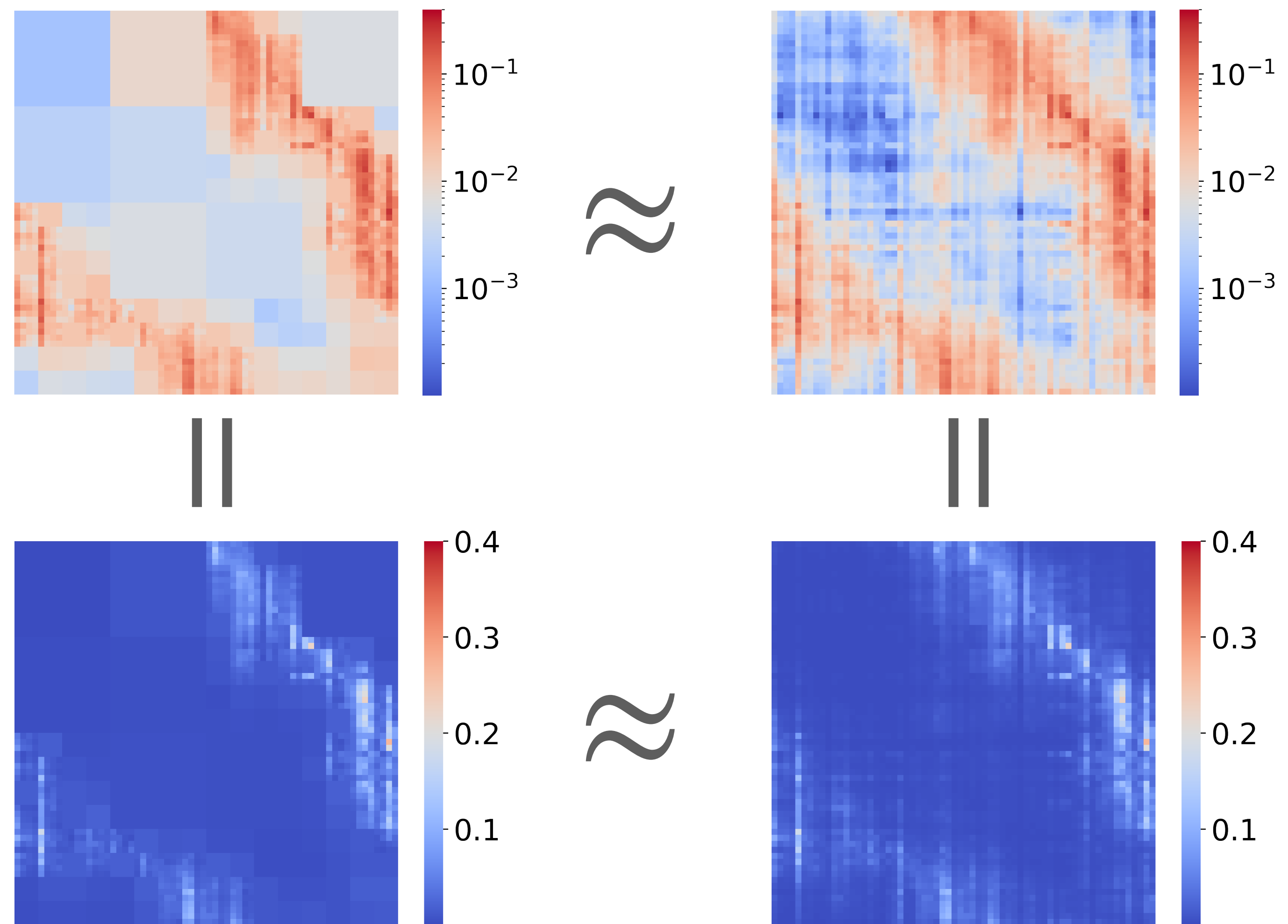
## Approximation Result





# Multi-Resolution Approximation

## Approximation Result

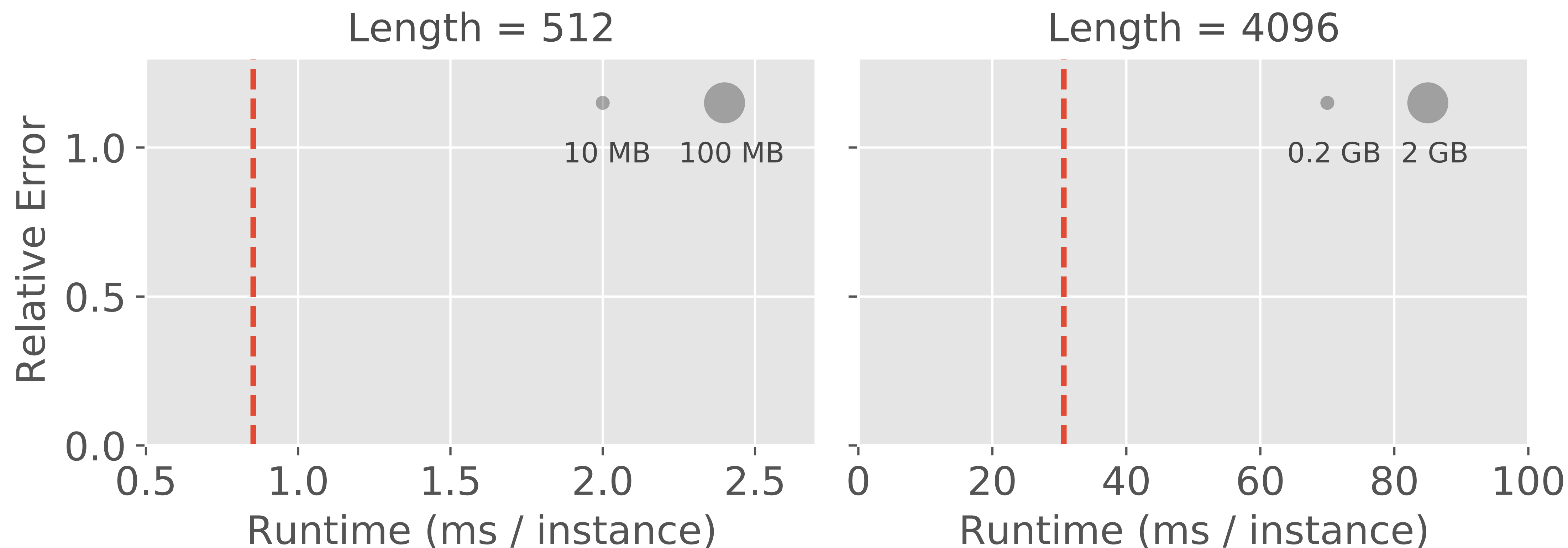
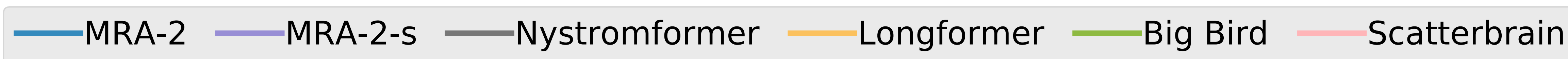


# **Evaluations**

## **Efficiency versus Approximation**

# Evaluations

## Efficiency v.s. Approximation

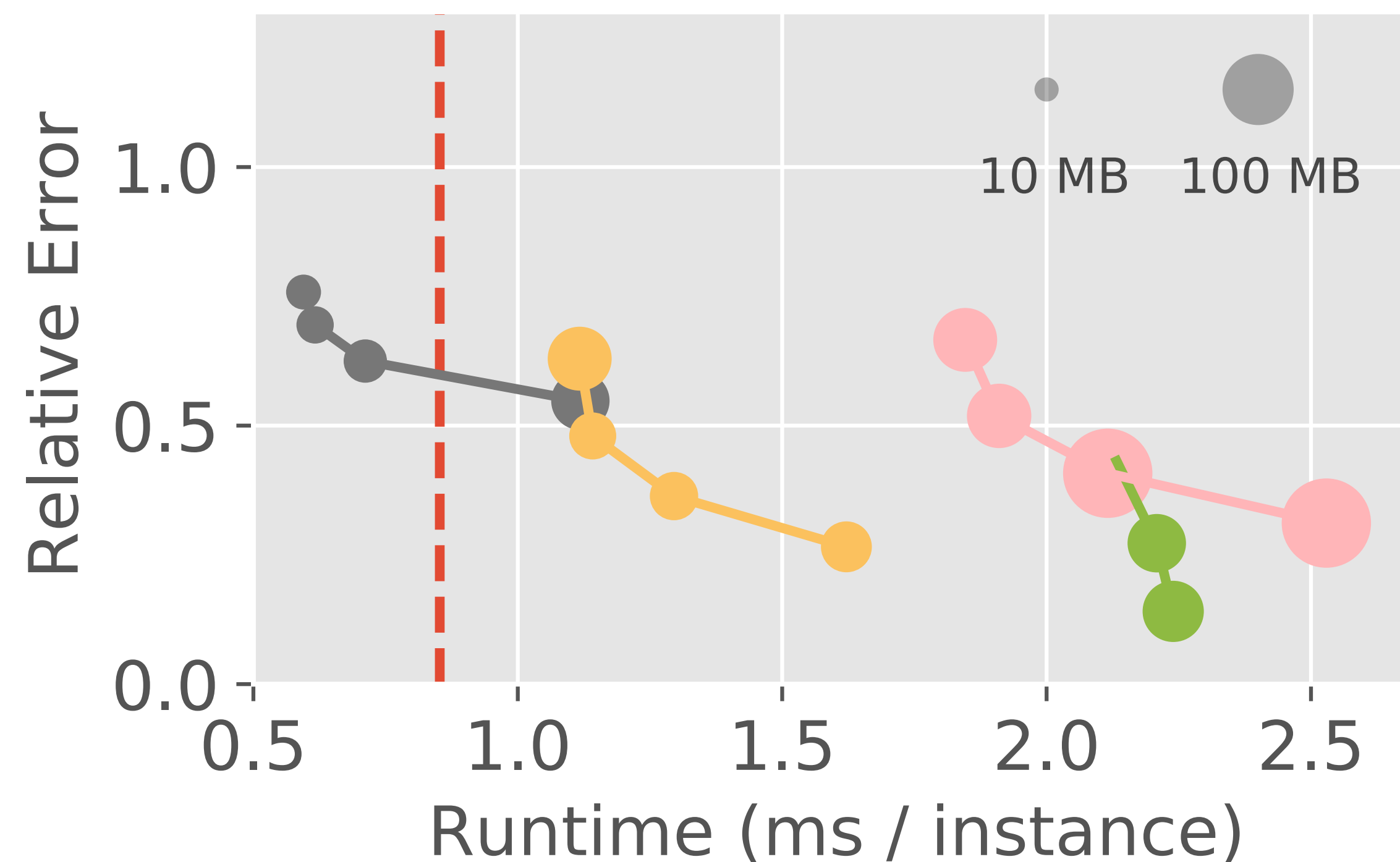


# Evaluations

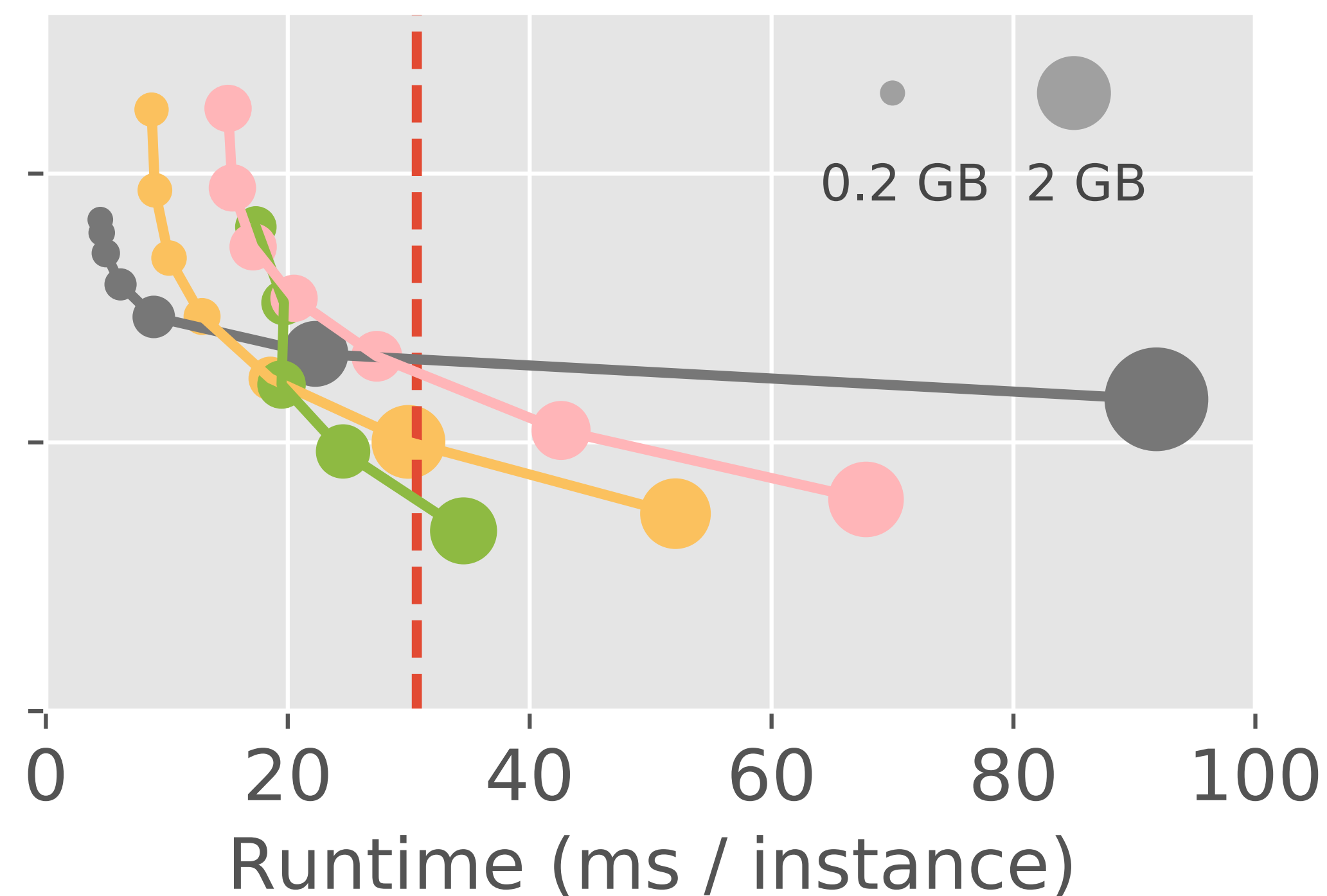
## Efficiency v.s. Approximation

MRA-2 MRA-2-s Nystromformer Longformer Big Bird Scatterbrain

Length = 512



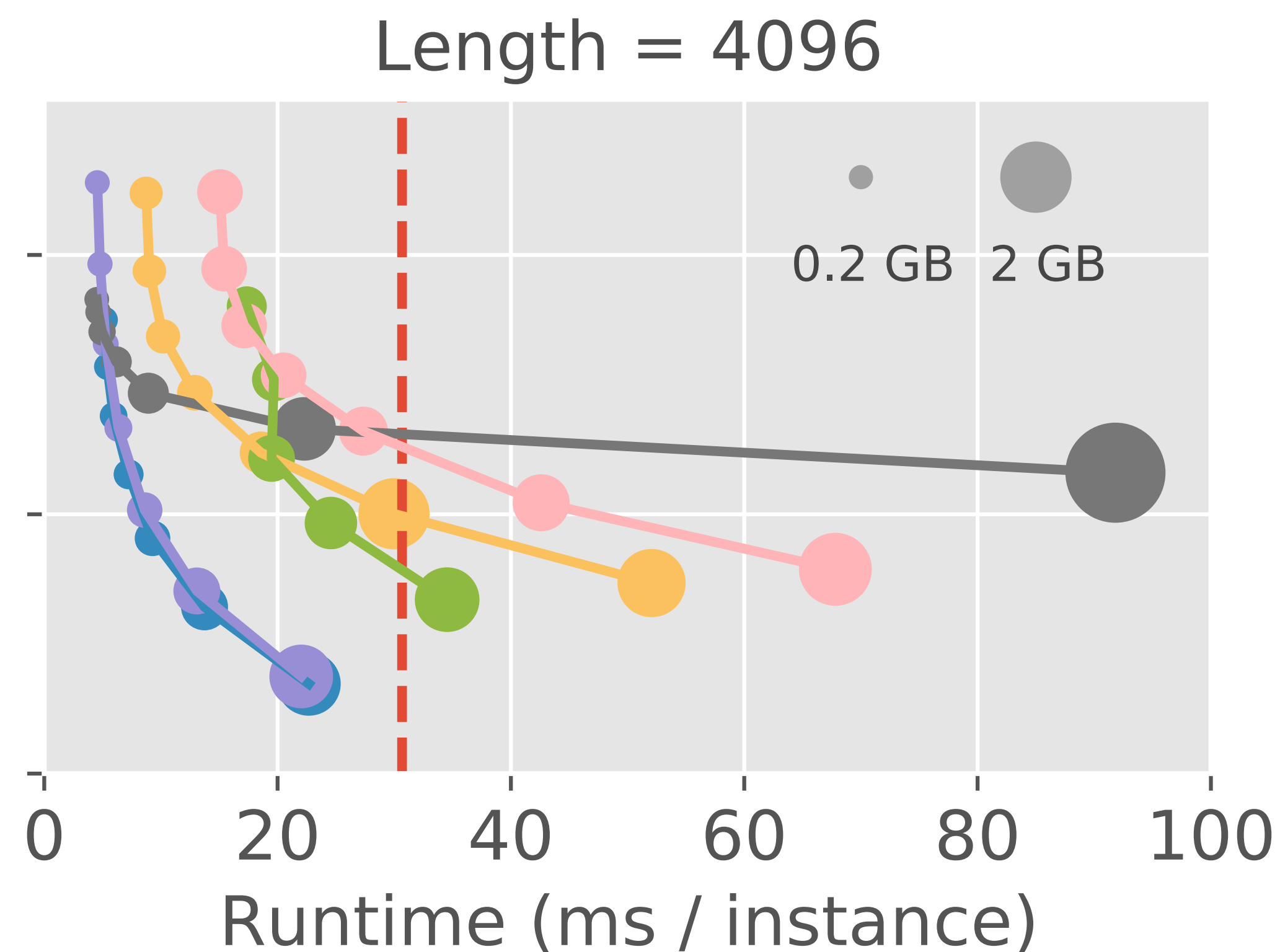
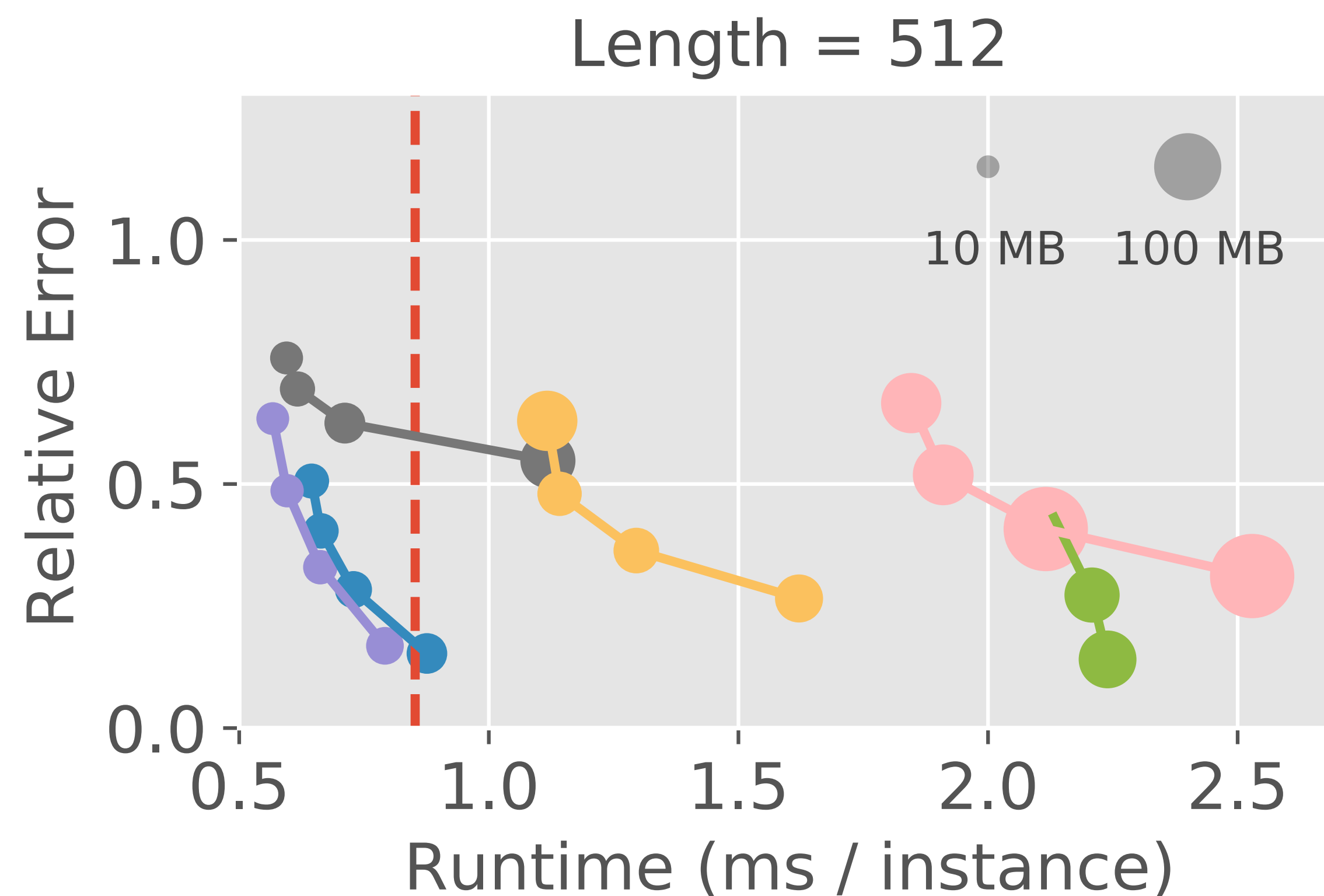
Length = 4096



# Evaluations

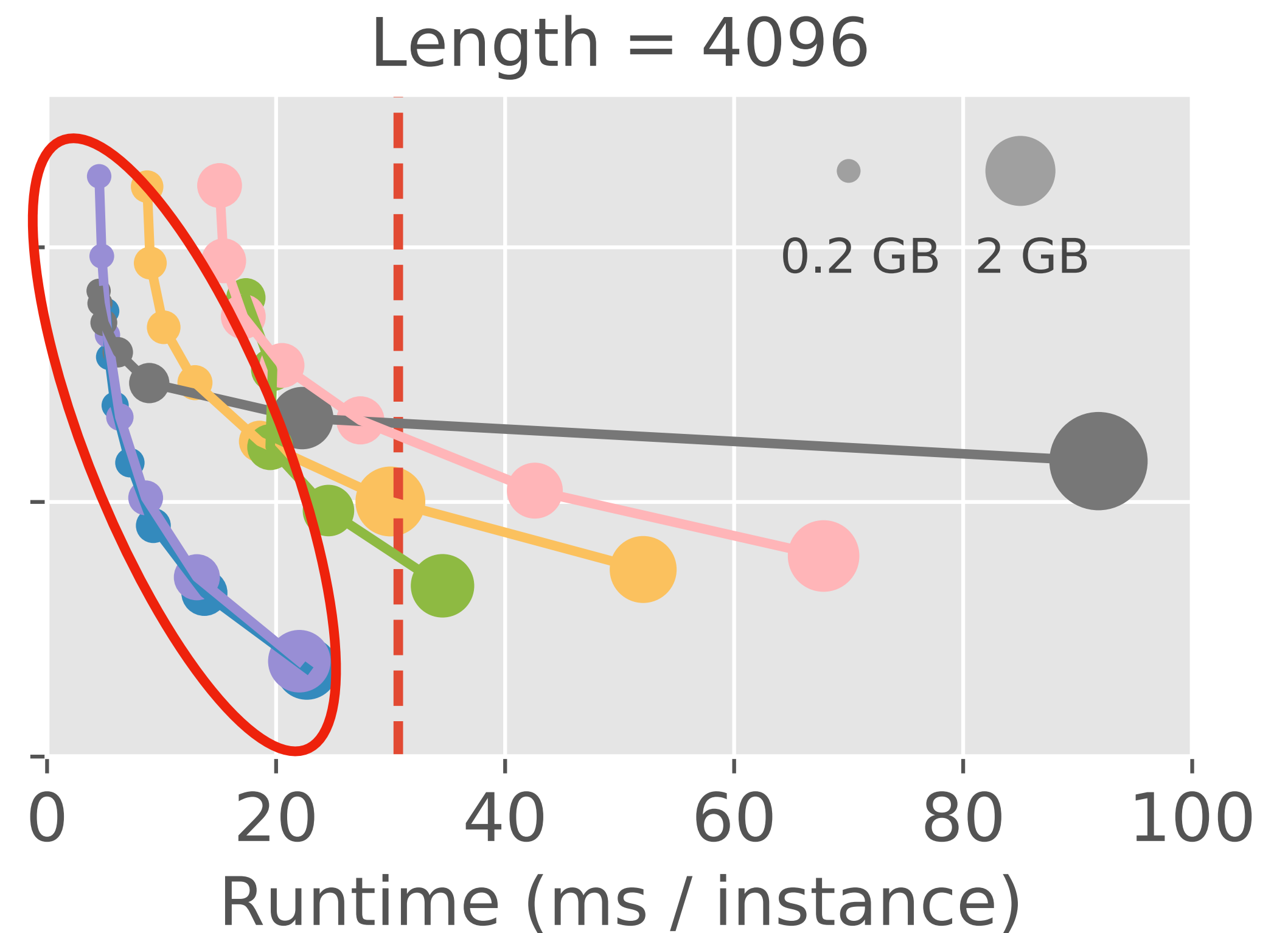
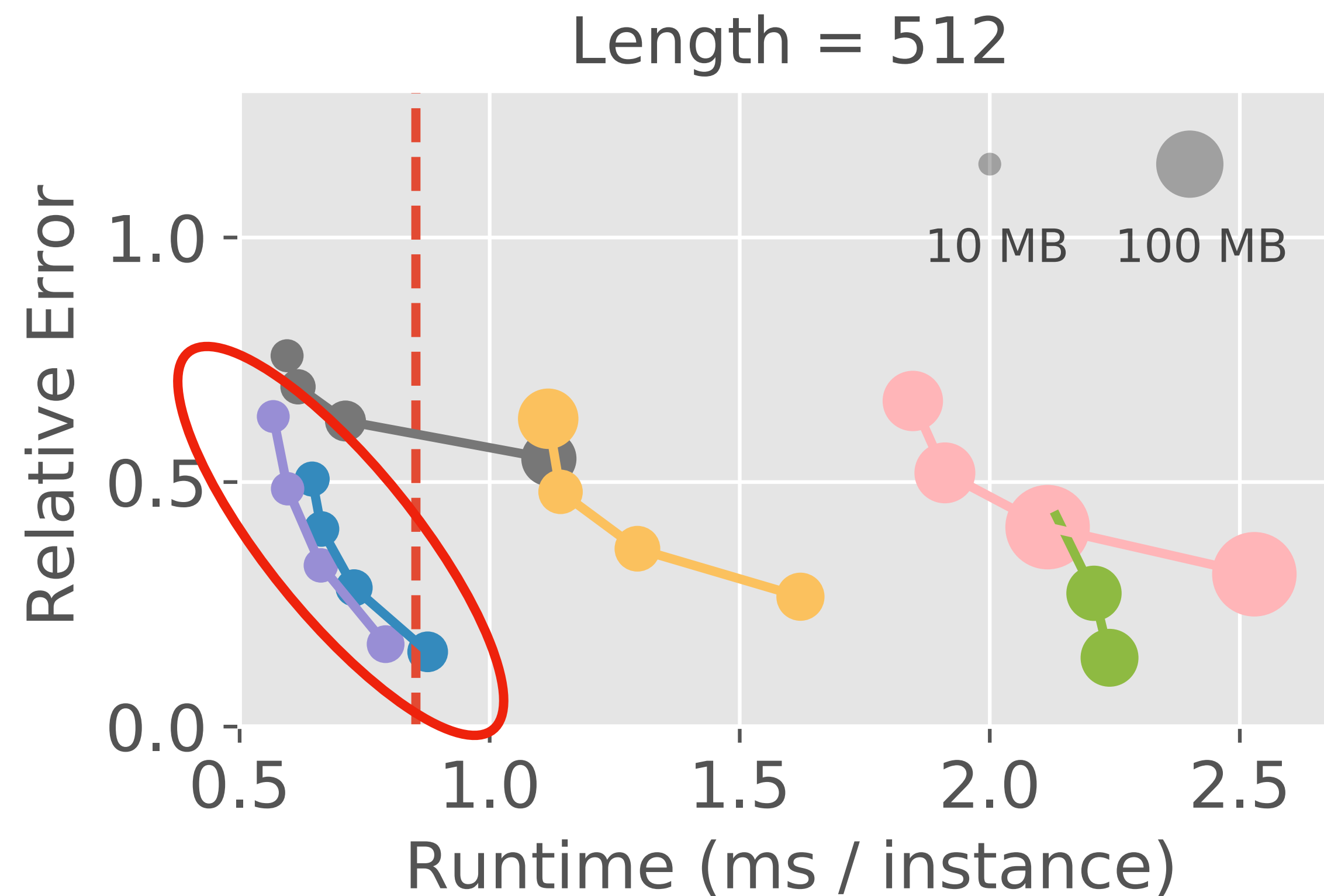
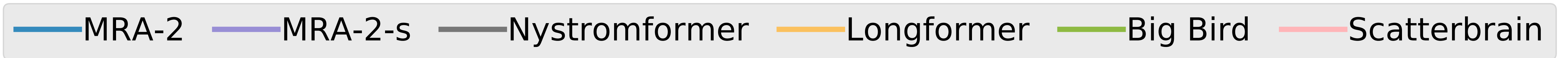
## Efficiency v.s. Approximation

MRA-2 MRA-2-s Nystromformer Longformer Big Bird Scatterbrain



# Evaluations

## Efficiency v.s. Approximation





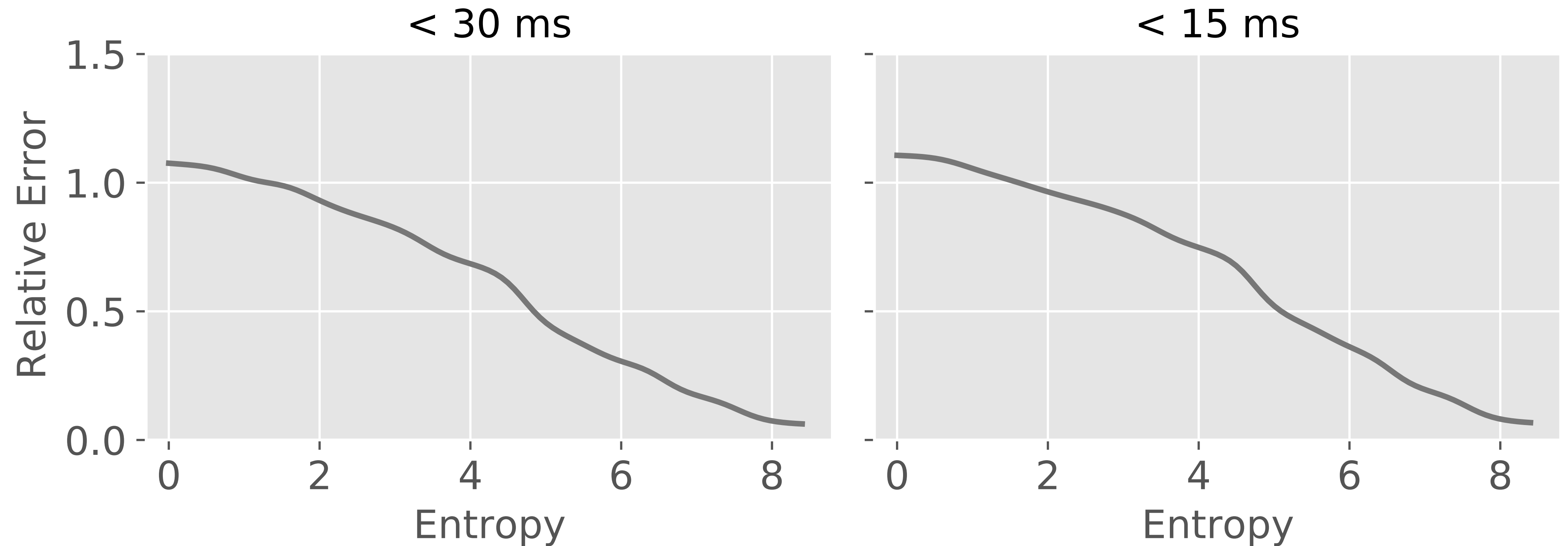
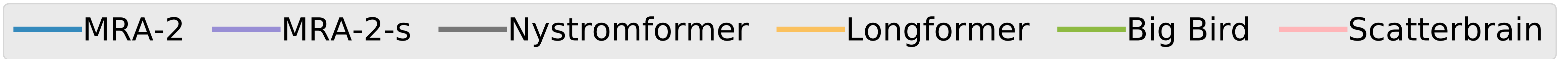
# **Evaluations**

## **Entropy versus Approximation**



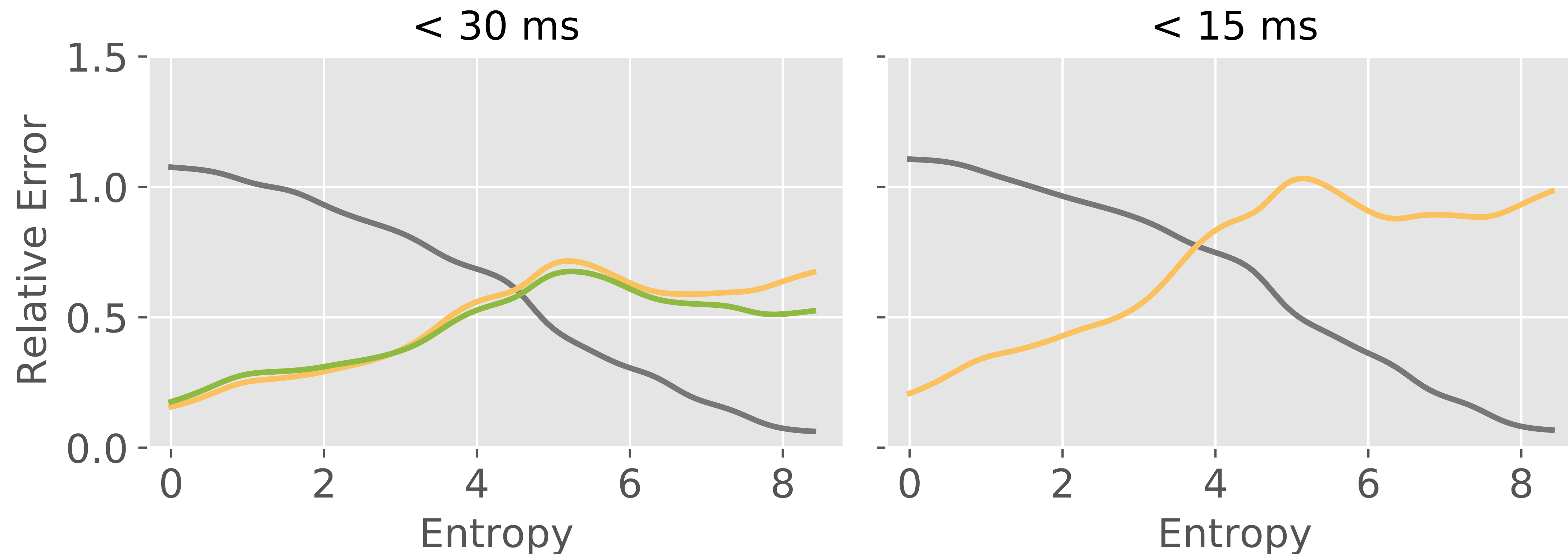
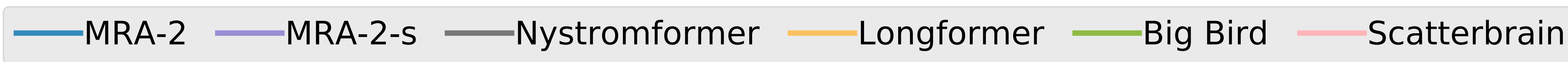
# Evaluations

## Entropy v.s. Approximation



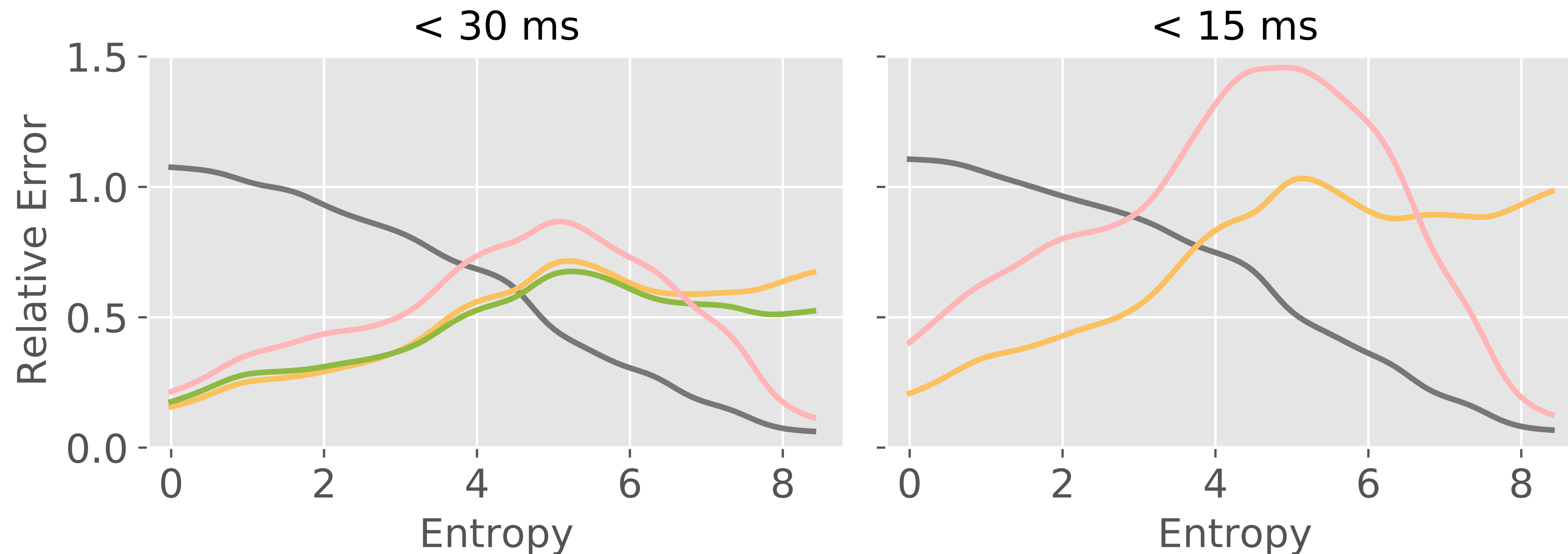
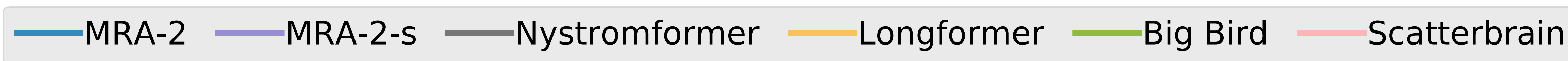
# Evaluations

## Entropy v.s. Approximation



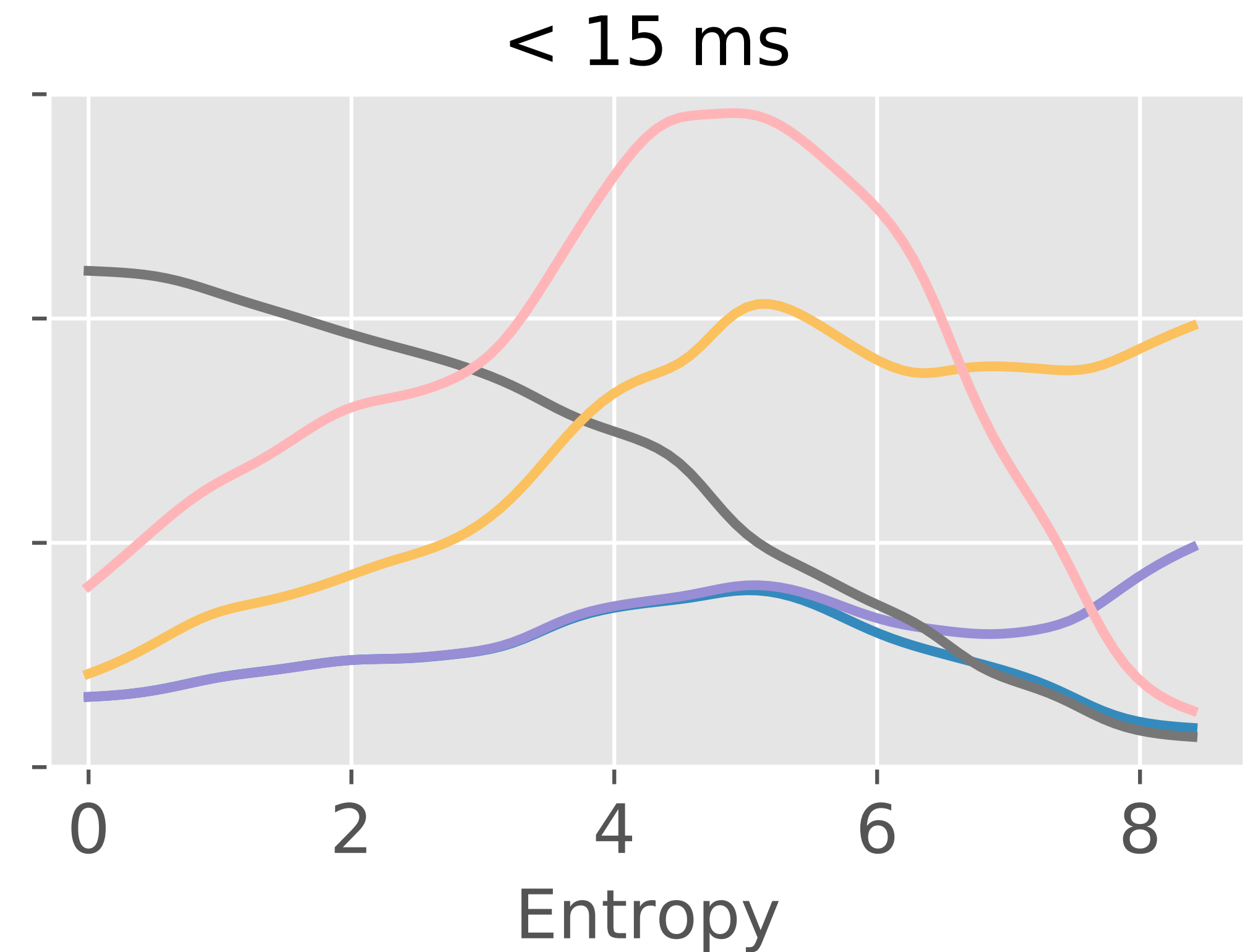
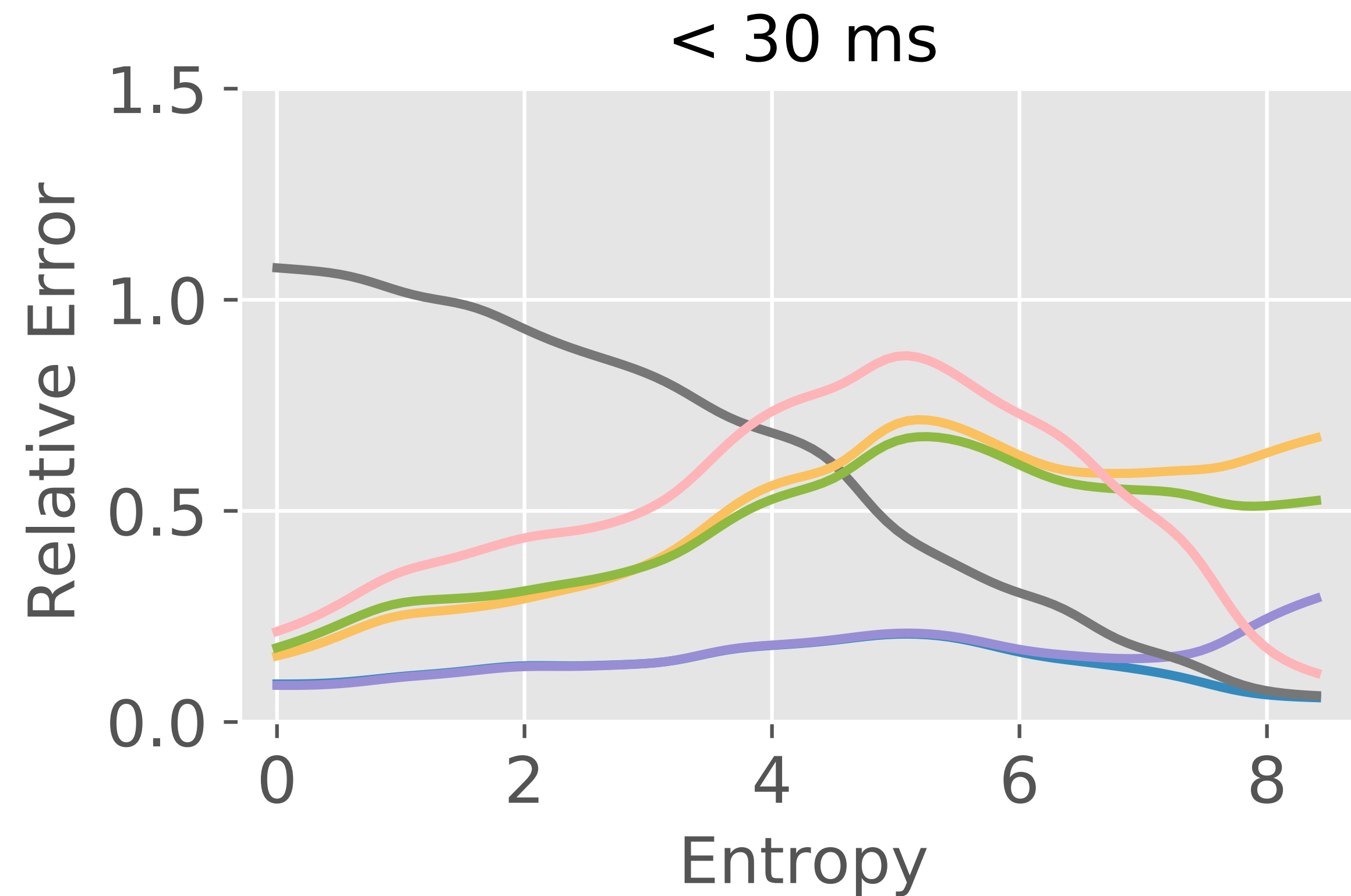
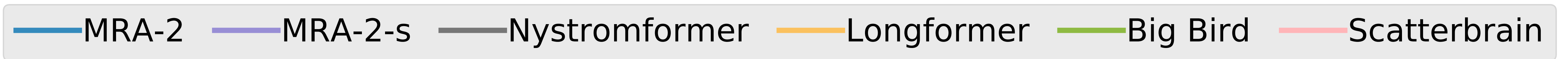
# Evaluations

## Entropy v.s. Approximation



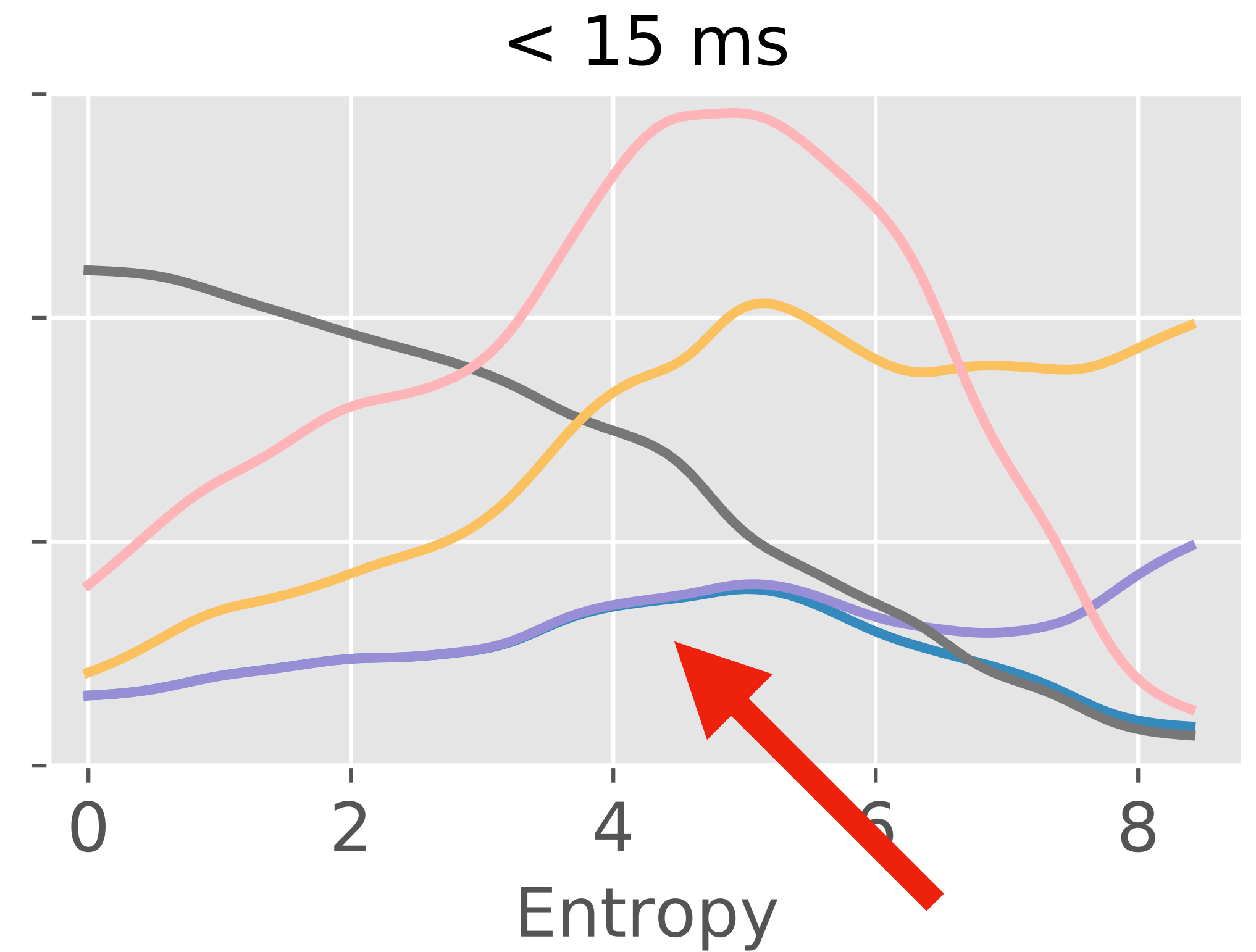
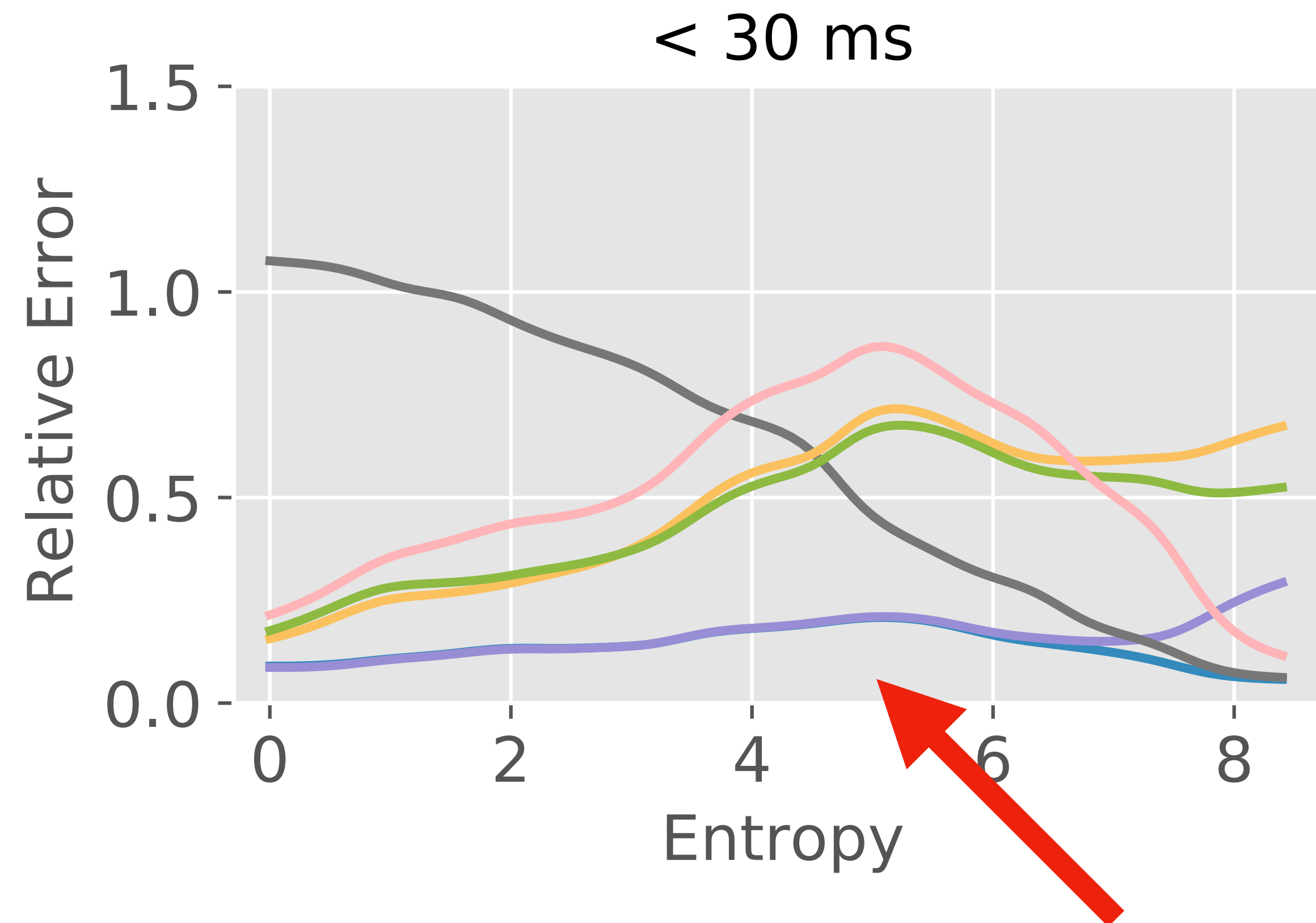
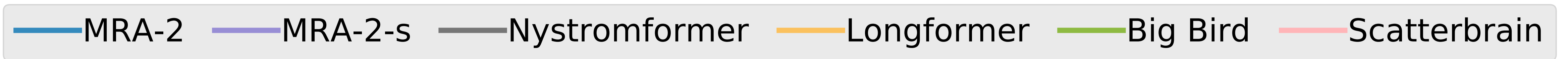
# Evaluations

## Entropy v.s. Approximation



# Evaluations

## Entropy v.s. Approximation



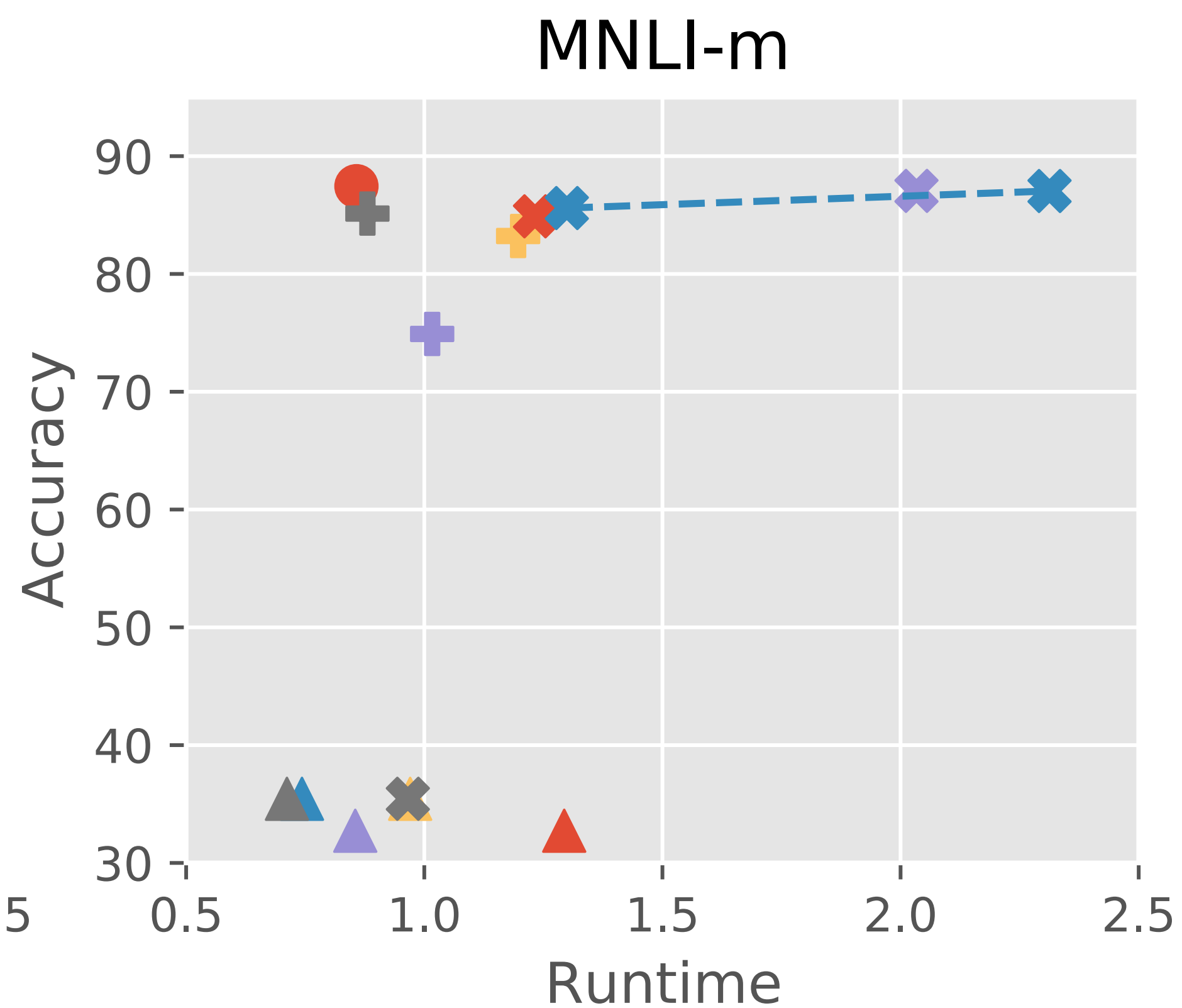
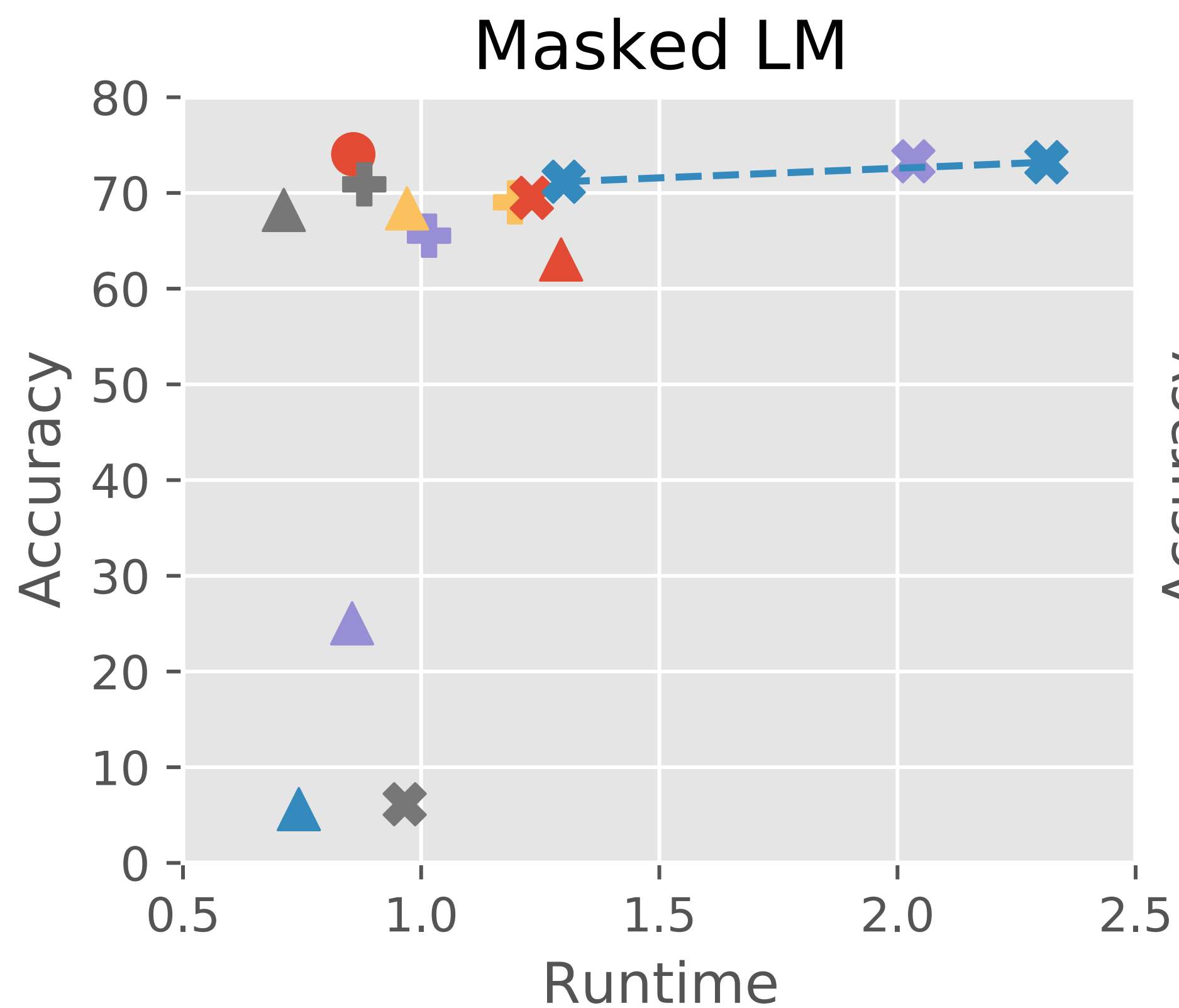


# **Evaluations**

## **RoBERTa-base-512**

# Evaluations

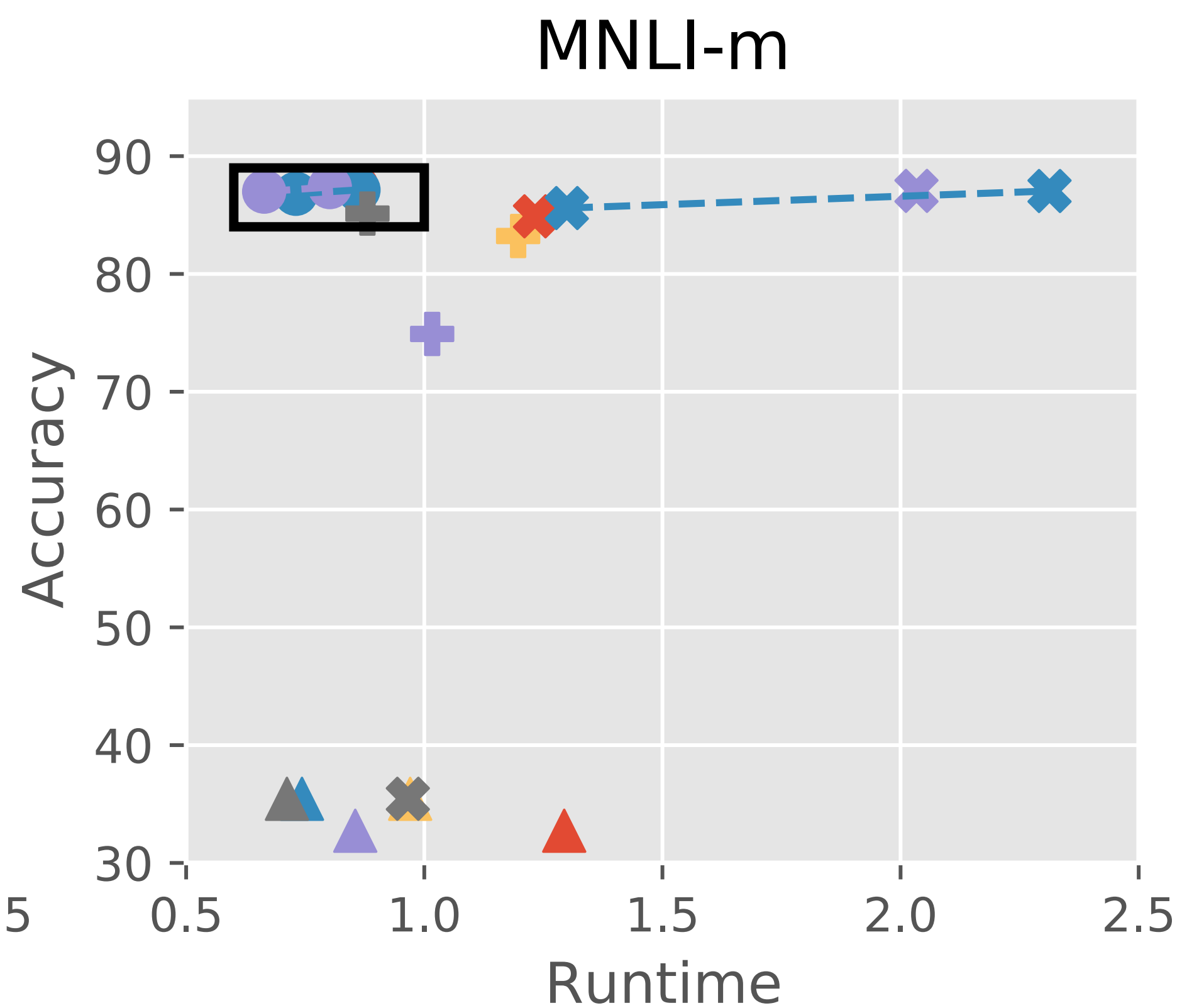
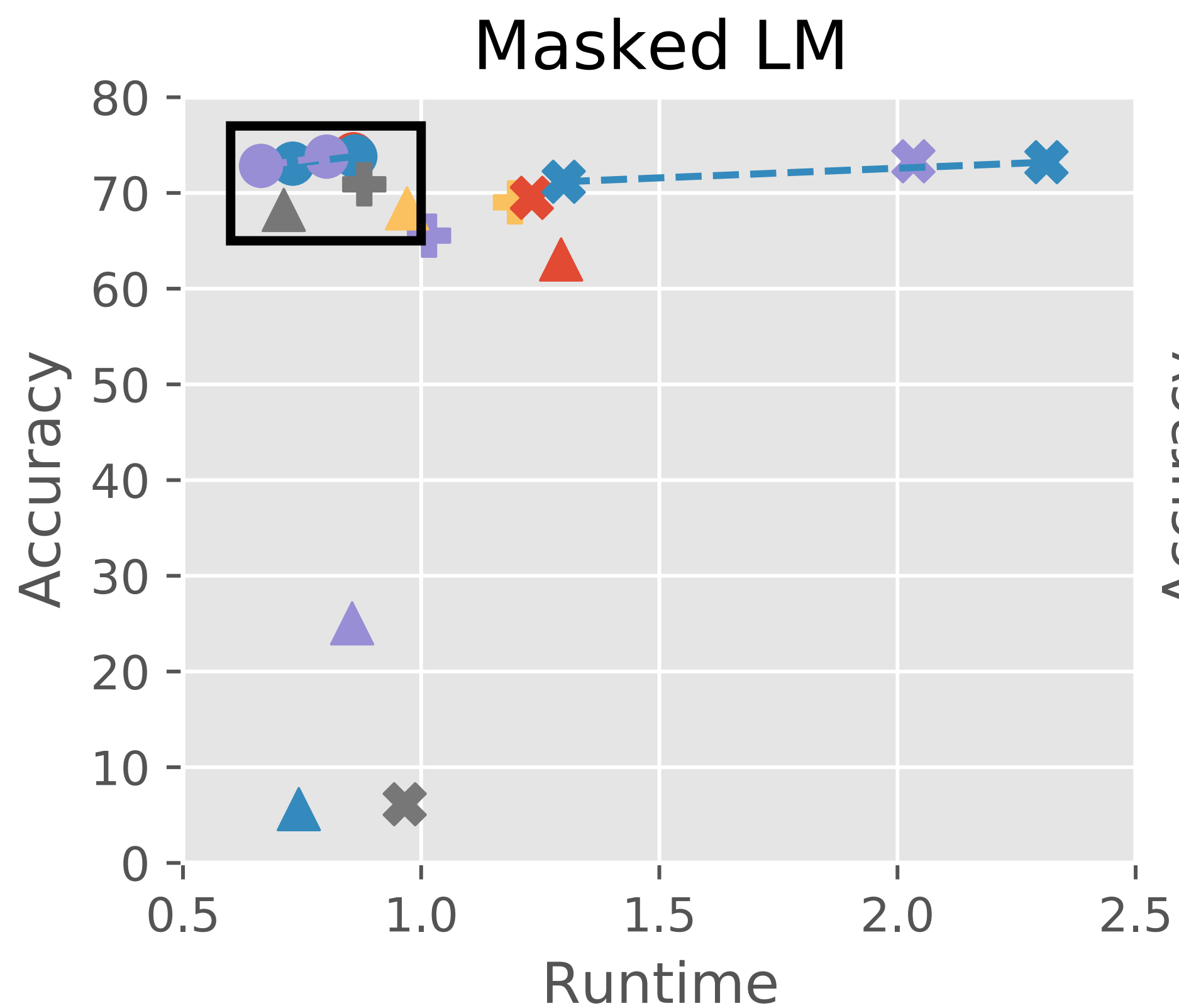
## RoBERTa-base-512



- Transformer
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

# Evaluations

## RoBERTa-base-512

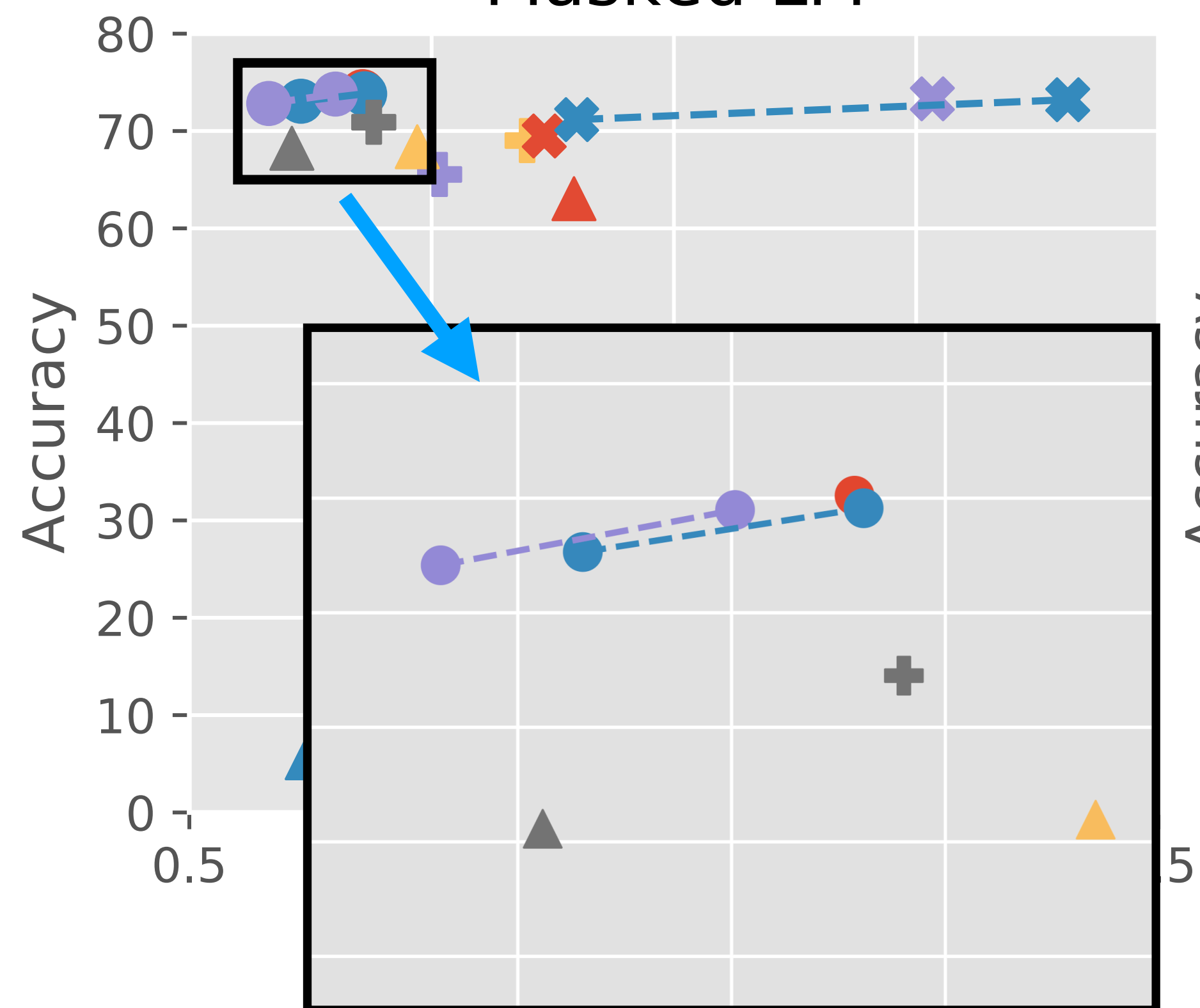


- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

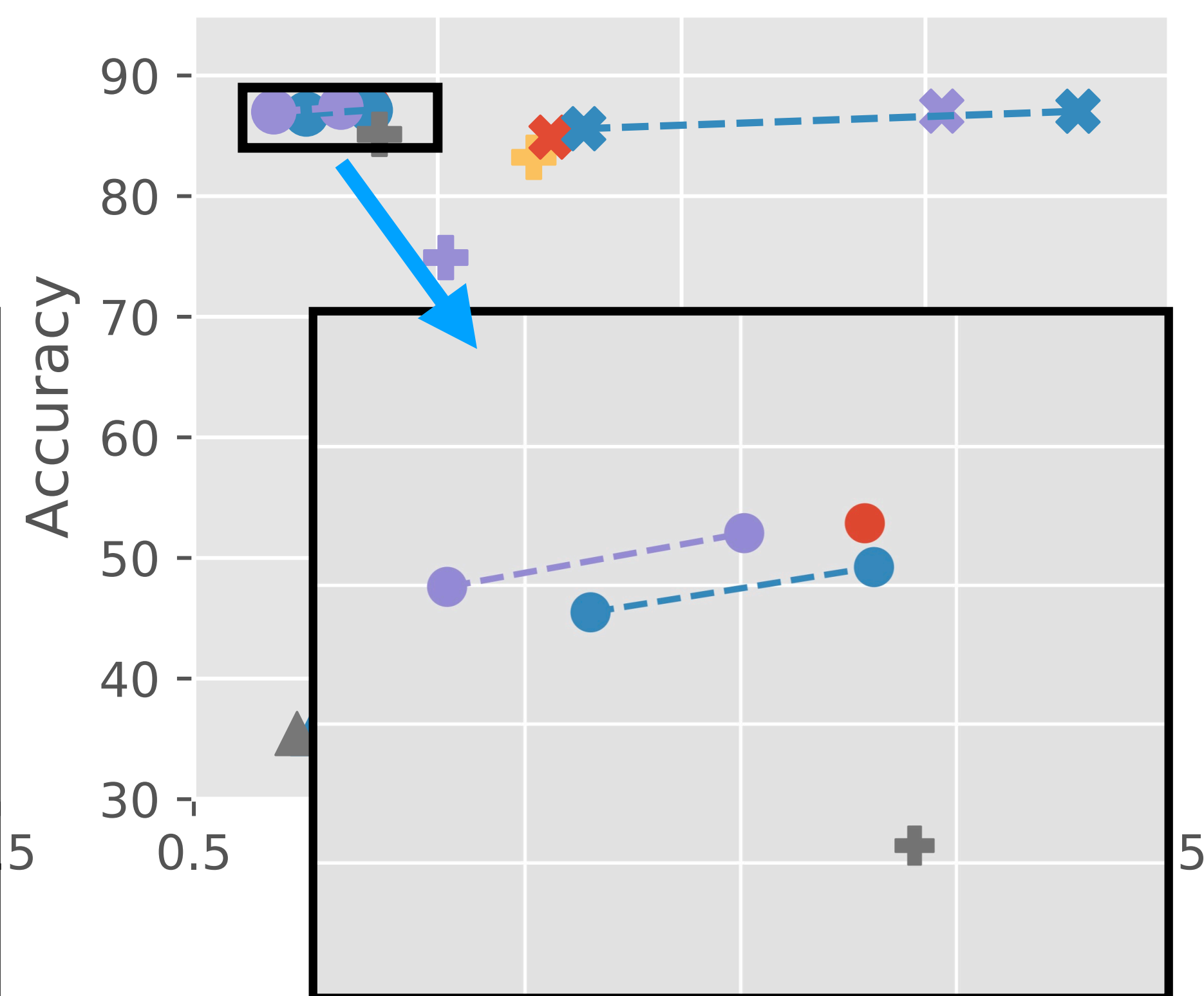
# Evaluations

## RoBERTa-base-512

Masked LM



MNLI-m

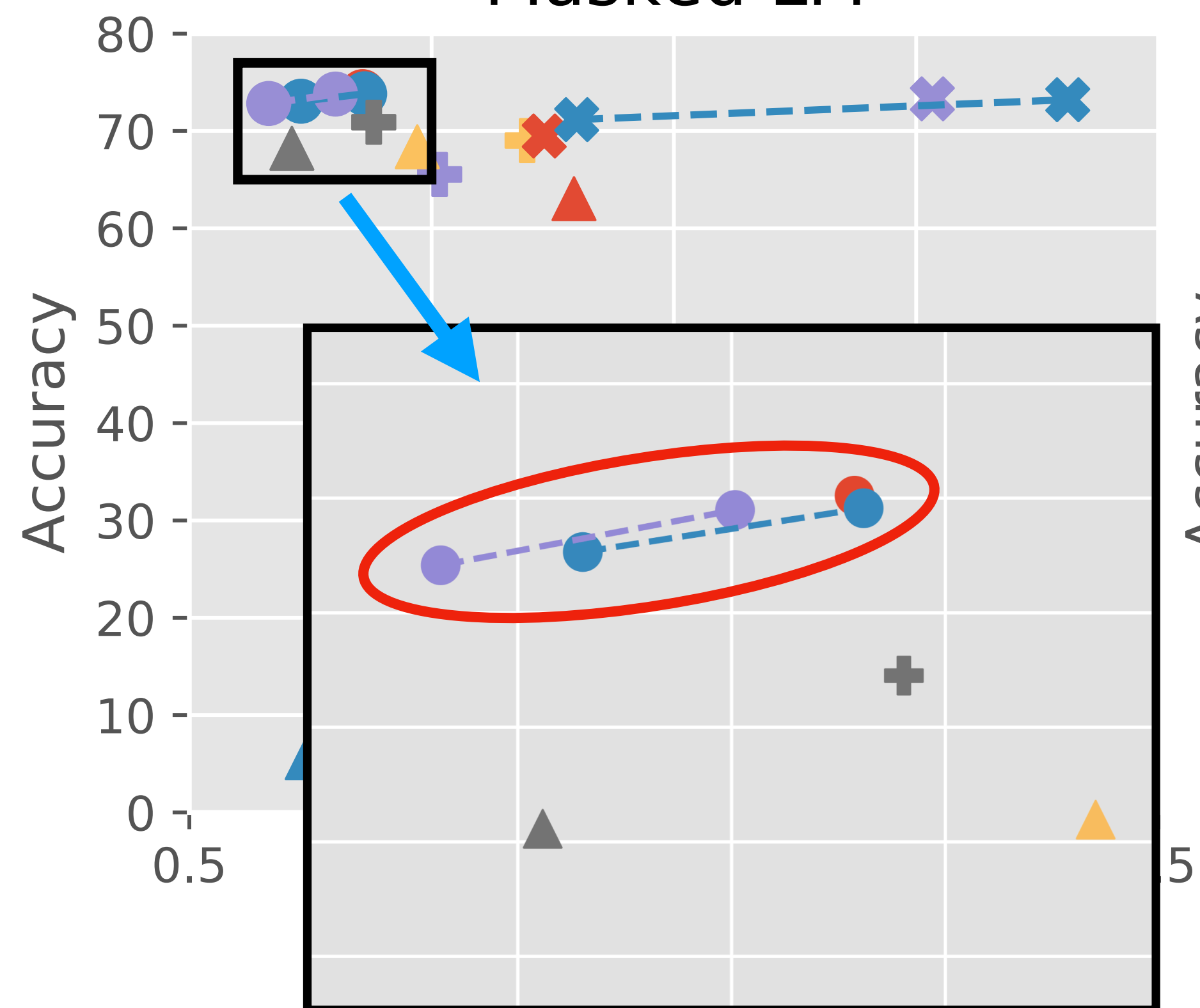


- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

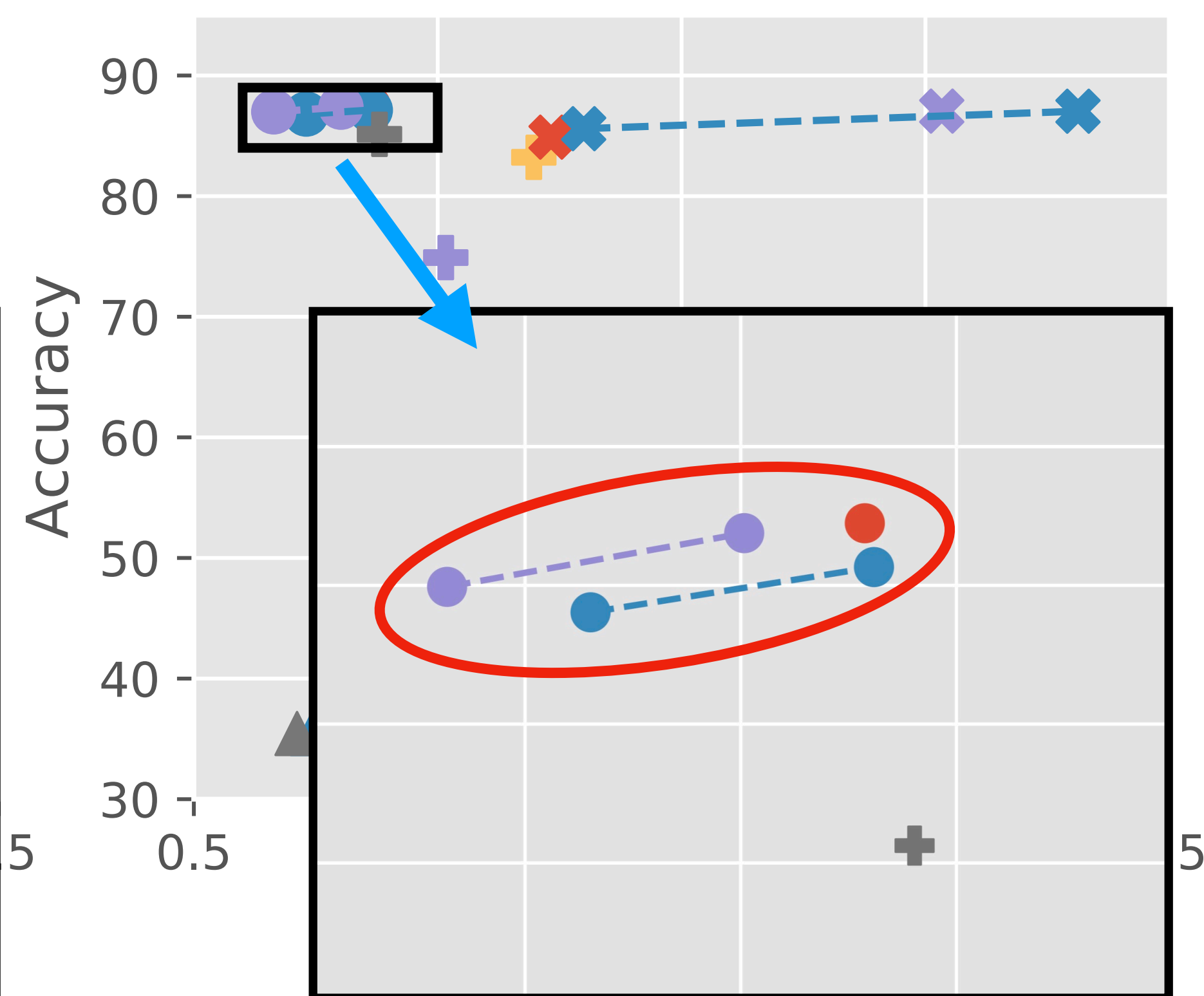
# Evaluations

## RoBERTa-base-512

Masked LM



MNLI-m



- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

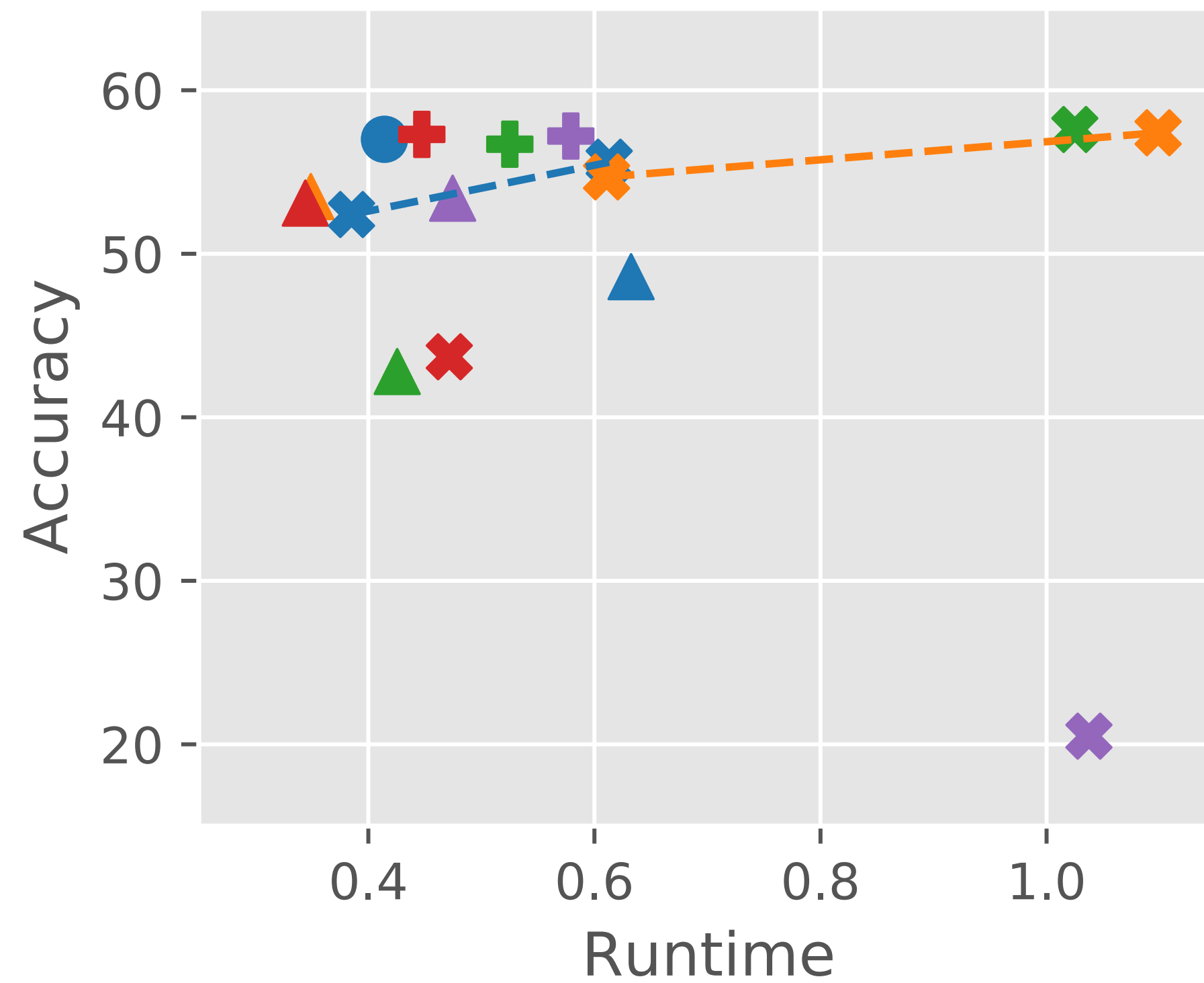
# **Evaluations**

## **RoBERTa-small-512**

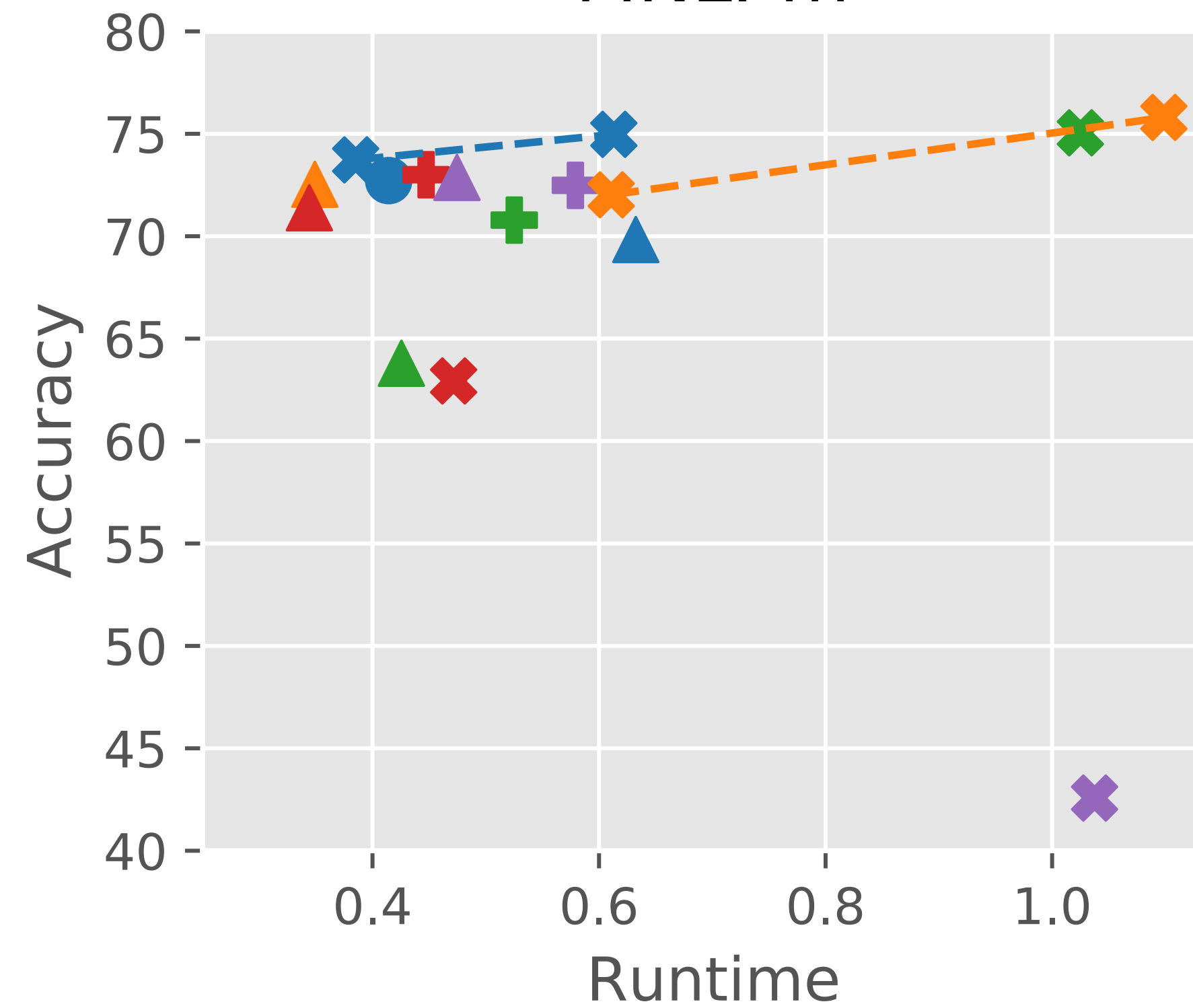
# Evaluations

## RoBERTa-small-512

Masked LM



MNLI-m

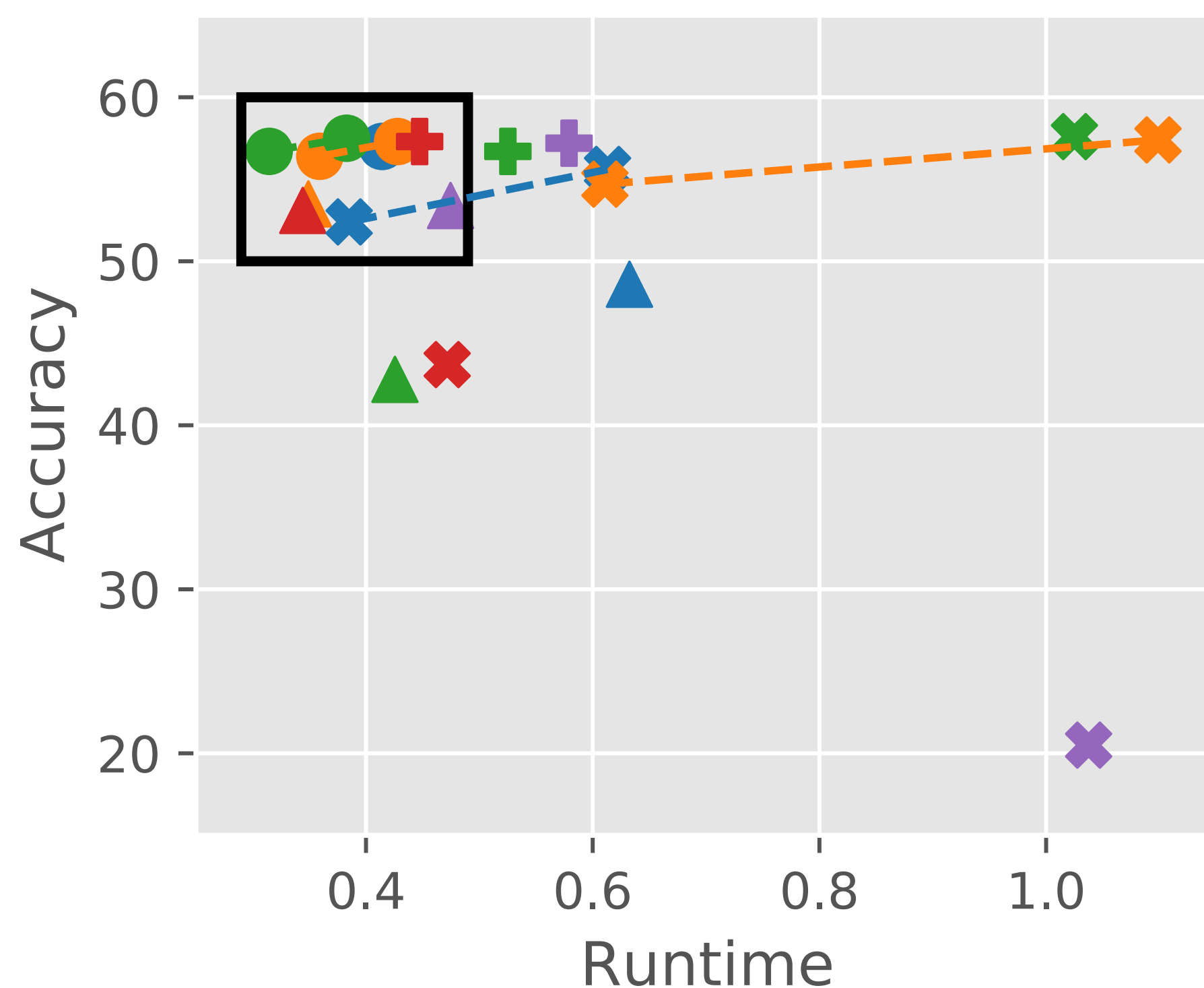


- Transformer
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

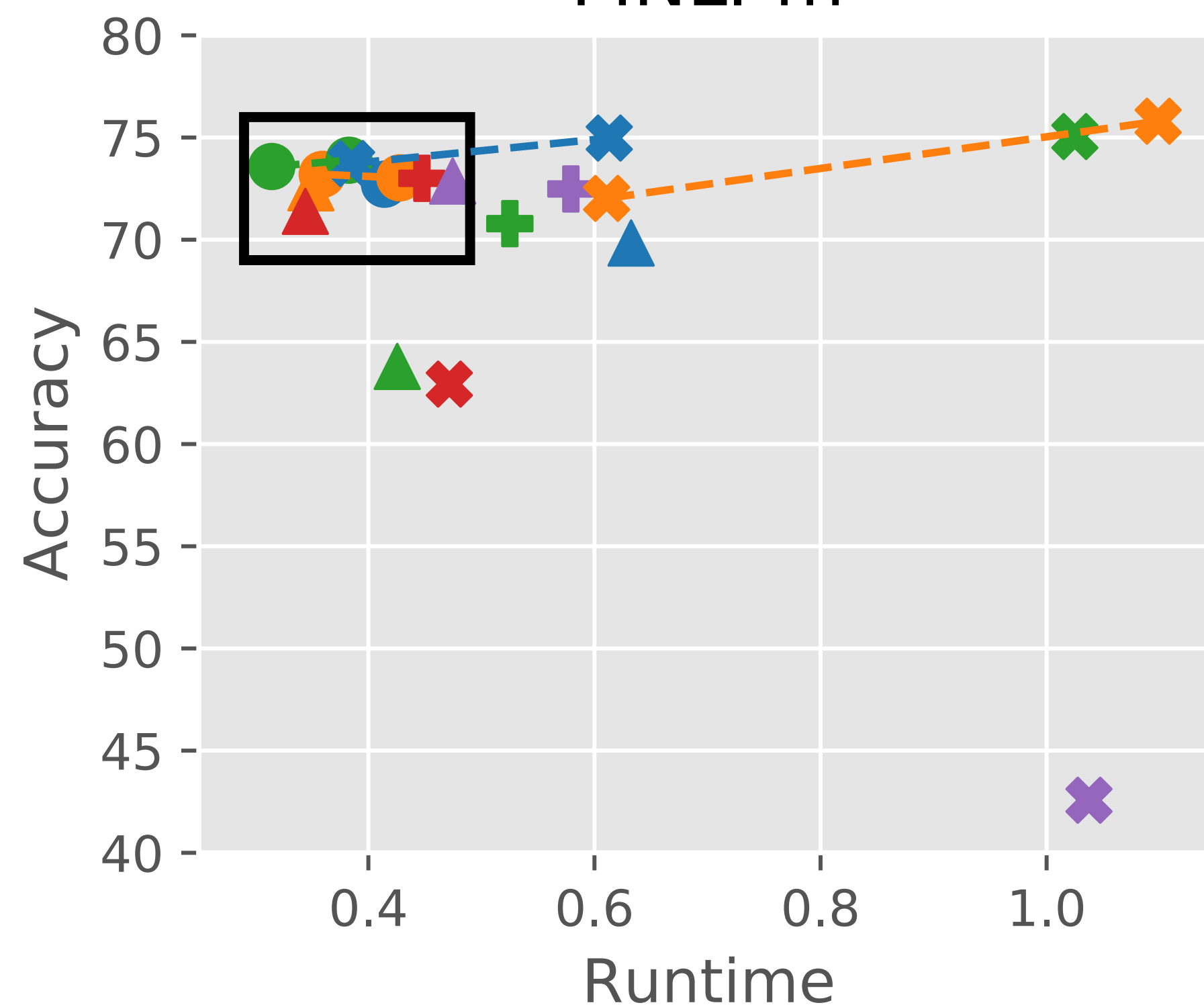
# Evaluations

## RoBERTa-small-512

Masked LM



MNLI-m



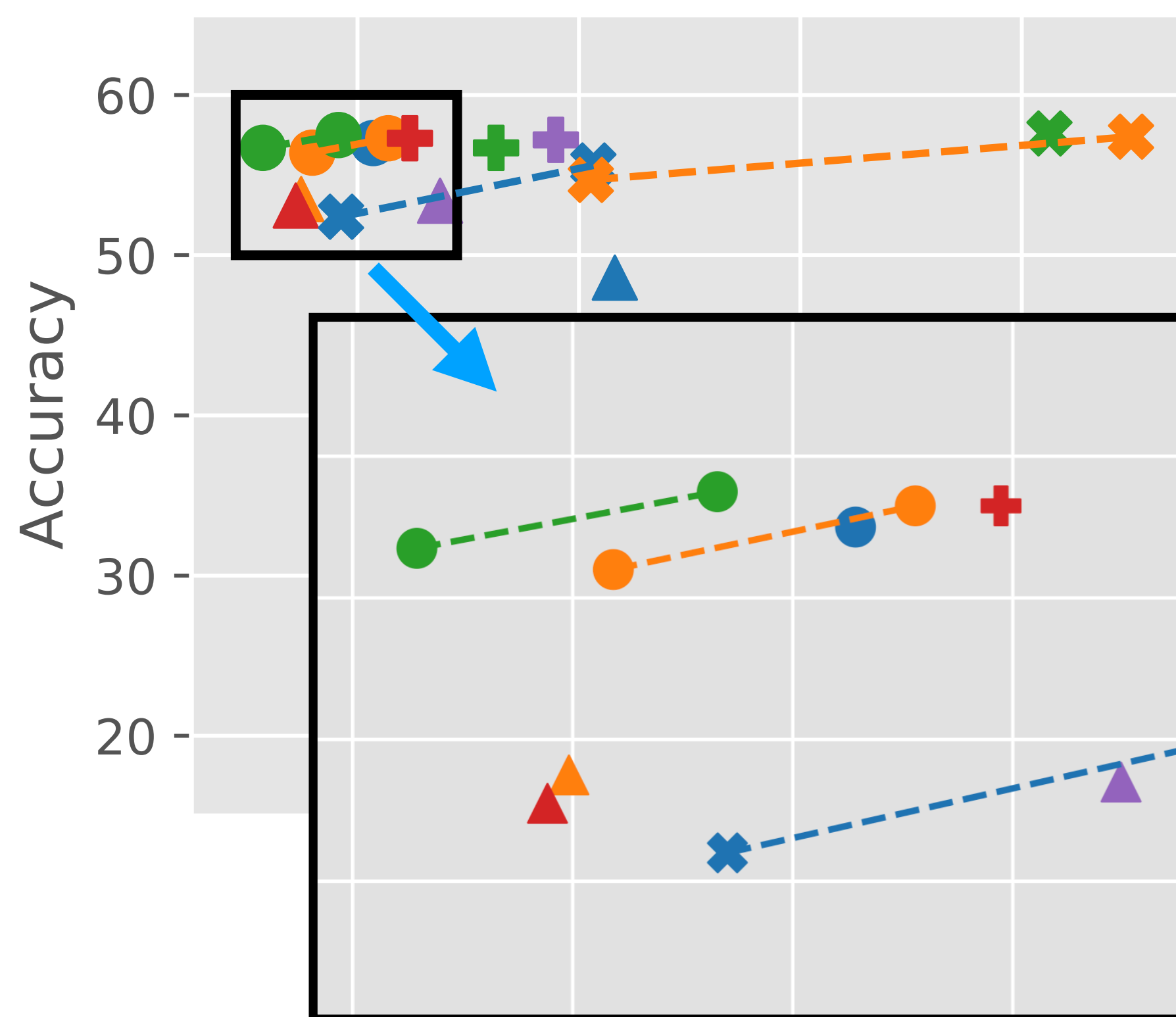
- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain



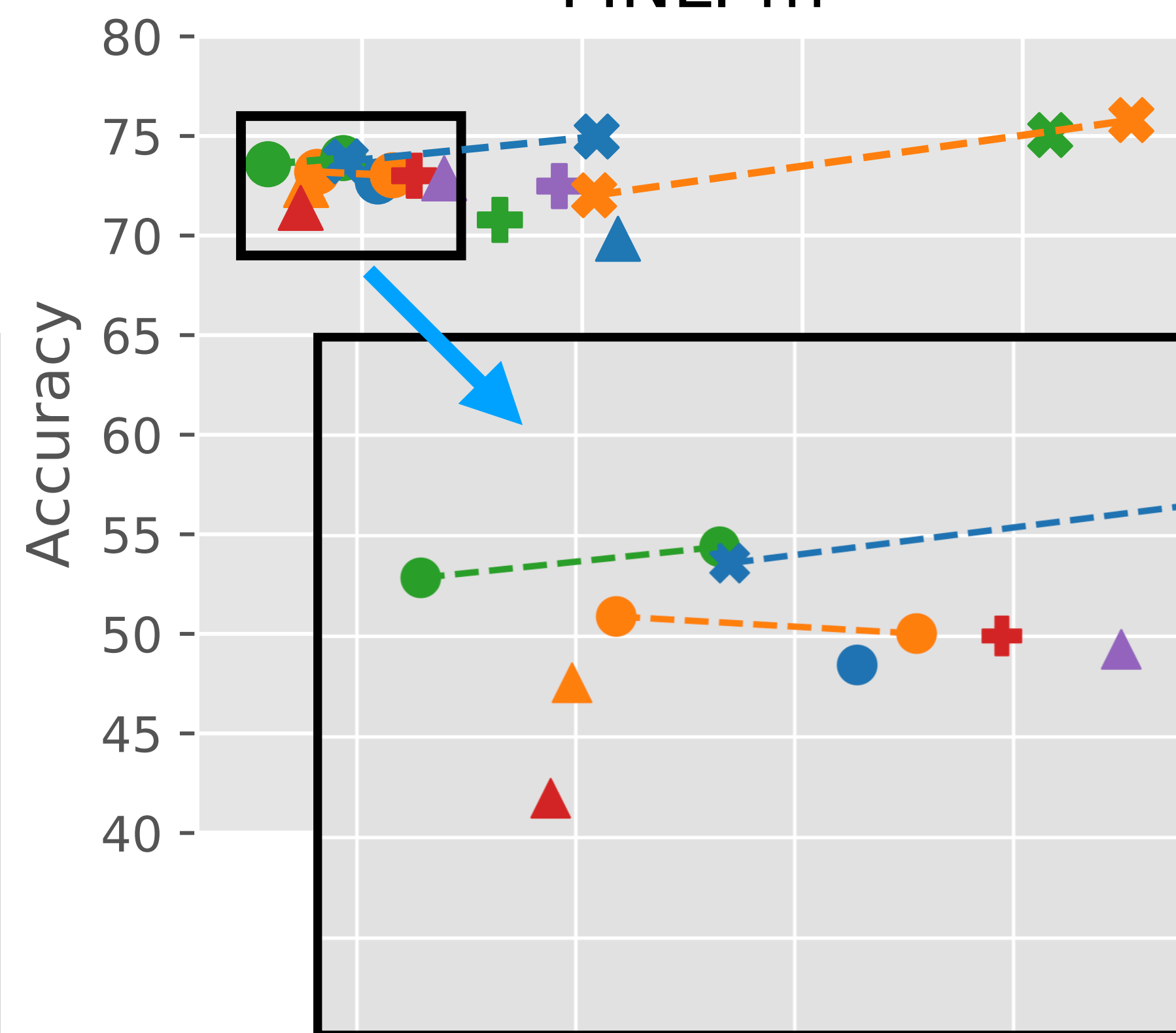
# Evaluations

## RoBERTa-small-512

Masked LM



MNLI-m

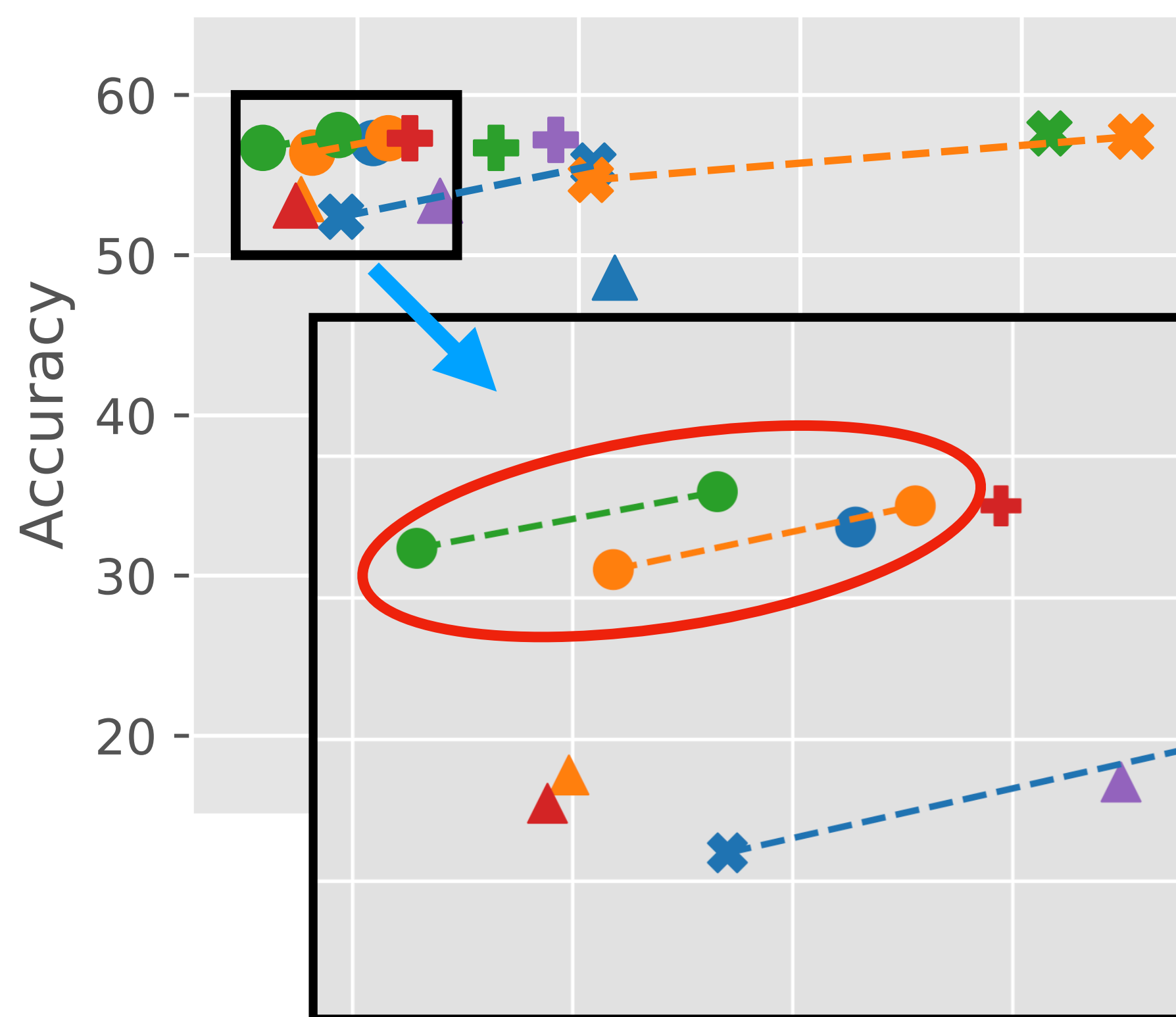


- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

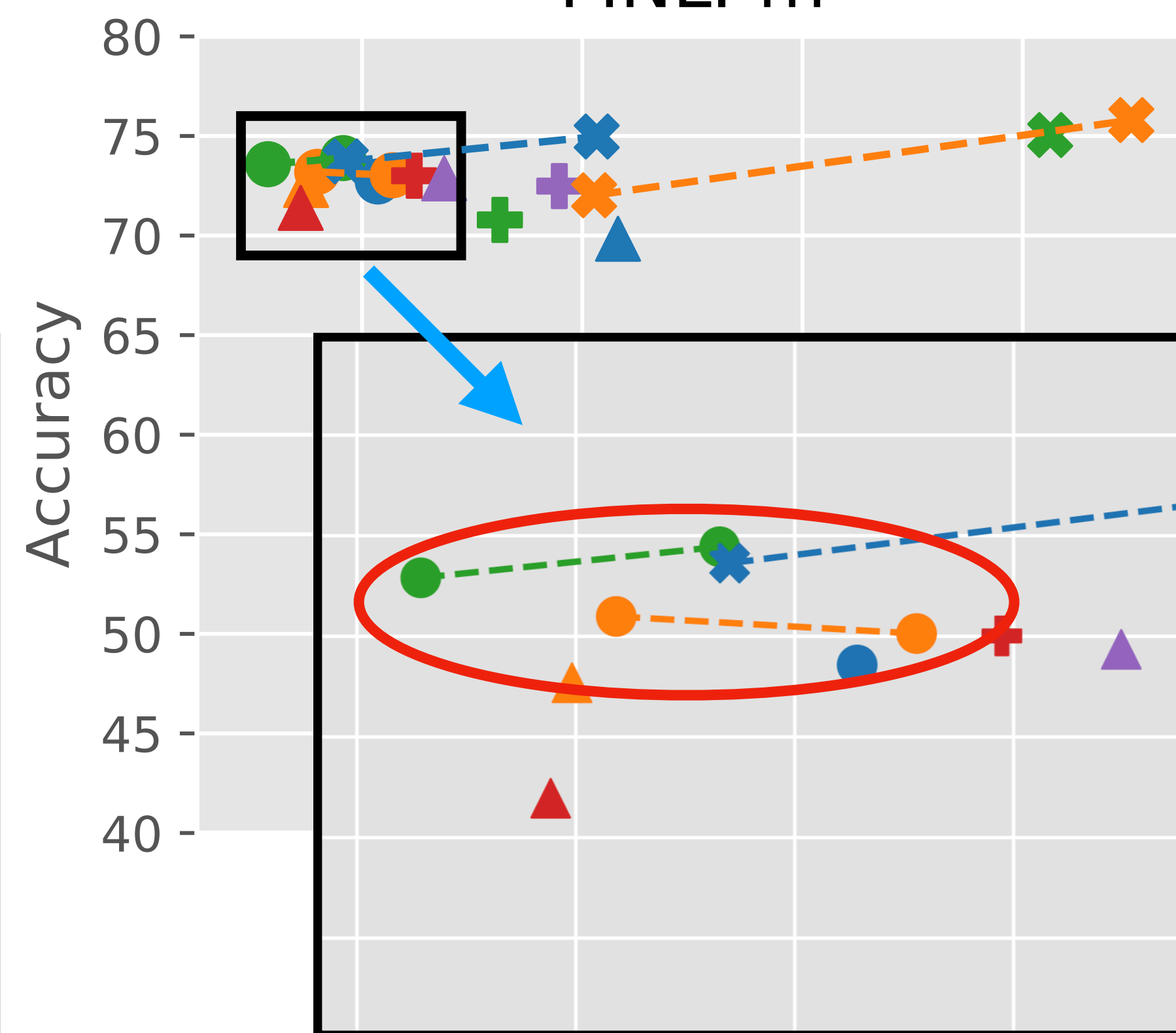
# Evaluations

## RoBERTa-small-512

Masked LM



MNLI-m



- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

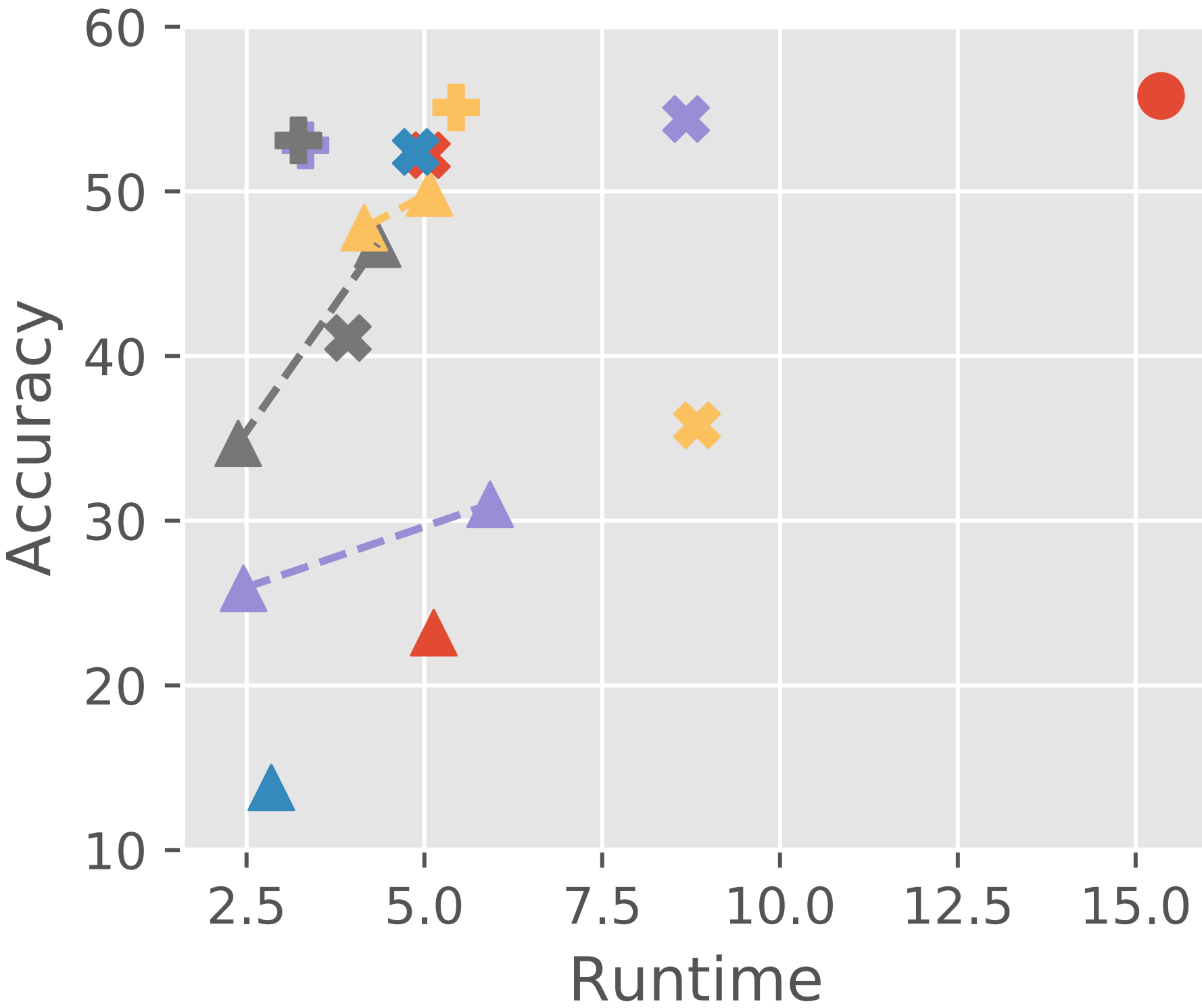
# **Evaluations**

## **RoBERTa-small-4096**

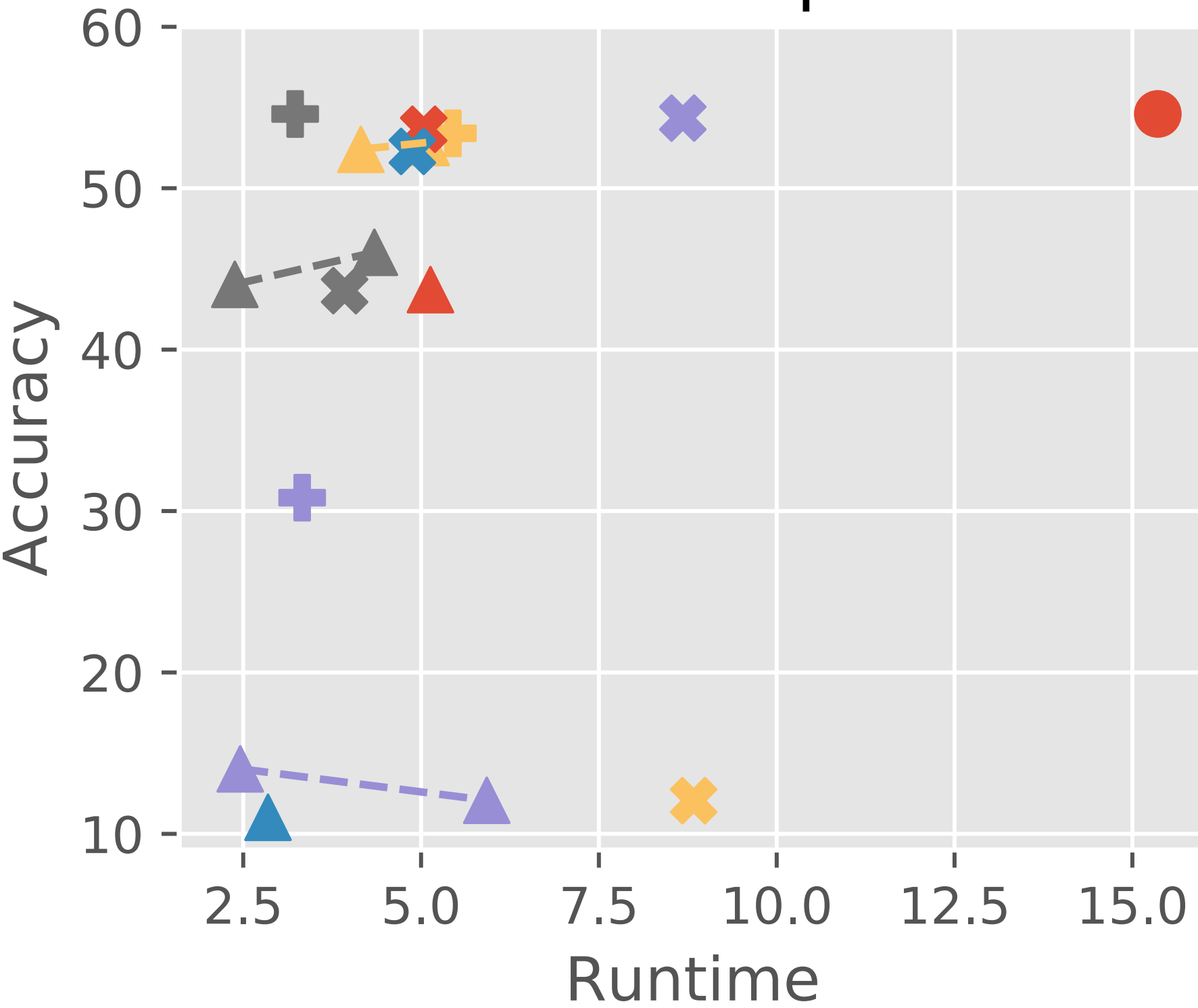
# Evaluations

## RoBERTa-small-4096

Masked LM



WikiHop

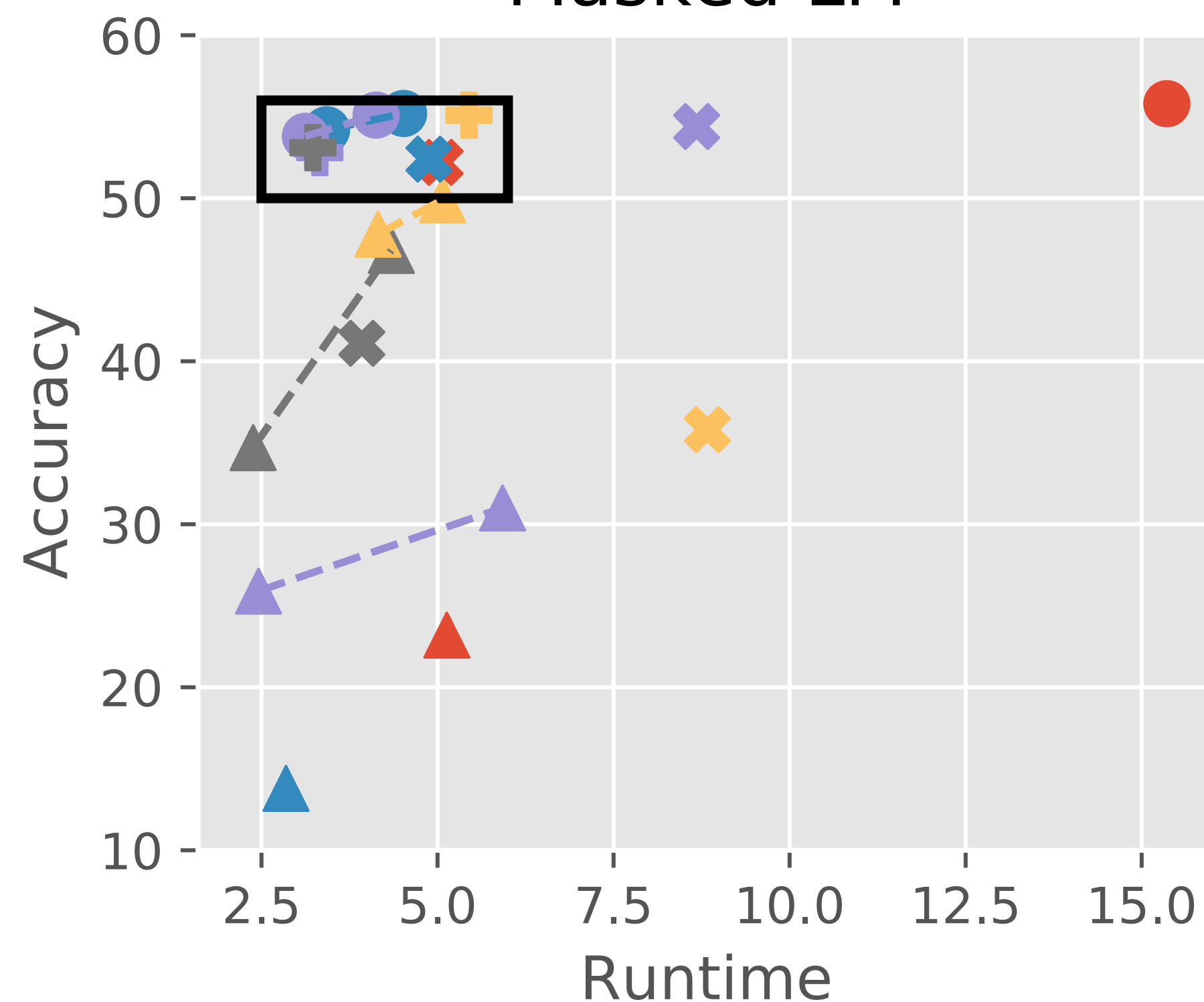


- Transformer
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

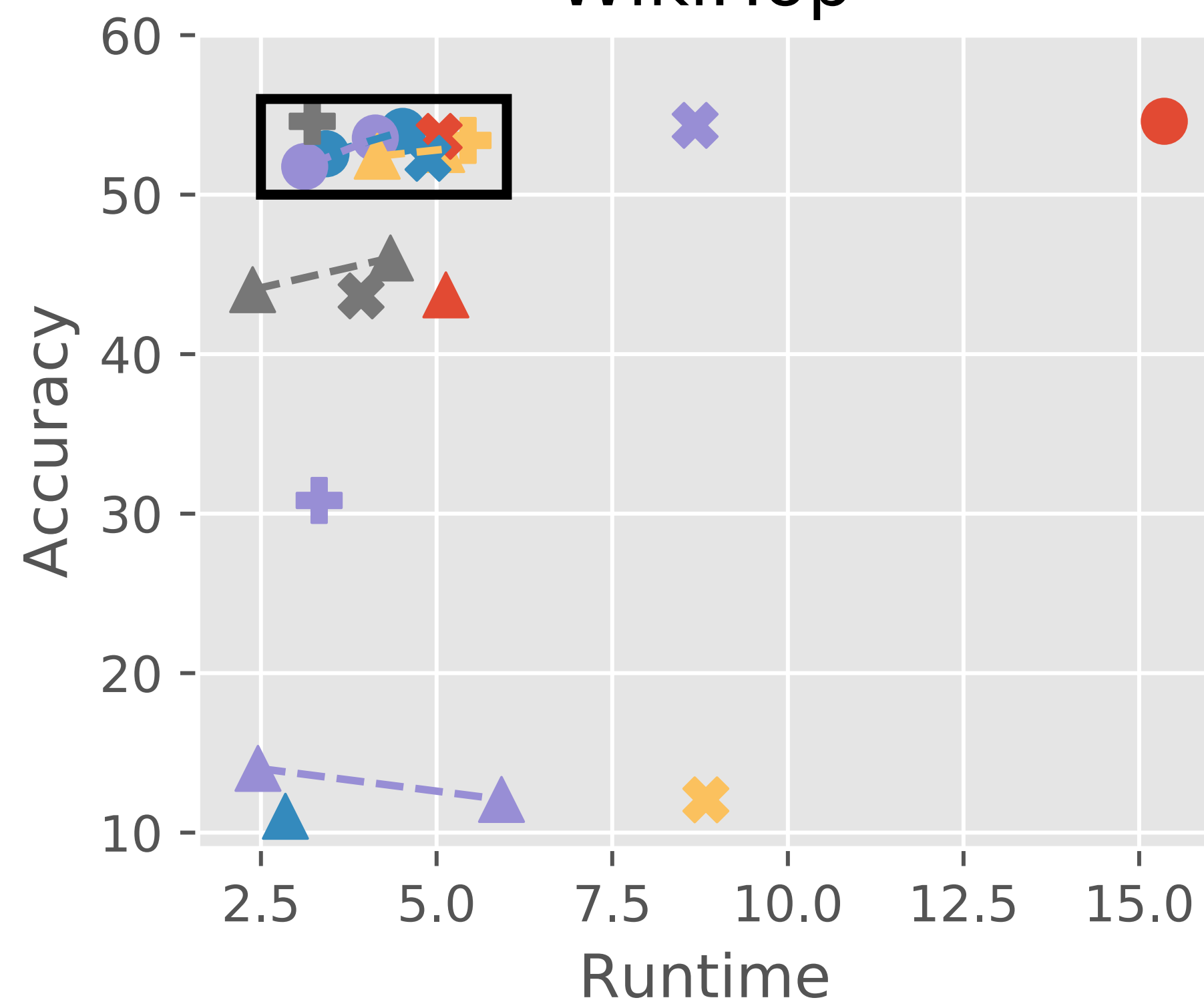
# Evaluations

## RoBERTa-small-4096

Masked LM



WikiHop

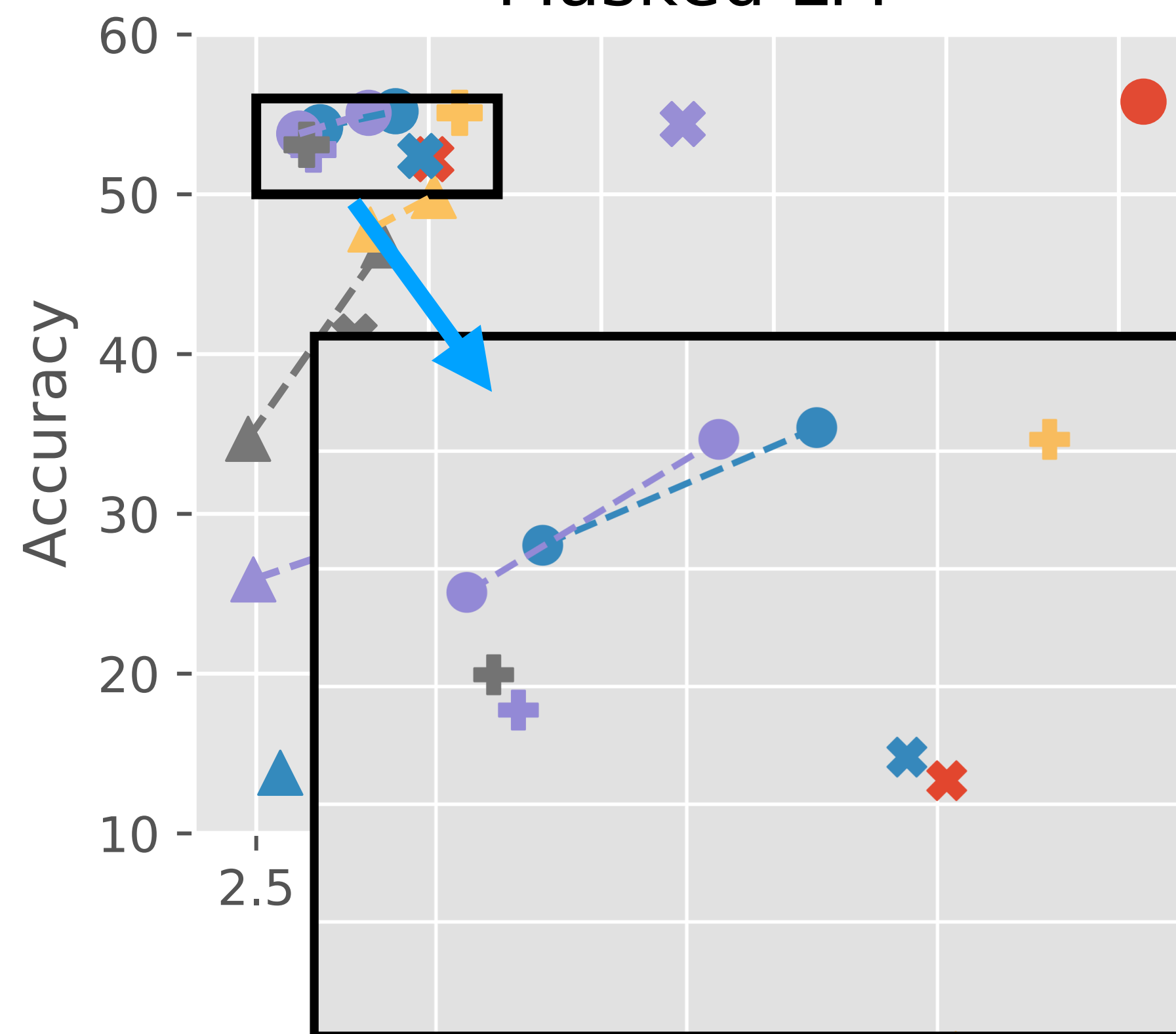


- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

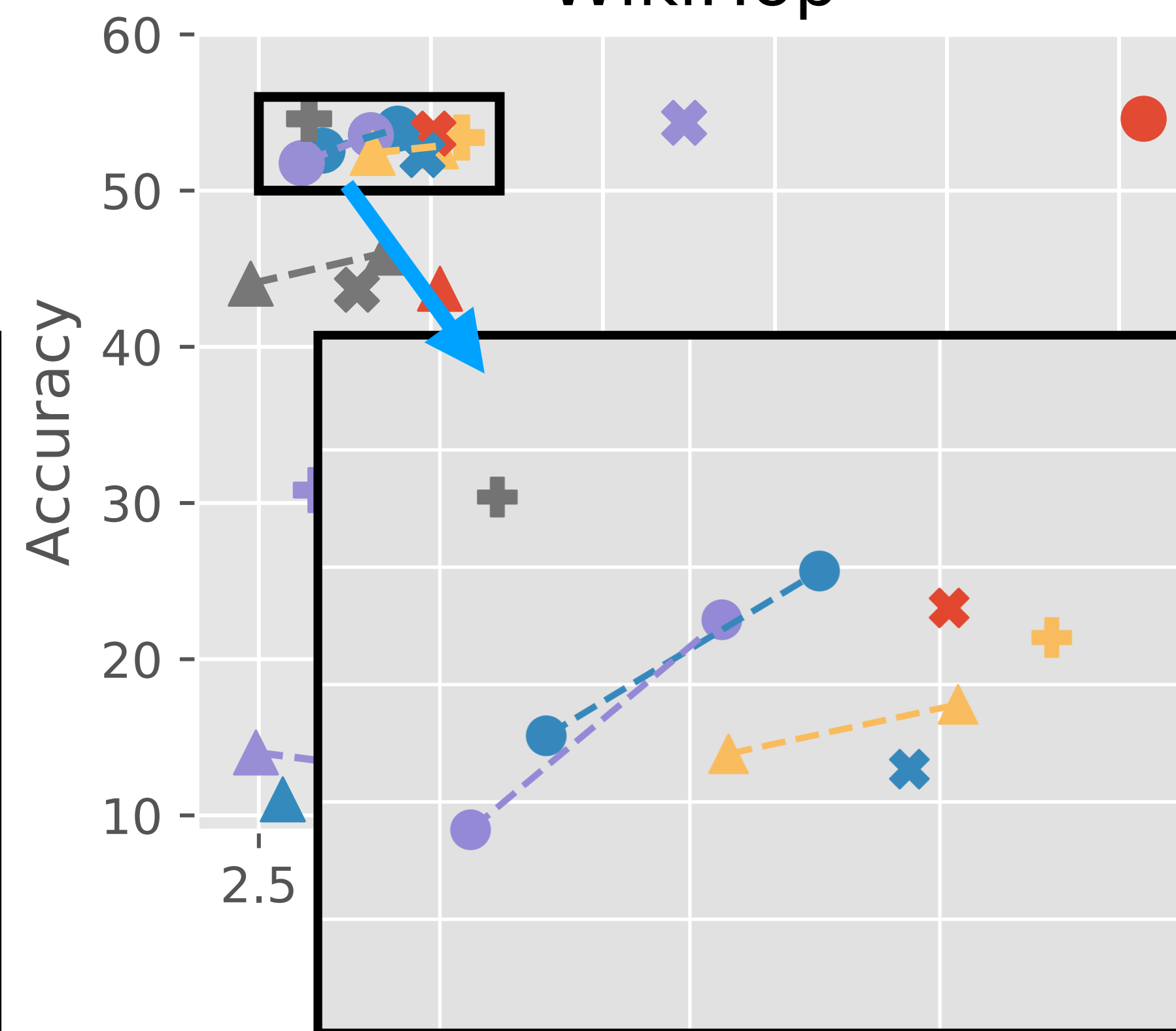
# Evaluations

## RoBERTa-small-4096

Masked LM



WikiHop

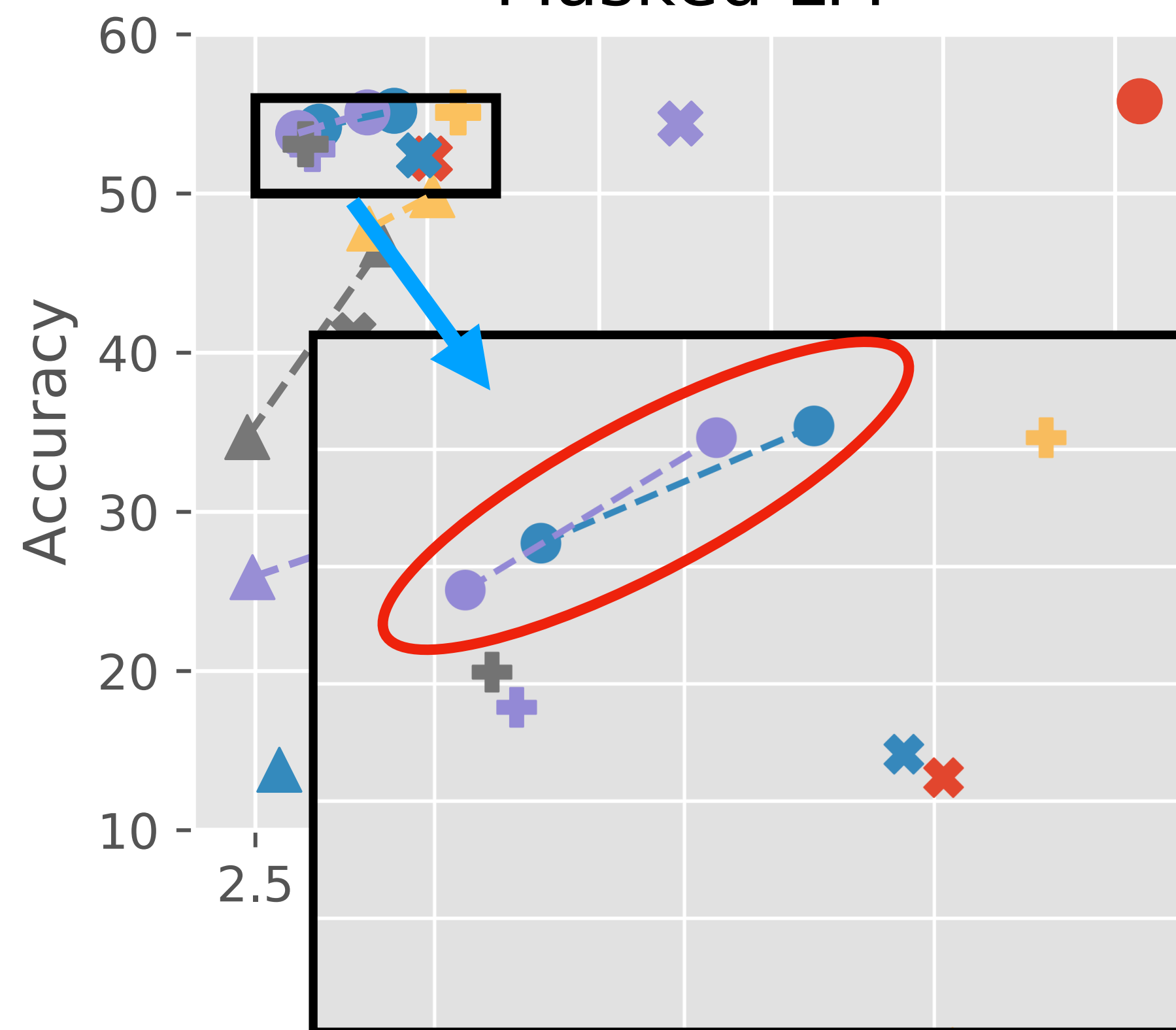


- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

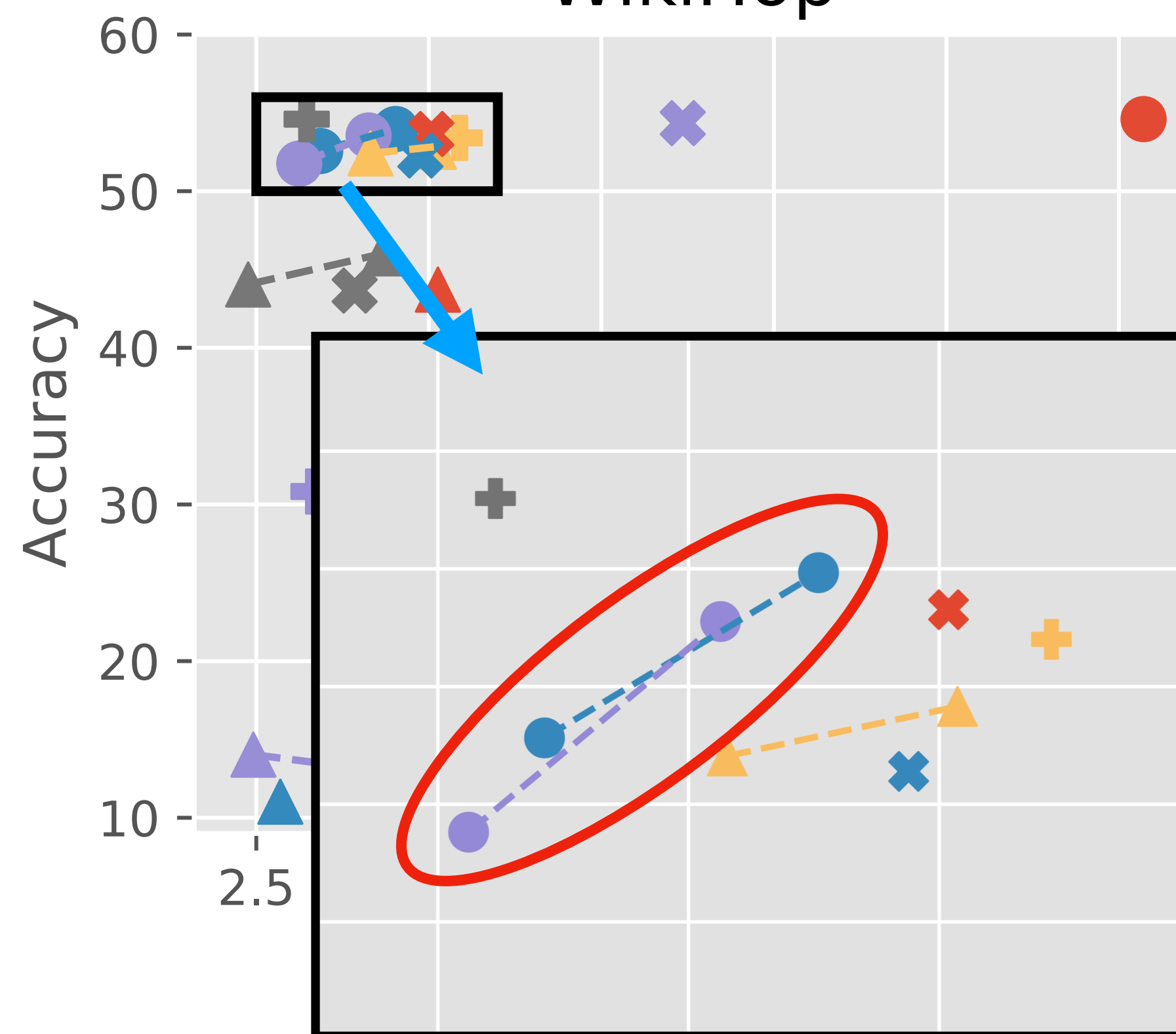
# Evaluations

## RoBERTa-small-4096

Masked LM



WikiHop



- Transformer
- MRA-2
- MRA-2-s
- Performer
- Linformer
- SOFT
- SOFT + Conv
- Nystromformer
- Nystrom + Conv
- YOSO
- YOSO + Conv
- Reformer
- Longformer
- Big Bird
- H-Transformer-1D
- Scatterbrain

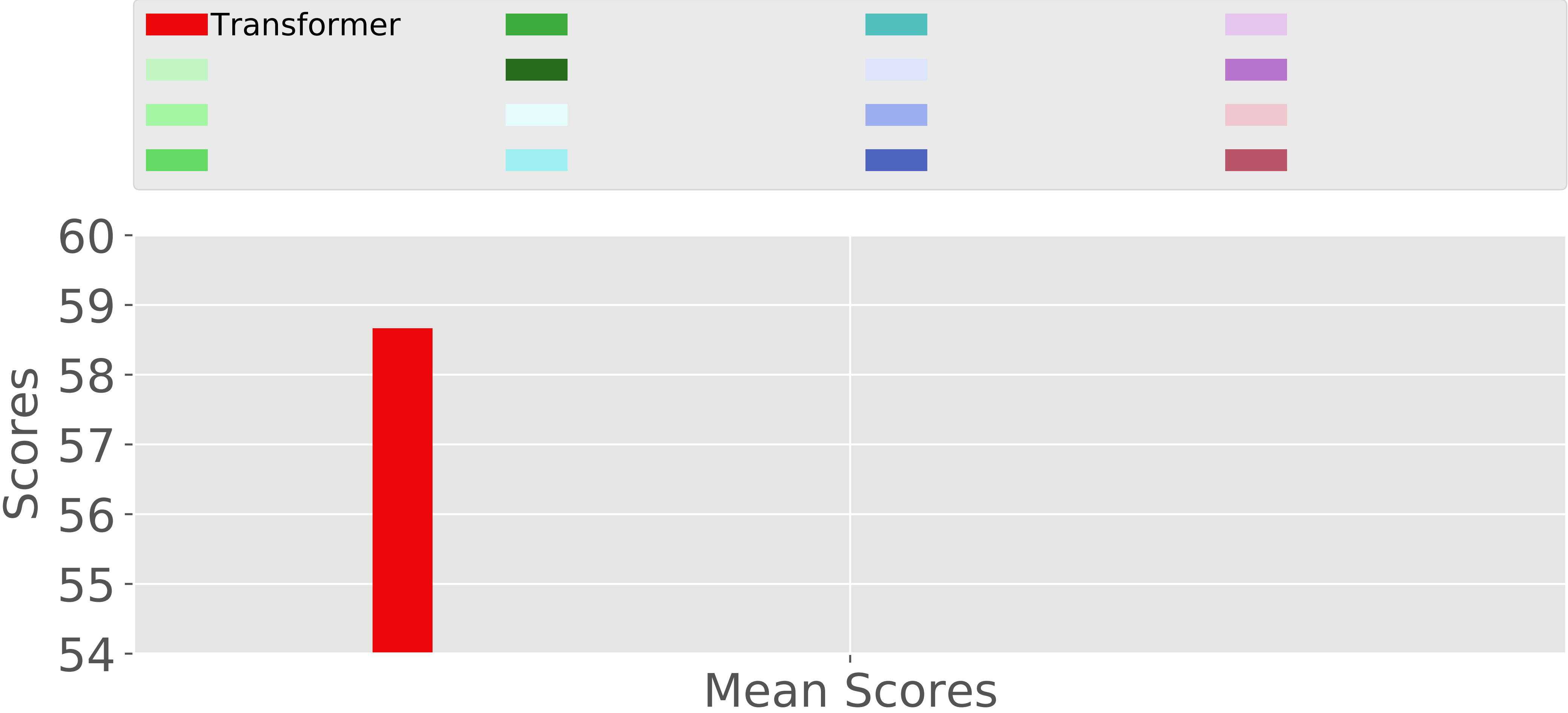
**Evaluations**

**Long Range Arena Benchmark**



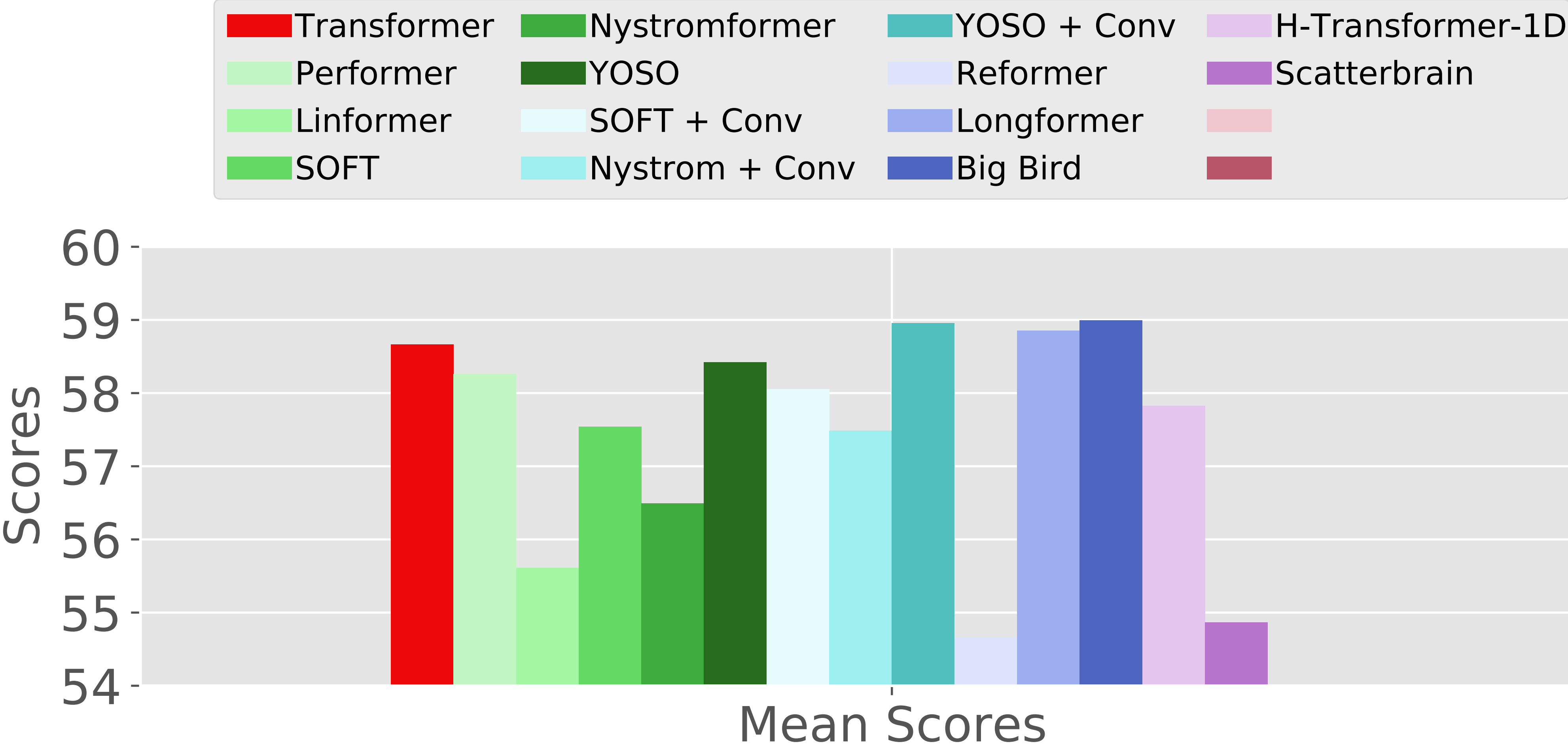
# Evaluations

## Long Range Arena Benchmark



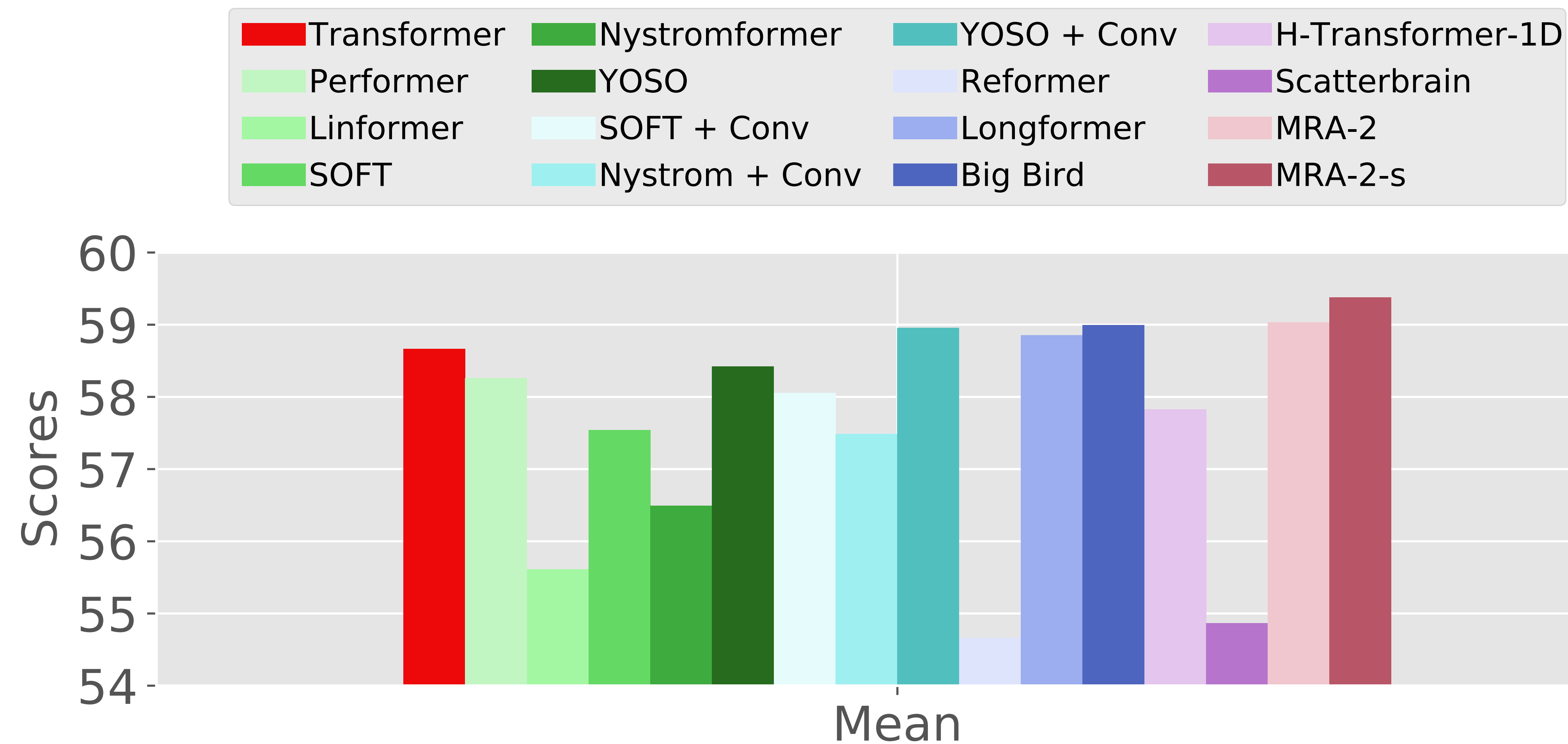
# Evaluations

## Long Range Arena Benchmark



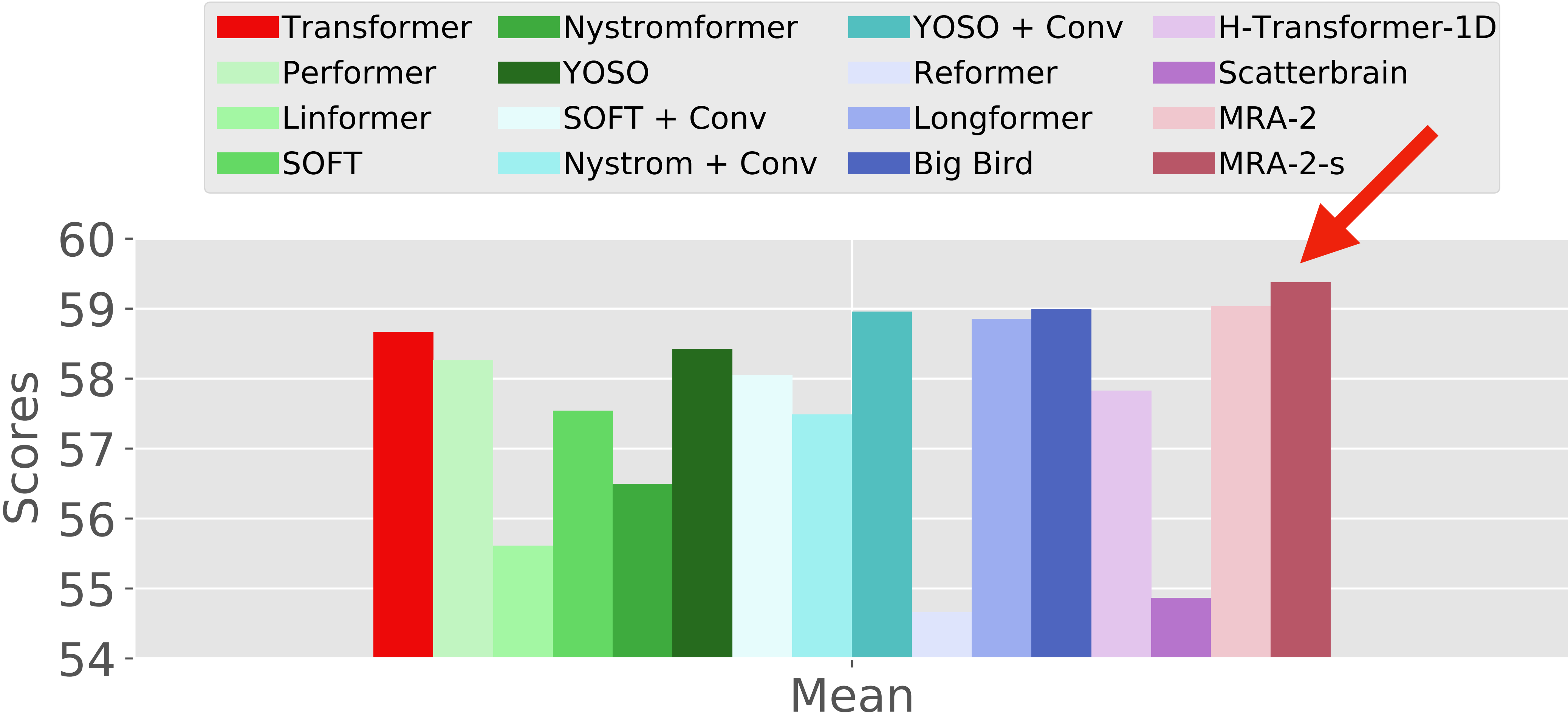
# Evaluations

## Long Range Arena Benchmark



# Evaluations

## Long Range Arena Benchmark

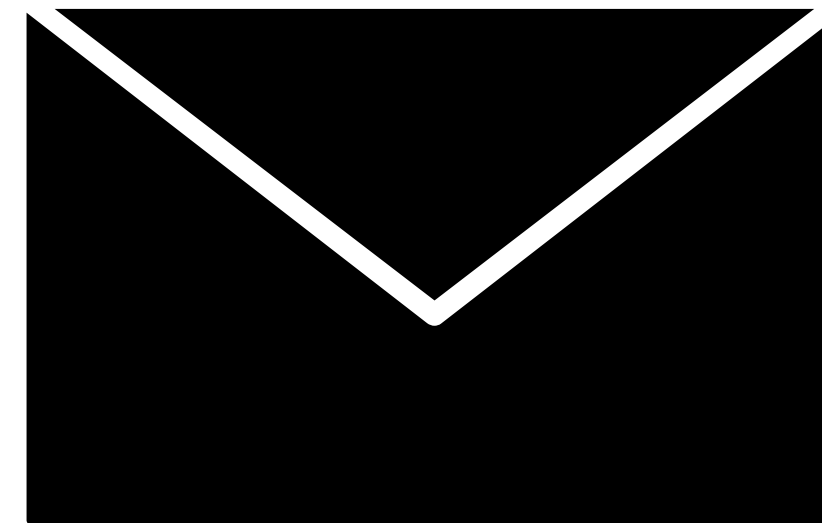


# Takeaway

- Multi-resolution analysis (MRA) ideas offer a win-win for efficient self-attention
  - potential access to a rich MRA theory
  - immediate practical benefits on efficiency and accuracy



mlpen/mra-attention



zzeng38@wisc.edu

main


1 branch

0 tags







Go to file

Add file

Code

 **mlpen** update

5fc2033 21 minutes ago 4 commits

	ImageNet	update	21 minutes ago
	LRA	update	21 minutes ago
	RoBERTa	add files	25 minutes ago
	supplement_code	add files	25 minutes ago
	.DS_Store	update	21 minutes ago
	README.md	Update README.md	23 minutes ago

README.md

Official Repo for Multi Resolution Analysis (MRA) for Approximate Self-Attention

About

No description, website, or topics provided.

Readme

0 stars

1 watching

0 forks

Releases

No releases published  
[Create a new release](#)

Packages

No packages published  
[Publish your first package](#)

https://github.com/mlpen/mra-attention