FlashMo: Geometric Interpolants and Frequency-Aware Sparsity for Scalable Efficient Motion Generation

NeurIPS 2025


Zeyu Zhang[1]*


Yiran Wang[1]*


Danning Li[2]*


Dong Gong[3]


Ian Reid[2]


Richard Hartley[1]

[1]The Australian National University

[2]Mohamed bin Zayed University of Artificial Intelligence

[3]The University of New South Wales

*Equal Contribution.

Motion Generation Challenges

- **Efficiency**: Motion diffusion models are computationally expensive and slow during both training and inference. Existing acceleration methods often rely on teacher models or complex mechanisms, leading to overengineering and limited efficiency.
- **Scalability**: Most motion diffusion models still depend on U-Net architectures and small datasets, limiting their scalability, while transformer-based designs show greater potential for large-scale generation.

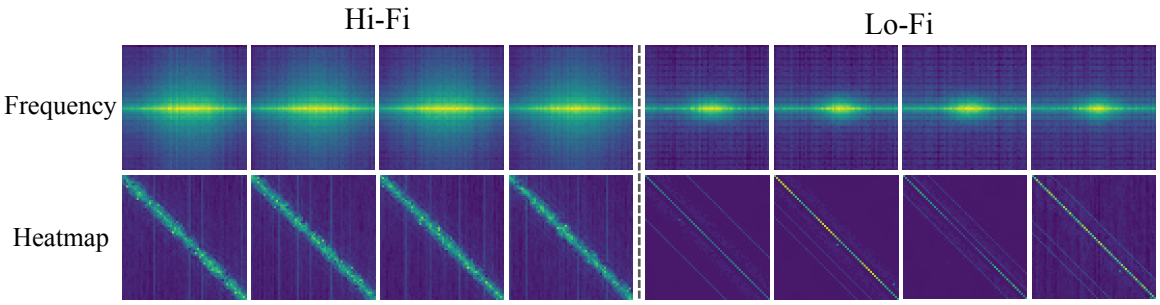
Motion Frequency Observation

A person is doing karate moves.

A person is dancing in the spot.

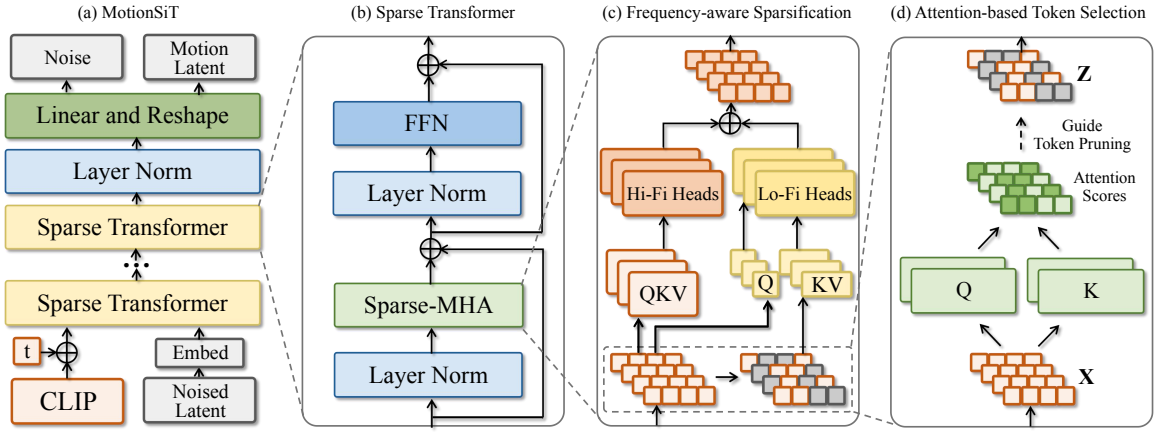
Frequency of human motion.

Frequency Magnitude Observation



Frequency magnitude vs. attention heatmap of attention heads.

FlashMo Architecture: MotionSiT



FlashMo architecture.

Geometric Interpolants

Temporal-Spatial Factorized Interpolant

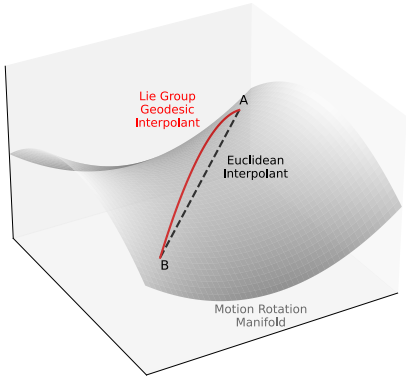
$$x_t = \alpha_T(t) \odot \alpha_S(t) \odot x^* + \sigma(t)\epsilon$$

Lie Group Geodesics Interpolant

$$x_t = \mathrm{Exp}(\alpha(t) \cdot \log(x^*) + \sigma(t) \cdot \epsilon)$$

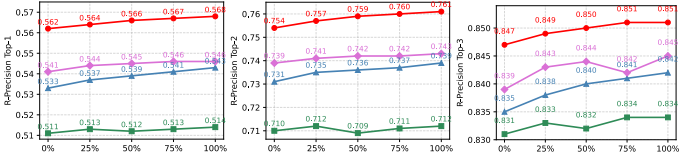
Geometric Factorized Interpolant

$$x_t^{\tau,s} = \mathrm{Exp}(\alpha_T(t)_\tau \cdot \alpha_S(t)_s \cdot \log(x^{*\tau,s}) + \sigma(t) \cdot \epsilon^{\tau,s})$$



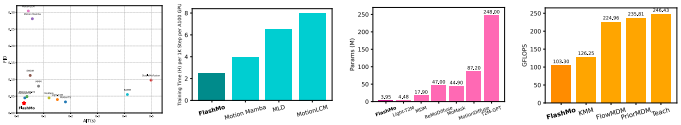
Euclidean interpolant vs. Lie group geodesics interpolant.

Scalability and Efficiency



(a) R-Precision Top-1 ↑ (b) R-Precision Top-2 ↑ (c) R-Precision Top-3 ↑

(d) FID ↓ (e) MM Dist ↓ (f) MModality ↑



(a) FID vs. AIT(s) ↙ (b) Training Time ↓ (c) Parameters ↓ (d) GFLOPS ↓

Visualization Results

End

Thank you.