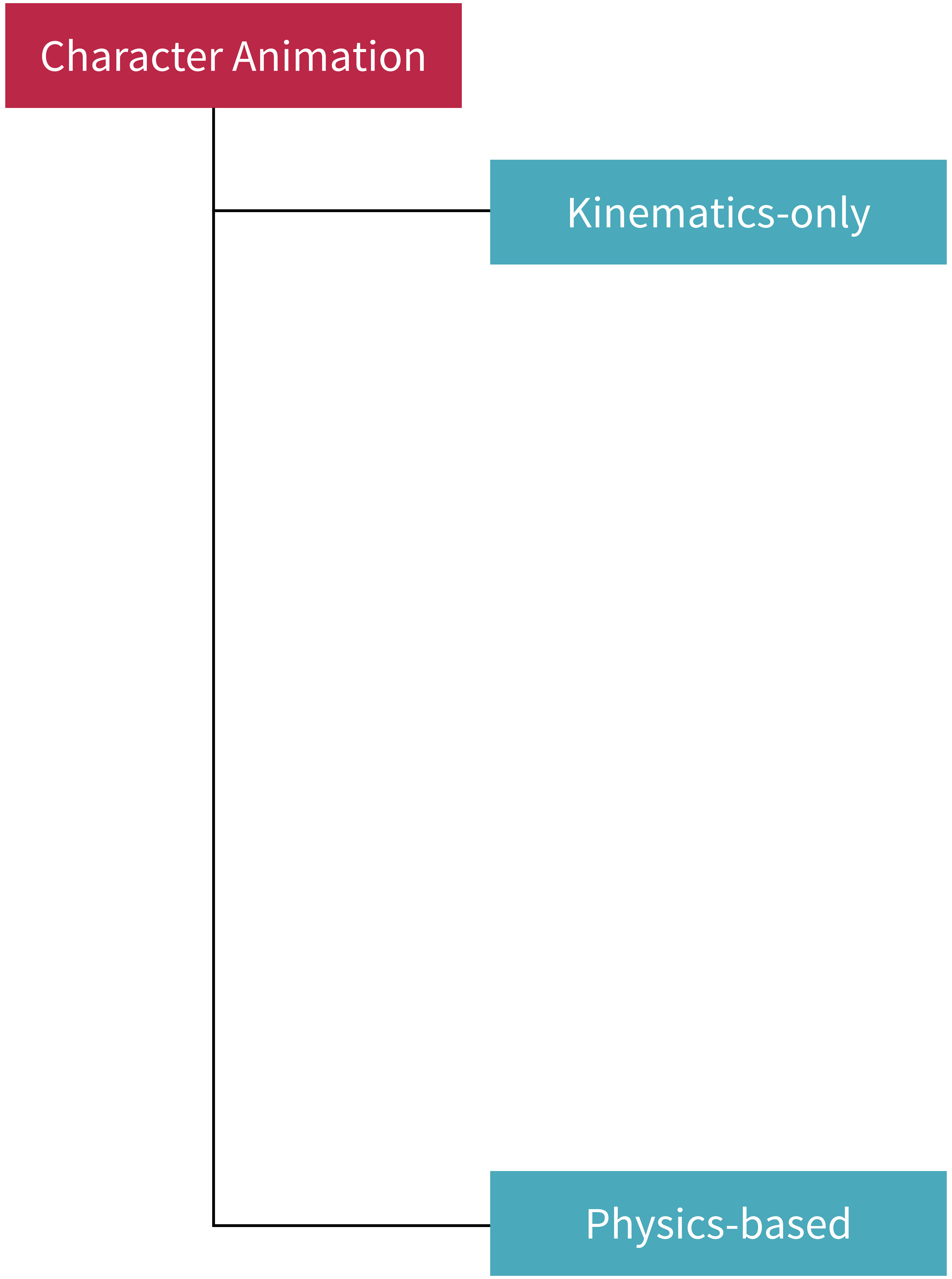
Lecture 17:

Data-driven Animation

FUNDAMENTALS OF COMPUTER GRAPHICS

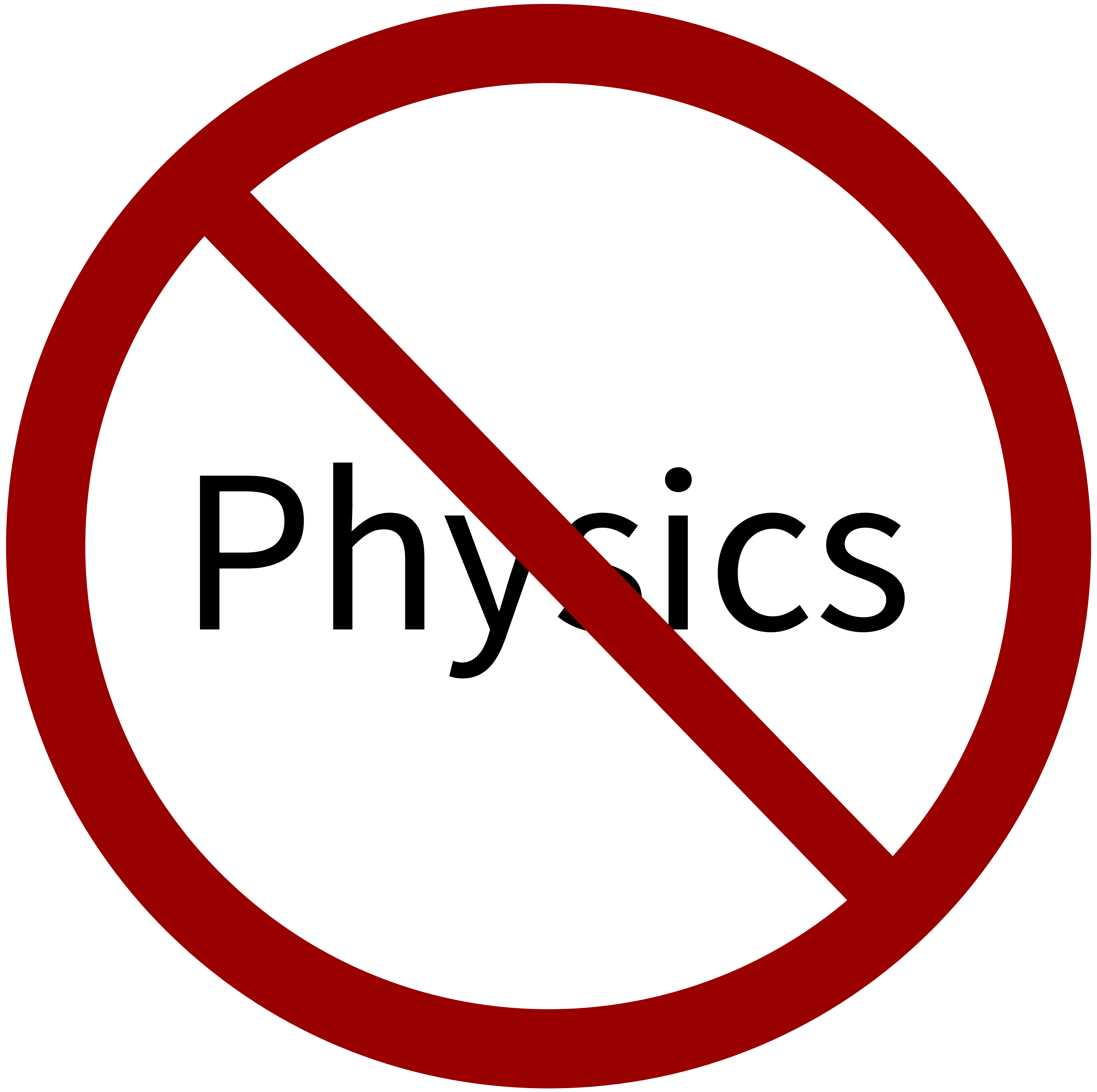
Animation & Simulation

Stanford CS248B


Character Animation
Kinematics-only
Physics-based

Character Animation

- Kinematics-only
- Physics-based



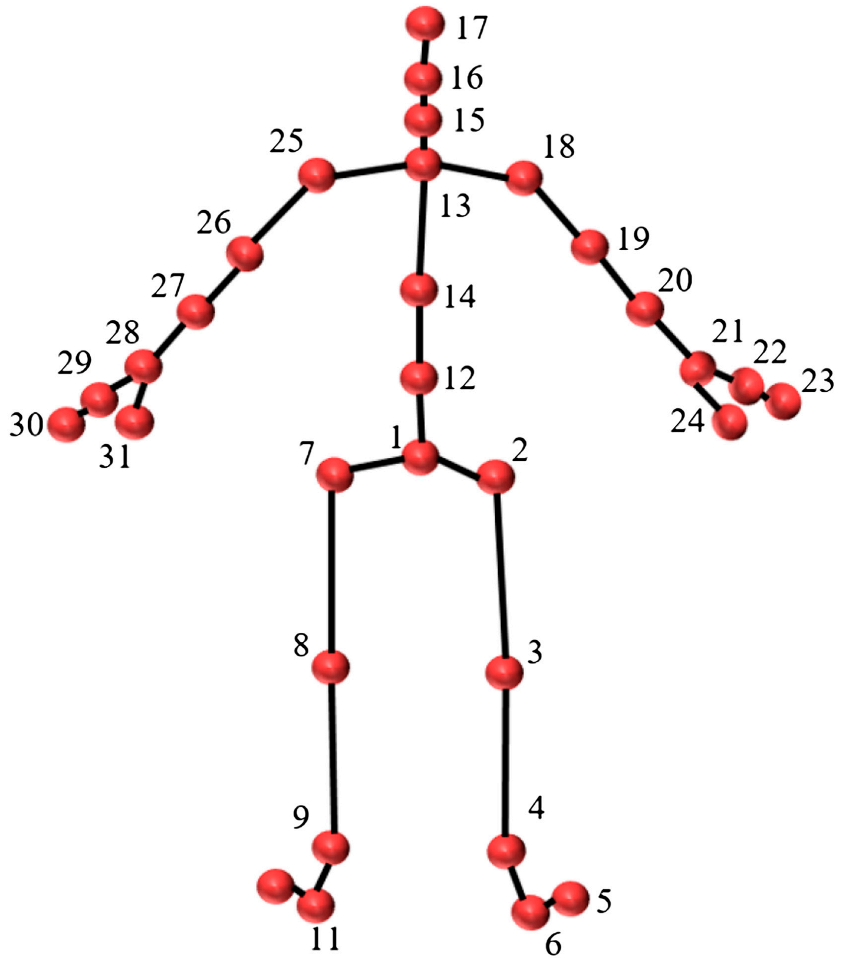
How to get data?

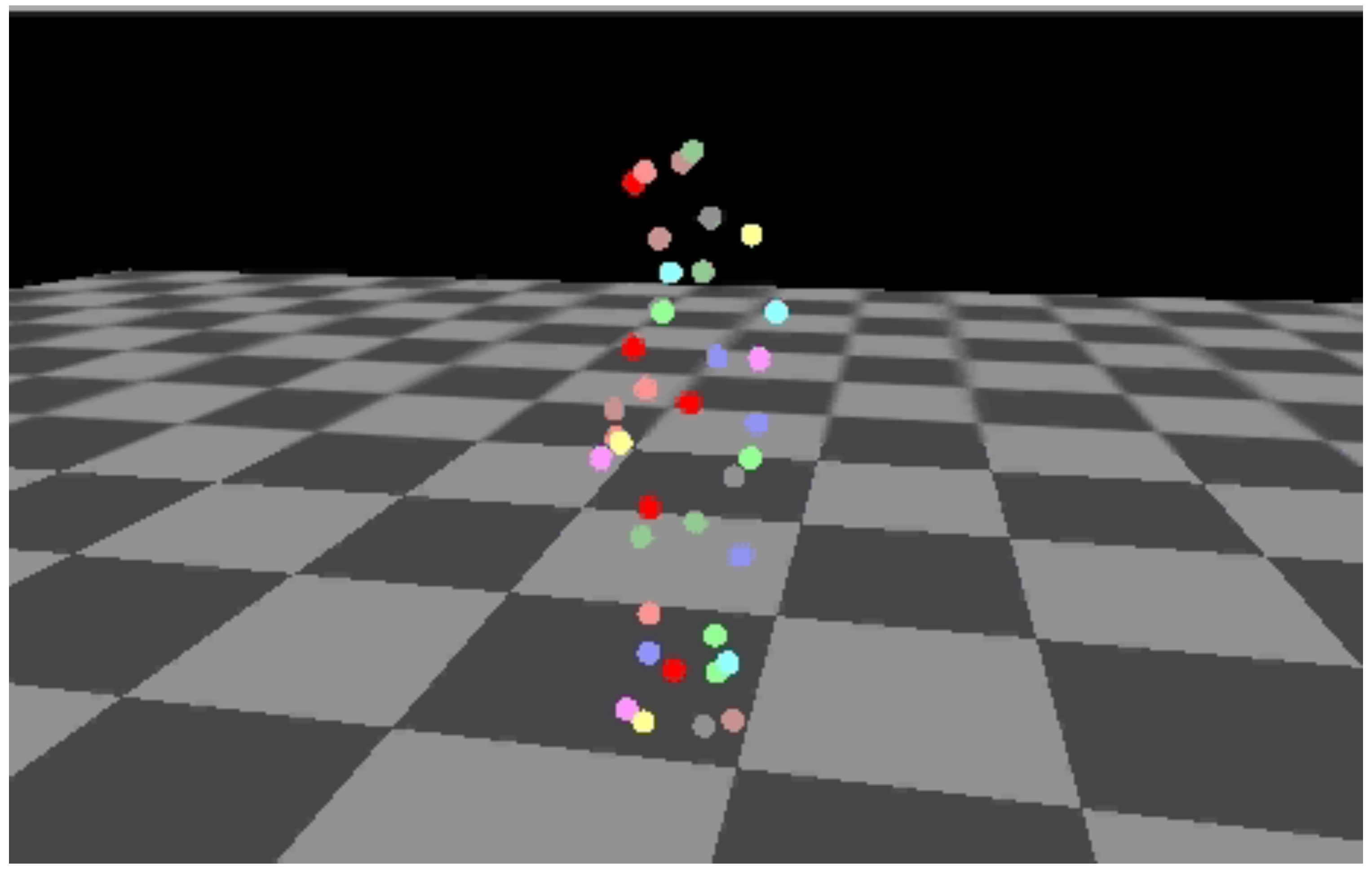
Optical systems

- Cameras
  - High temporal resolution (120+ fps)
  - Detect the locations of reflective markers
  - In principle, two cameras are sufficient to reconstruct the 3D location of a marker. In practice, more cameras can reduce occlusion and increase precision
- Markers:
  - passive: sensitive to infrared
  - active: emit LED light

Motion capture pipeline

Calibration → Capturing → Skeleton building → Marker labeling → Inverse kinematics → Trajectory smoothing



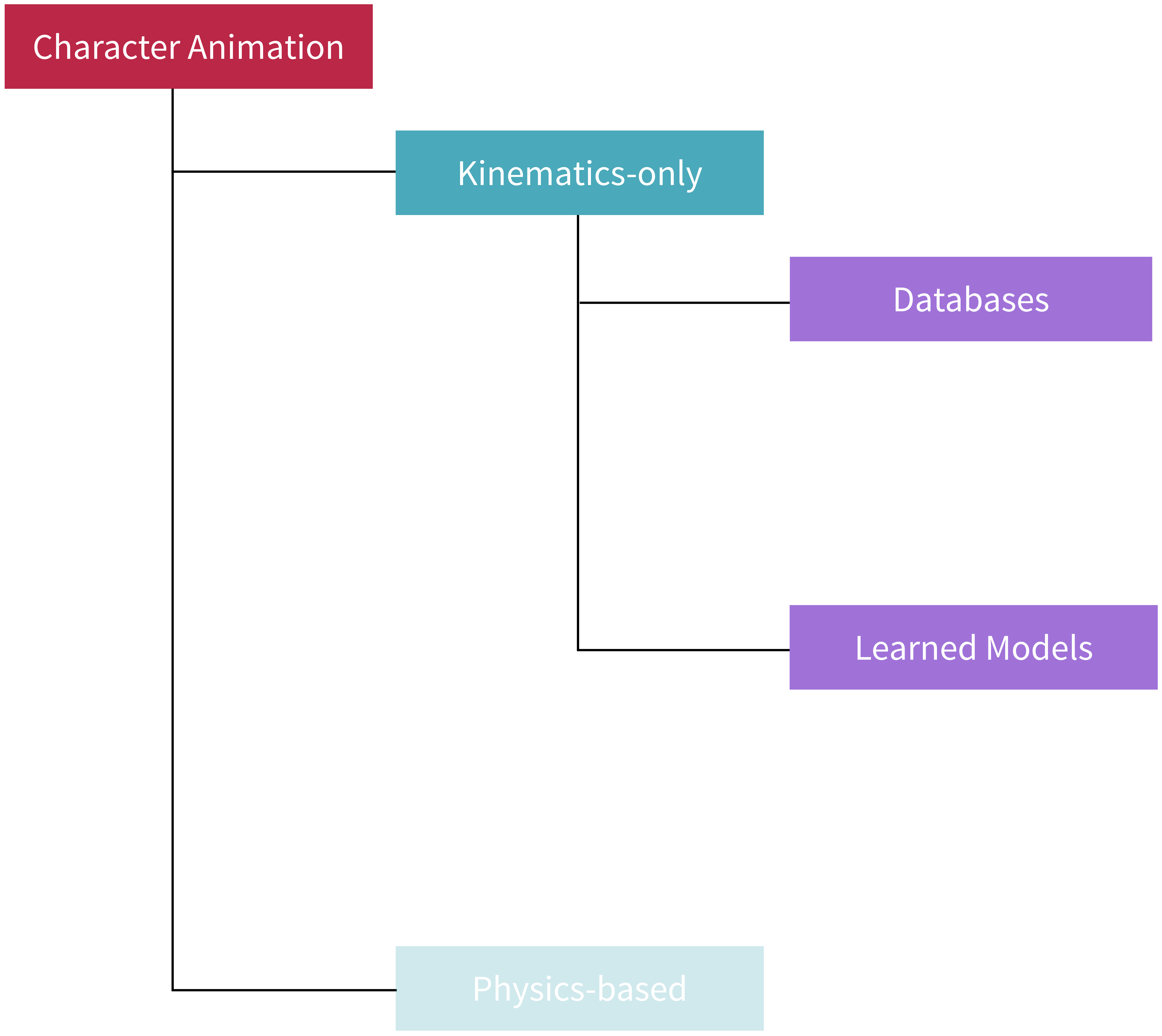


Motion databases

- A motion database consists of a collection of motion sequences with different motion types and lengths, but for the same character model parameterization.
- Existing motion databases, such as AMASS mocap database, provide motions for a wide variety of human activities.

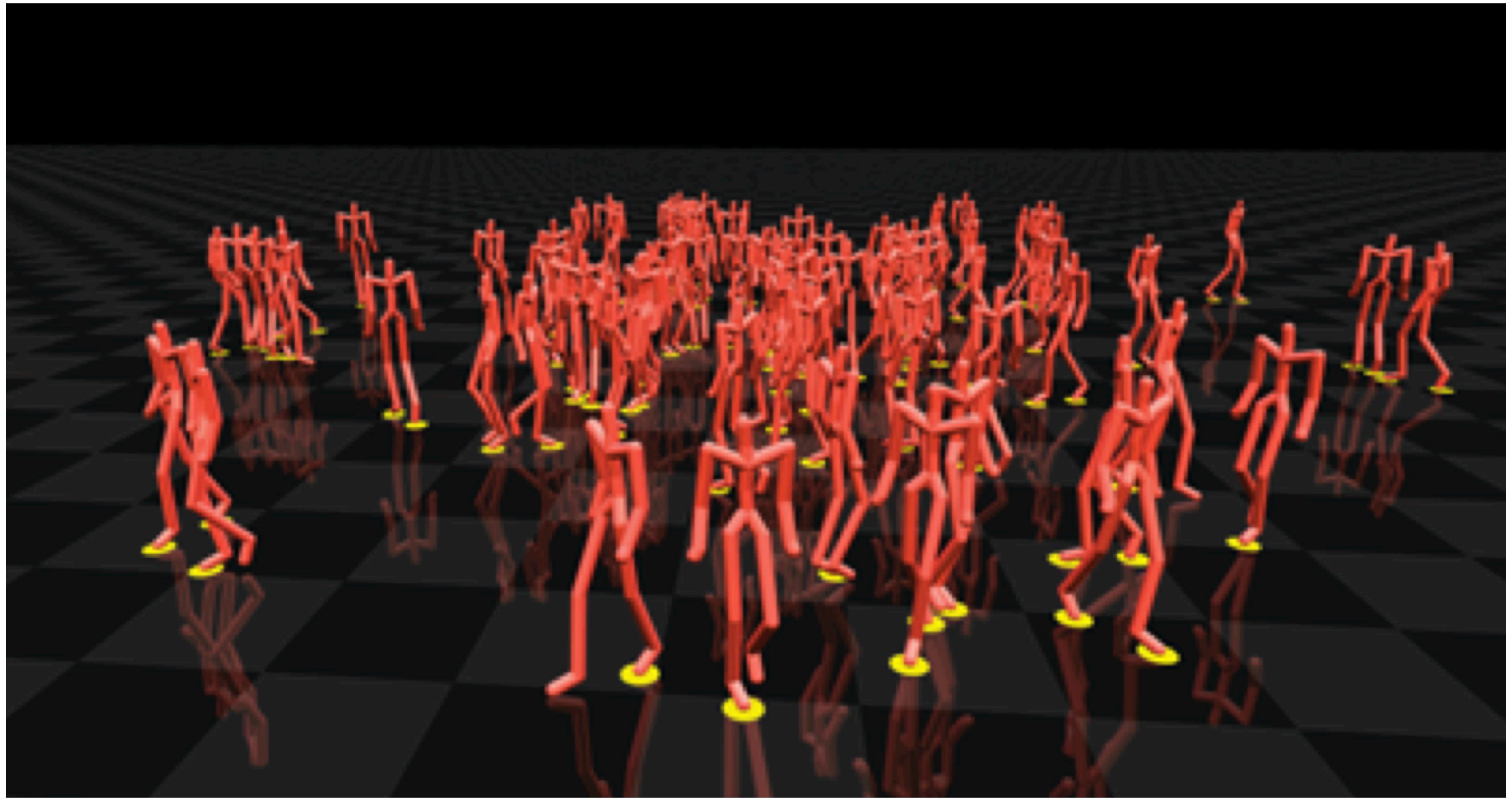
Dataset	Field of origin	License	Participants	Lower body kinematics	Upper body kinematics	GRF	EMG		
CMU	Graphics	Free to use + Redistribute	96	Yes	Yes	No	No	No	No
ACCAD	Graphics	CCv3.0	20	Yes	Yes	No	No	No	No
BMLmovi	Graphics	Free to redistribute	67	Yes	Yes	No	No	No	No
KIT	Graphics	Redistribute with citation	226	Yes	Yes	No	No	Yes	No
Eyes Japan	Graphics	CCv2.1	37	Yes	Yes	No	No	No	No
MPI HDM05	Graphics	CCv3.0	4	Yes	Yes	No	No	No	No
SFU	Graphics	Free for research purpose	7	Yes	Yes	No	No	No	No
TotalCapture	Graphics	Free for research purpose	5	Yes	Yes	No	No	Yes	No
Carmago et al	Biomechanics	Redistribute with citation	25	Yes	No	Yes	Yes	Yes	No
Fukuchi et al	Biomechanics	Redistribute with citation	49	Yes	Yes	Yes	No	No	No
Moore et al	Biomechanics	Redistribute with citation	15	Yes	Yes	Yes	No	No	No
Macaluso et al	Biomechanics	Redistribute with citation	10	Yes	No	Yes	Yes	No	Yes
Embry et al	Biomechanics	Redistribute with citation	10	Yes	No	Yes	Yes	No	No
Tiziana et al	Biomechanics	Redistribute with citation	50	Yes	Yes	Yes	Yes	No	No
Collins lab data	Unpublished	???	???	Yes	Yes	Yes	No	Yes	Yes

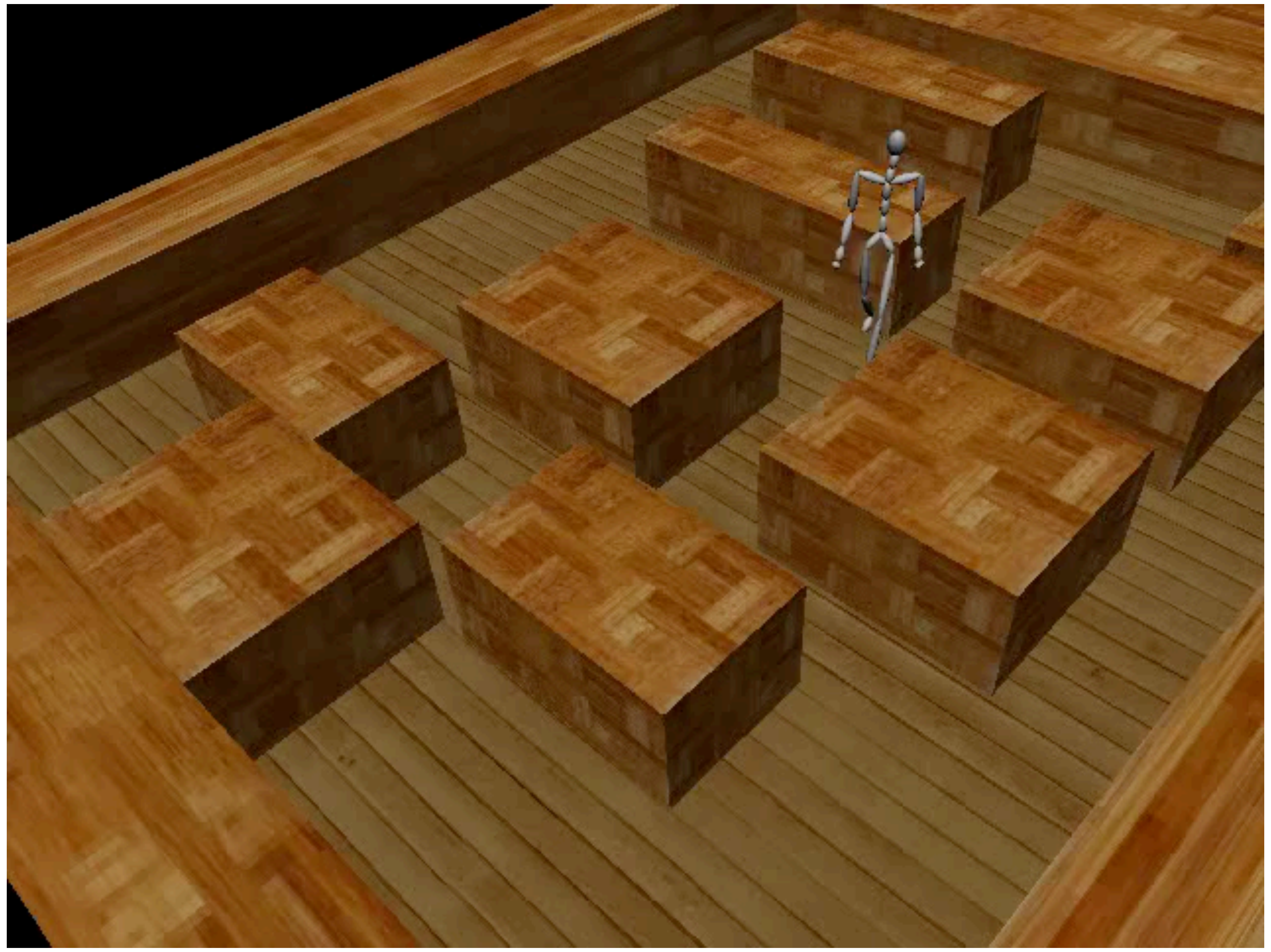
How to use data?

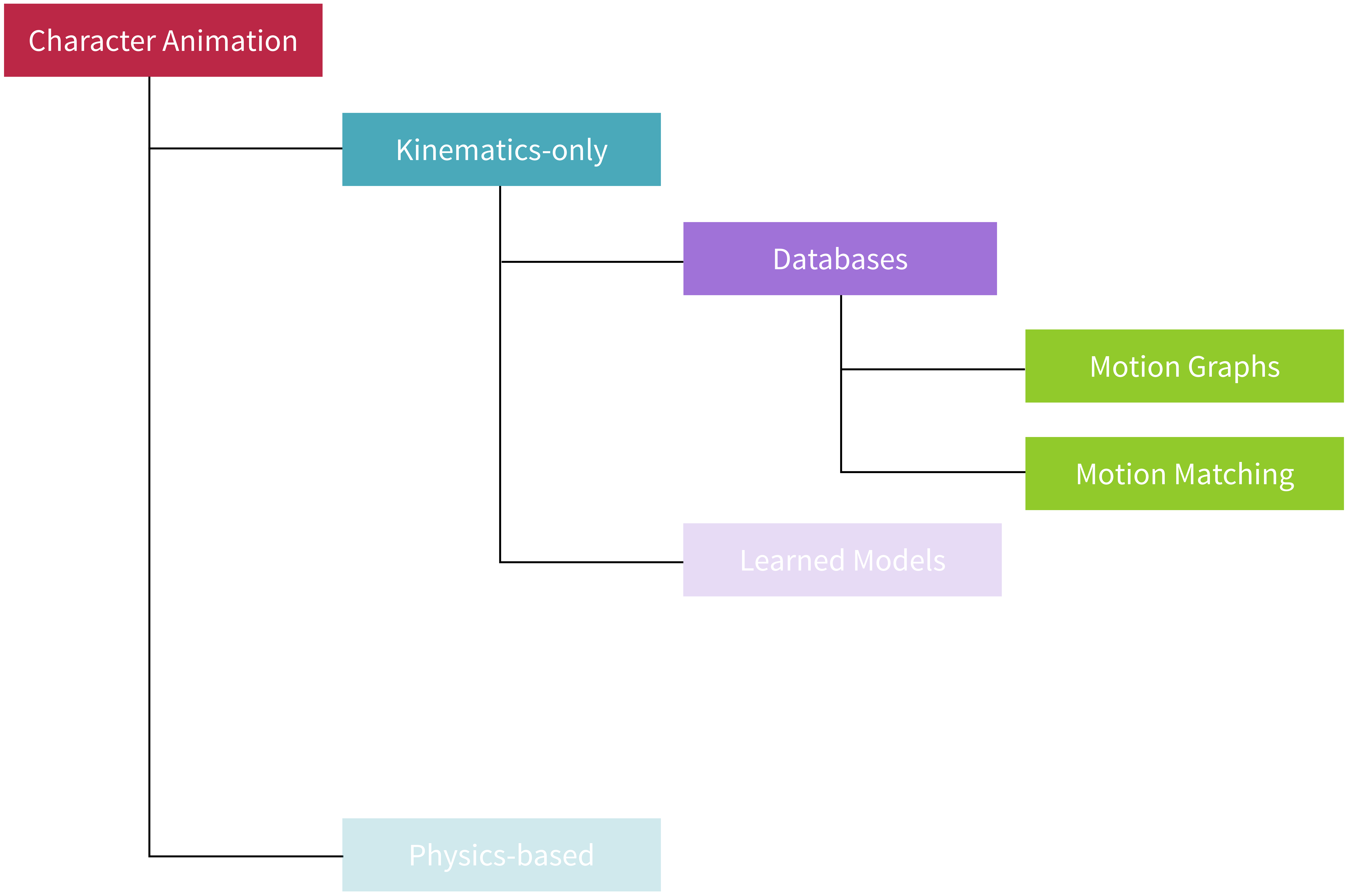

Character Animation
Kinematics-only
Databases
Learned Models
Physics-based

Synthesize animation from a motion database

- Which frame to play next, given the current frame and user constraints?
- How to make transitions smooth?

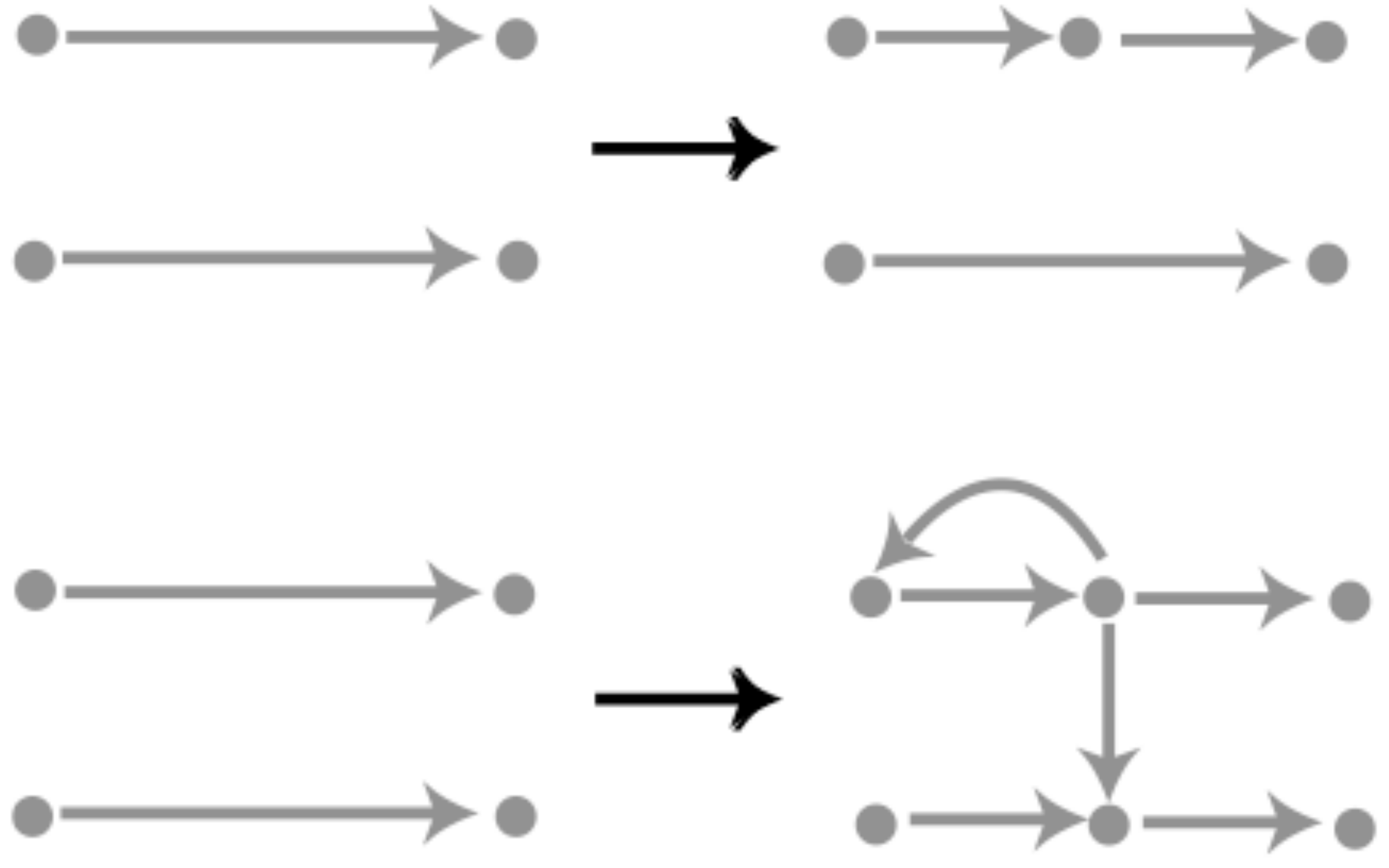





Character Animation
Kinematics-only
Databases
Motion Graphs
Motion Matching
Learned Models
Physics-based

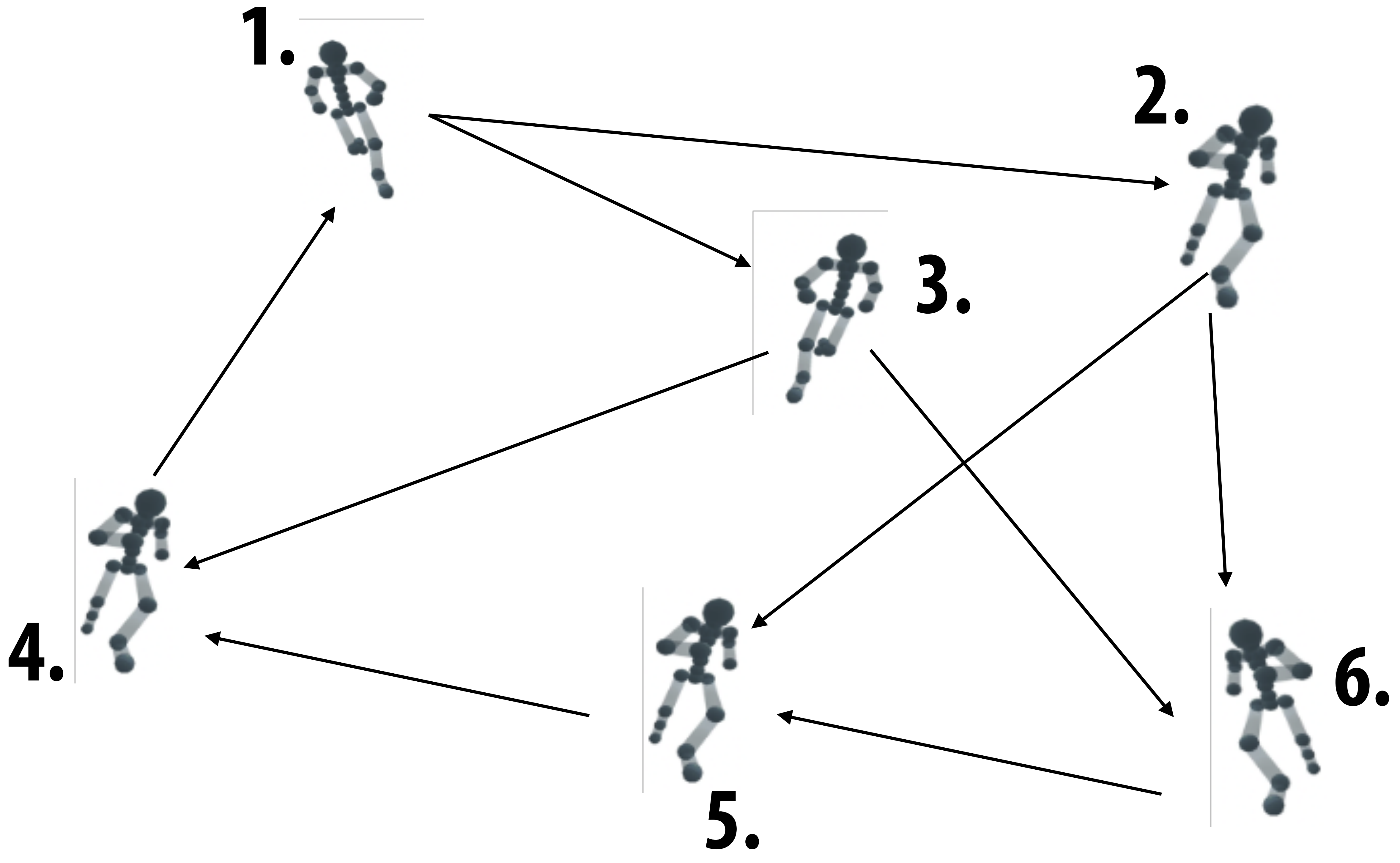
Motion graphs

- Each edge corresponds to a motion clip
- Each node corresponds to a choice point connecting motion clips



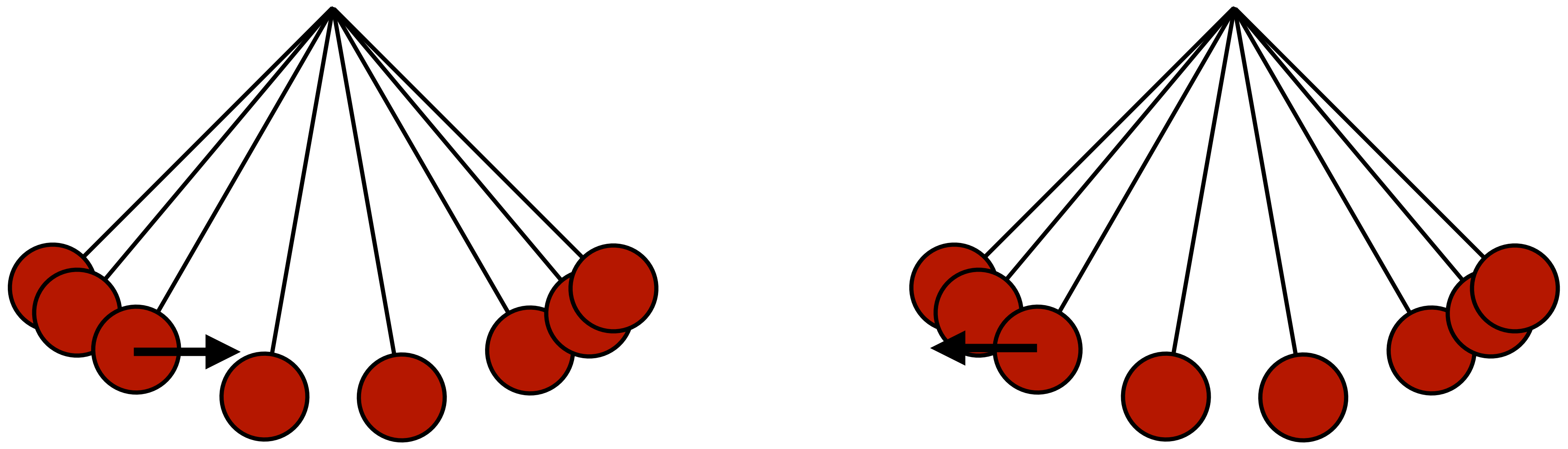
Motion graphs

- Each node corresponds to a motion clip
- Each edge means a smooth transition between clips



Similarity metrics

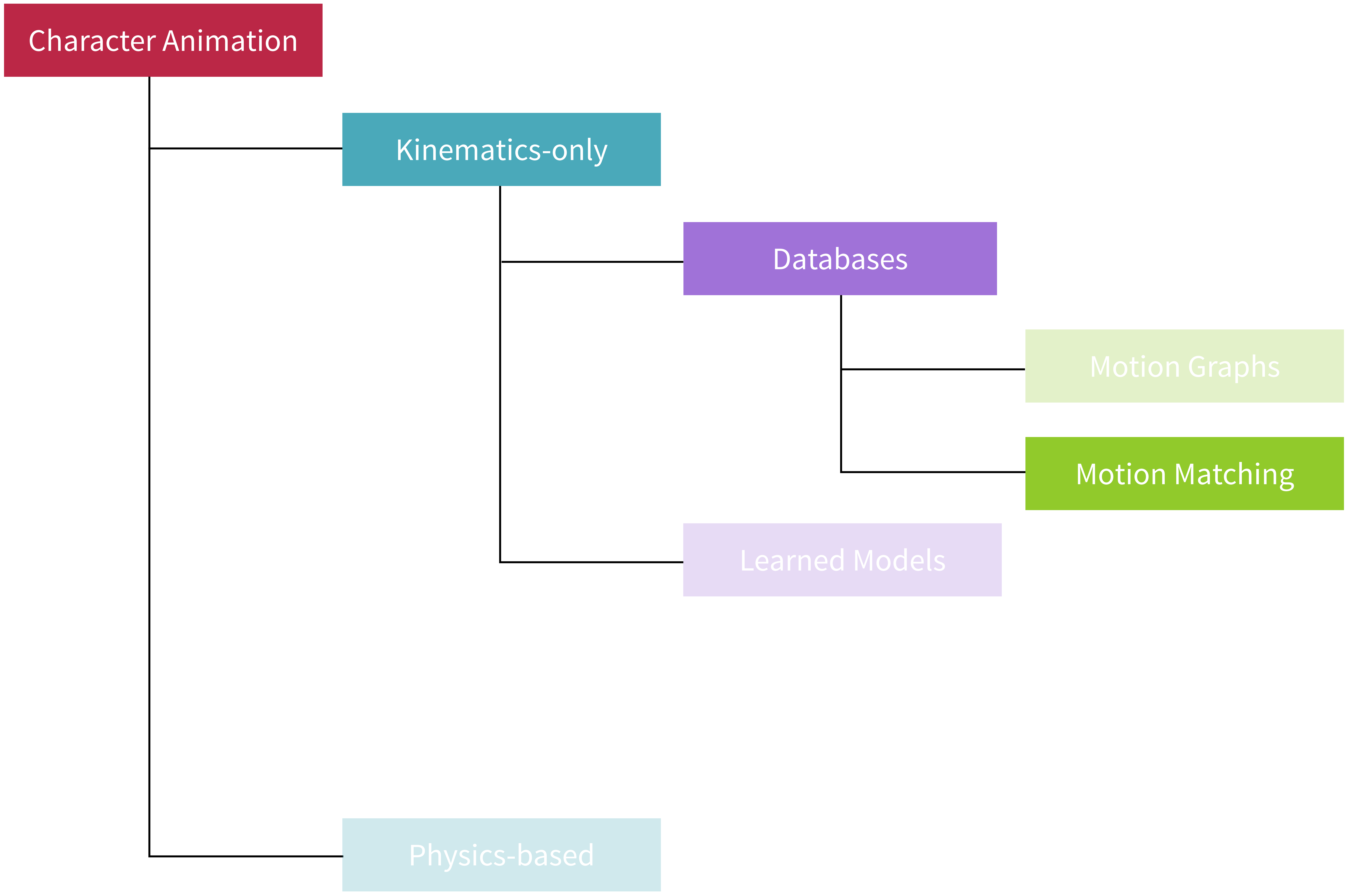
- **Transition is allowed when two motions are very similar**
  - **Naive similarity metric:** $\|\mathbf{q}_i - \mathbf{q}_j\|_2$
- **Parameters have different overall effects on the character**
  - **Position and orientation are in different units:** $(\mathbf{x}, \mathbf{R}, \theta, \phi, \sigma, \cdots)$
- **Velocity also needs to be considered:** $\|\dot{\mathbf{q}}_i - \dot{\mathbf{q}}_j\|_2$



Search on motion graphs

- Given constraints/goals at specific time instances on the graph, use a search algorithm to find paths connecting the constraints

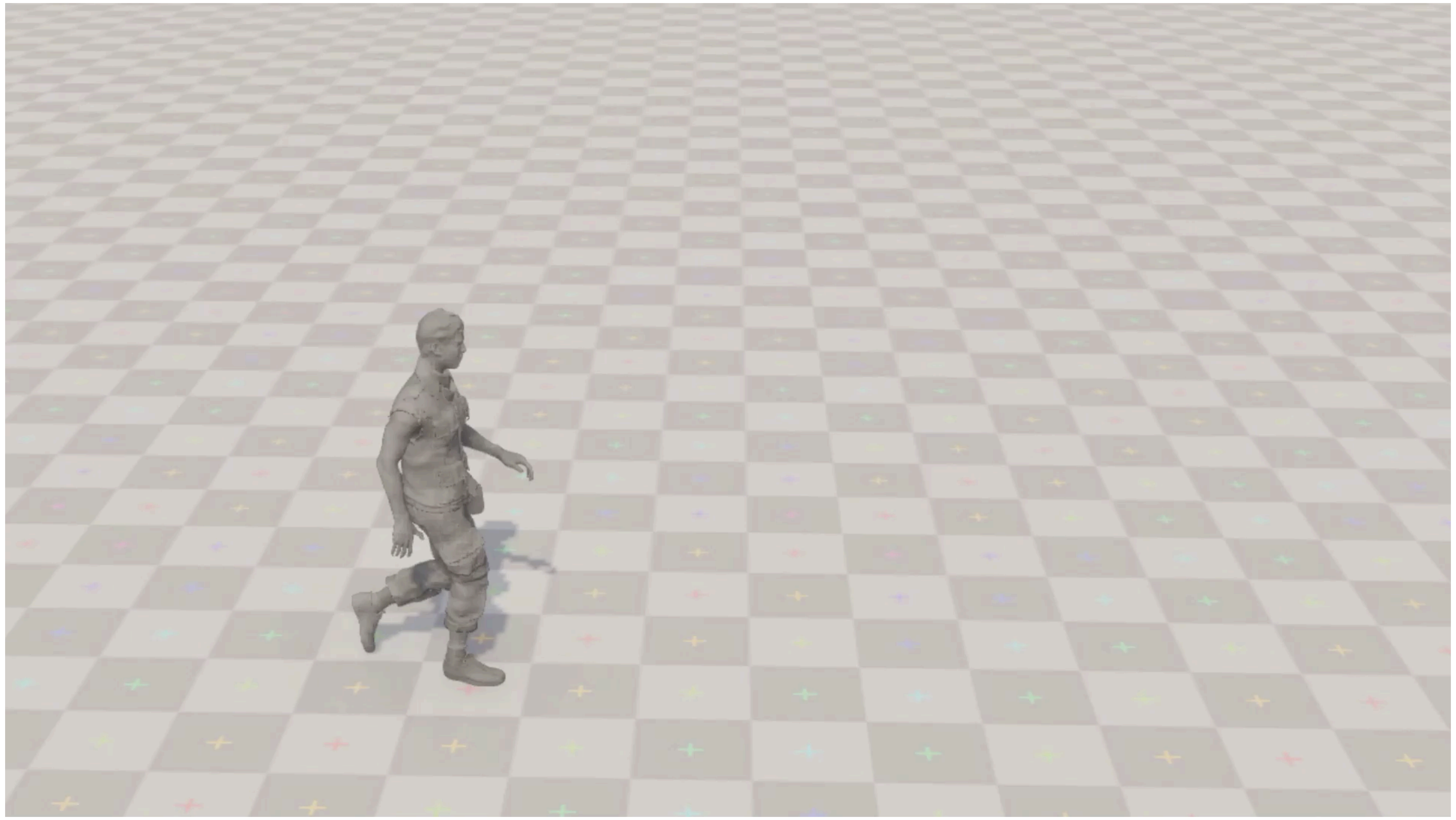




Character Animation
Kinematics-only
Databases
Motion Graphs
Motion Matching
Learned Models
Physics-based

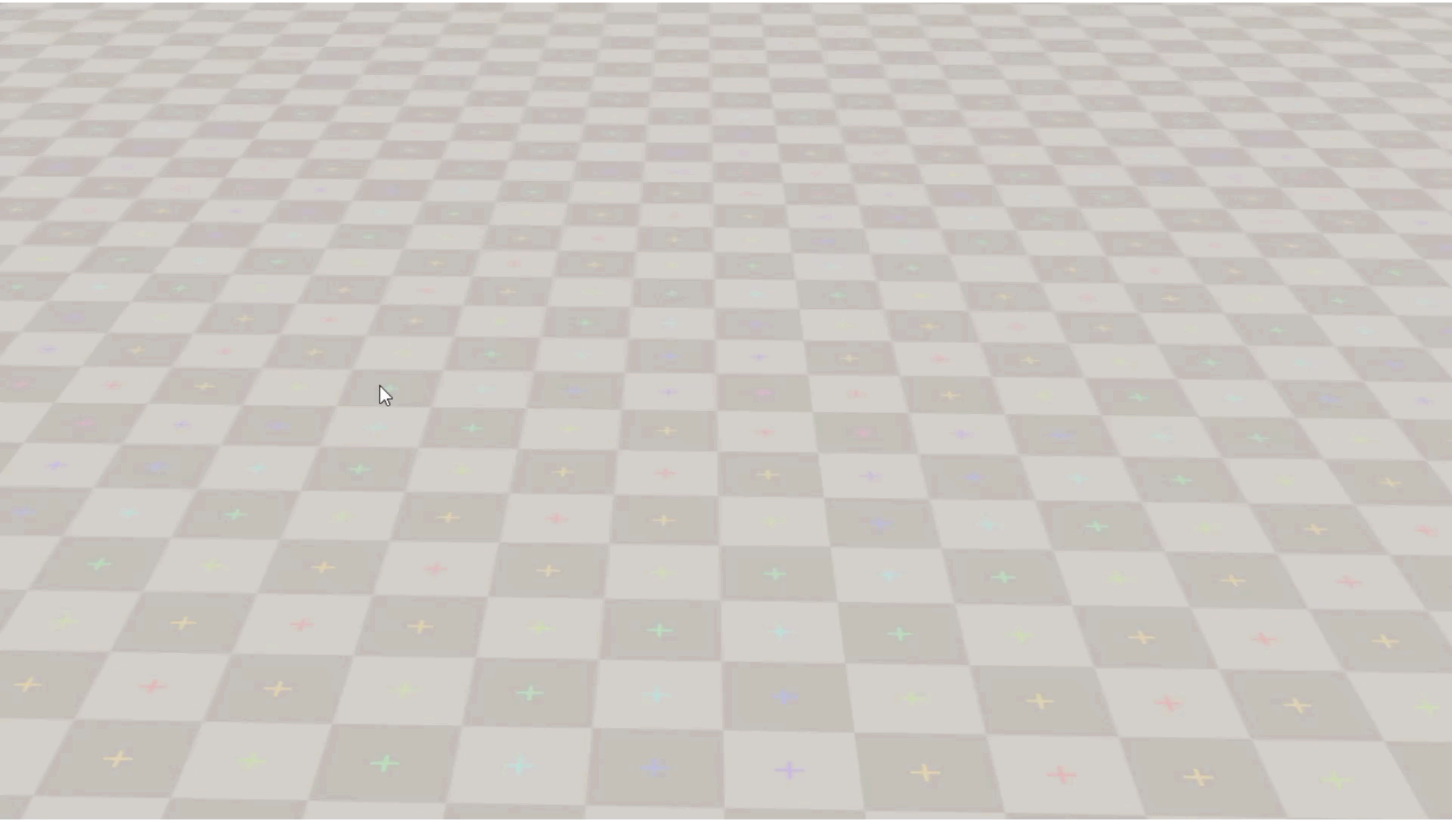
Motion matching

- **Which frame to play next, given the current frame and user constraints?**
- **How to make transitions smooth?**



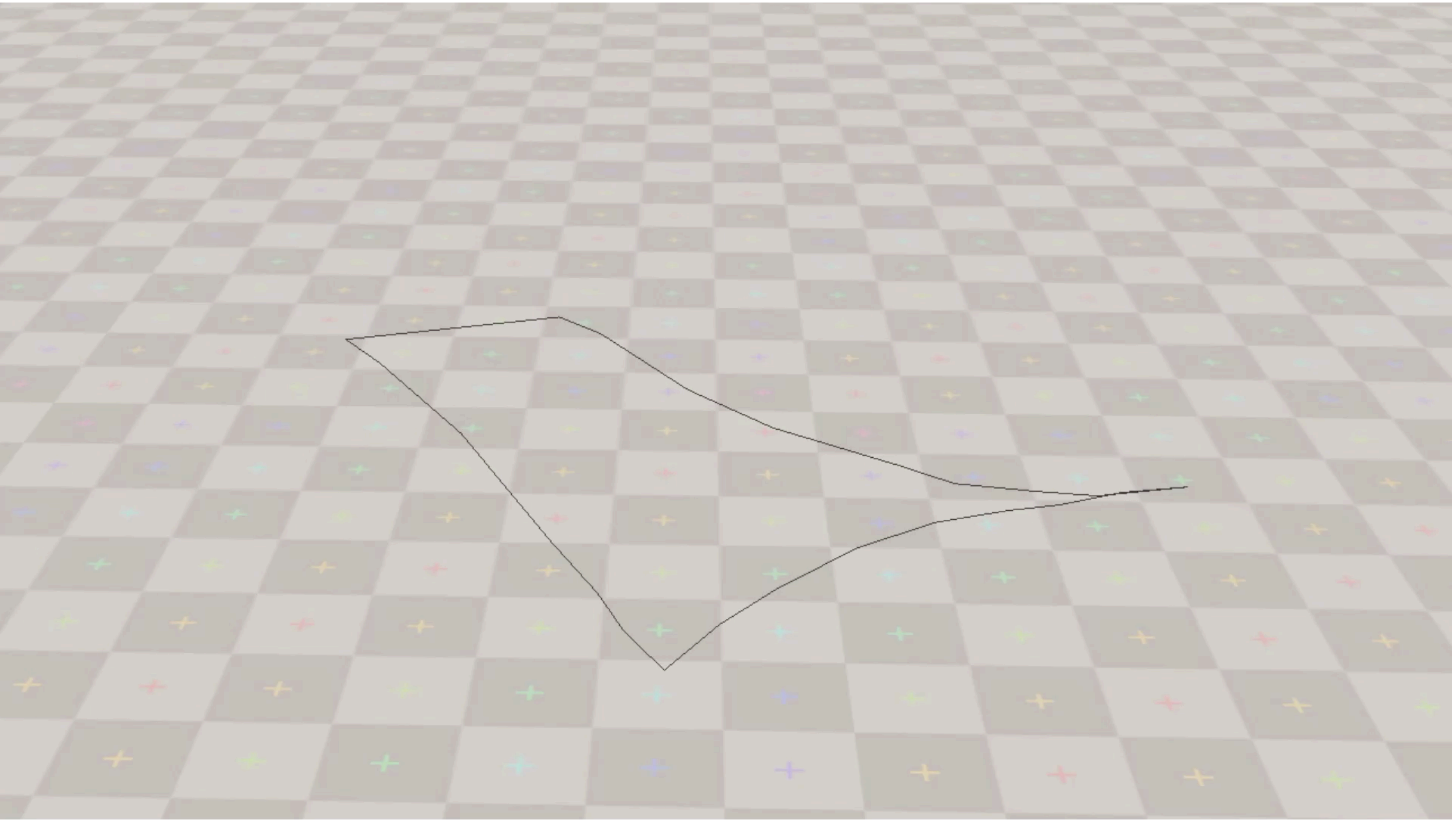
Motion Matching works best with lots of motion capture data.

Path following



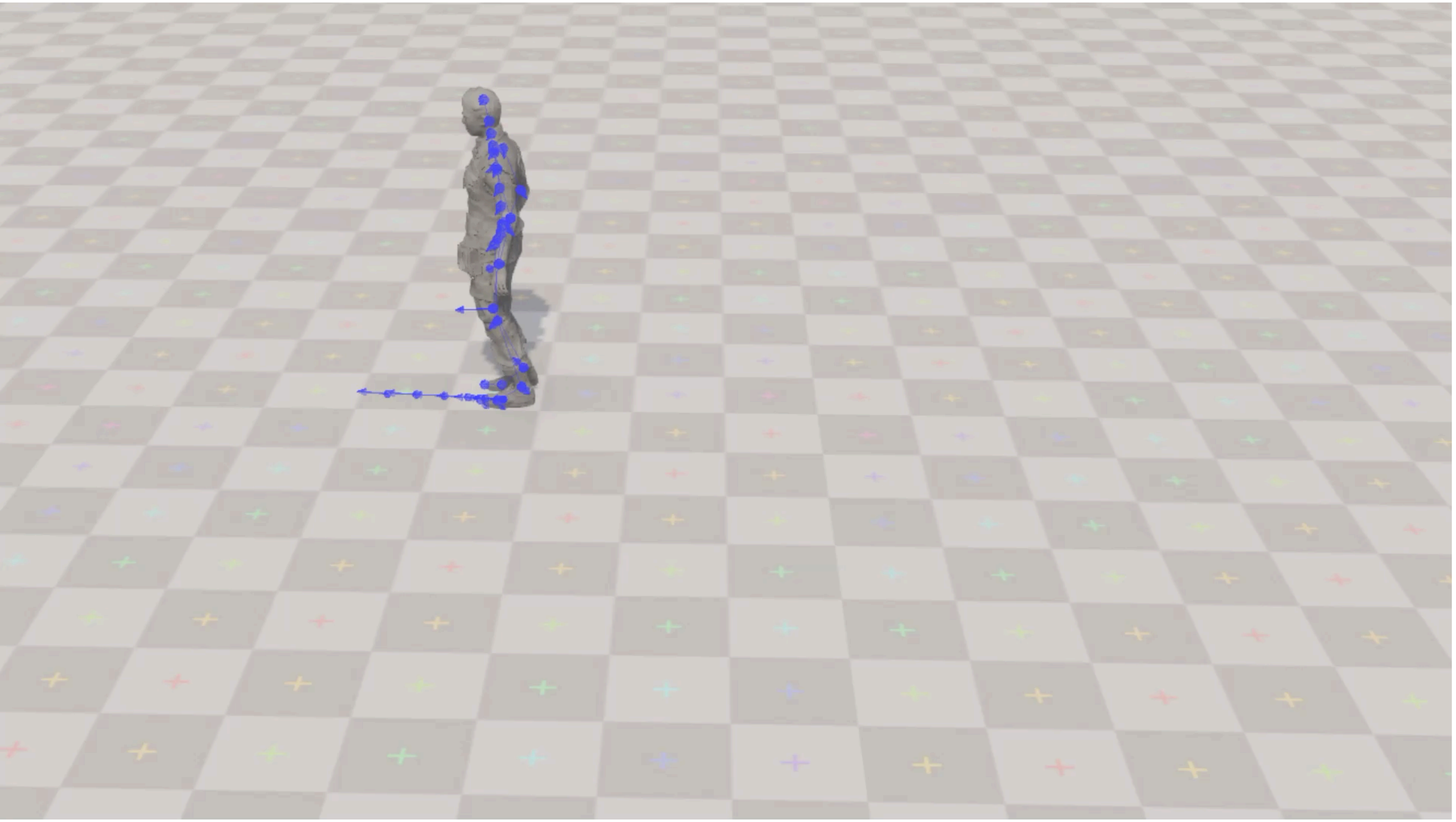
How to animate a character to make it follow this new path?

Patch motion clips



Repeatedly searching the dataset for a clip that, if played from the current location, would do a better job of keeping the character on the path than the current clip.

When to switch?

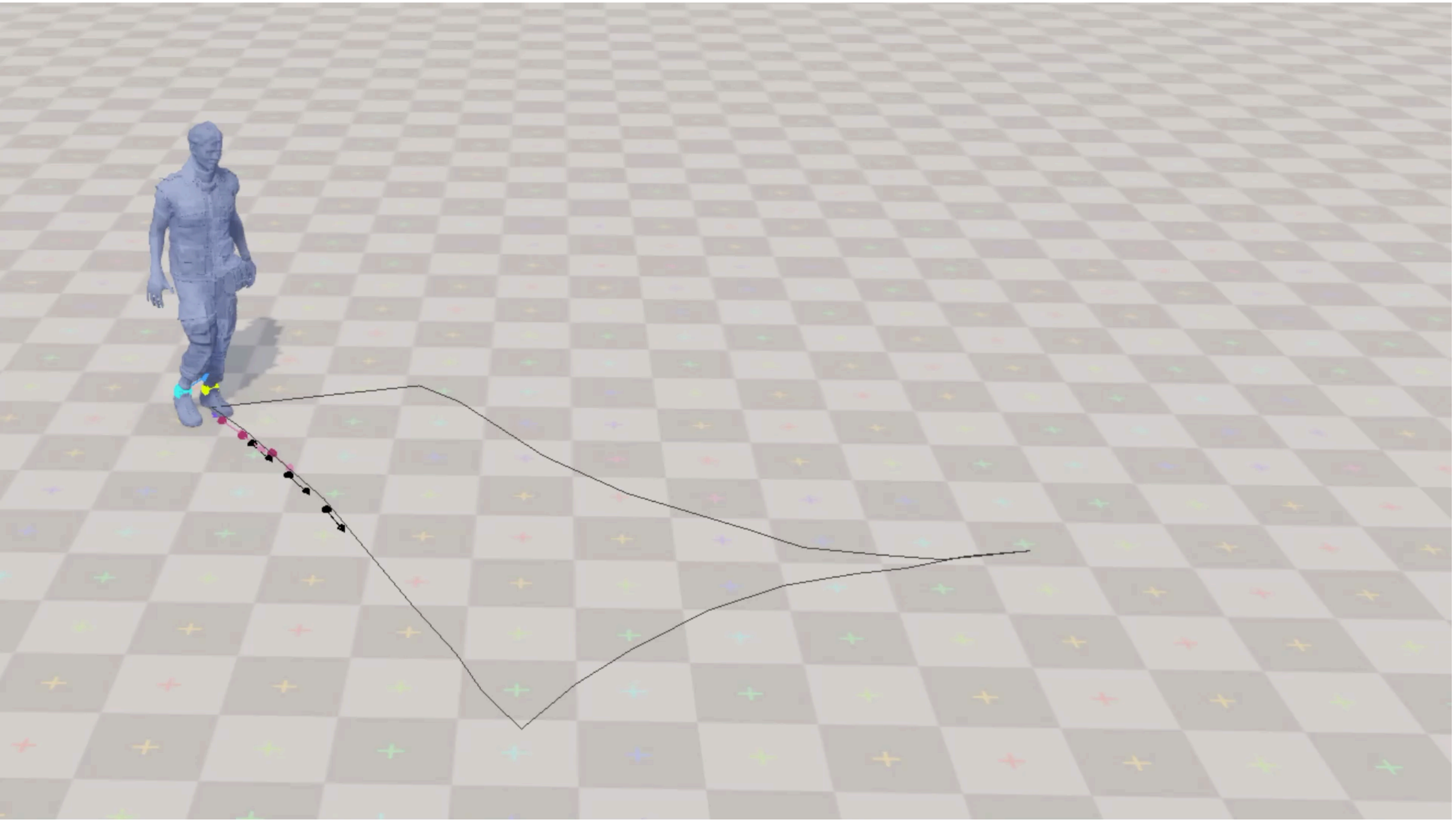


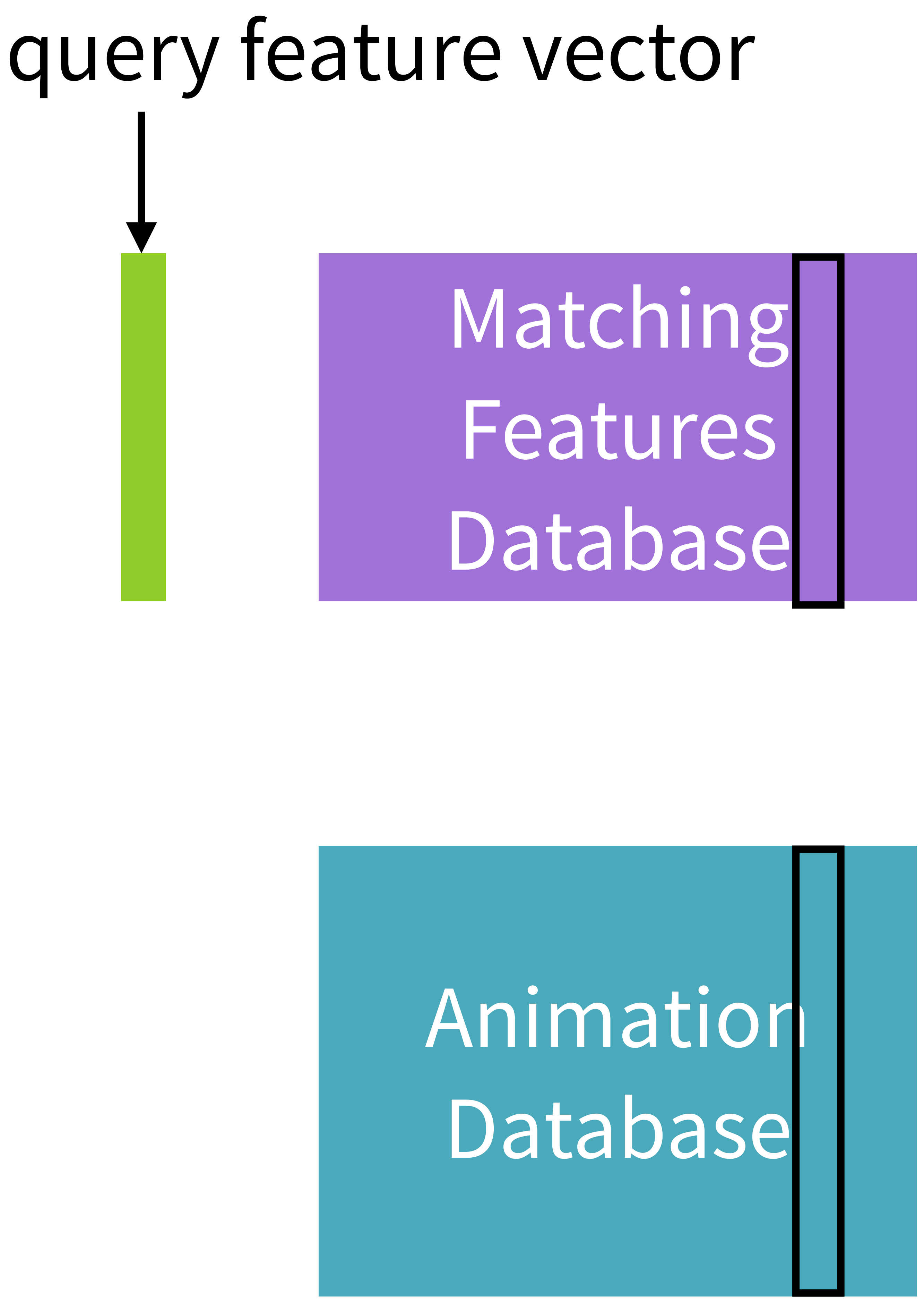
A single frame of animation contains a lot of information – and some of the information might not be useful for this decision.

Feature selection

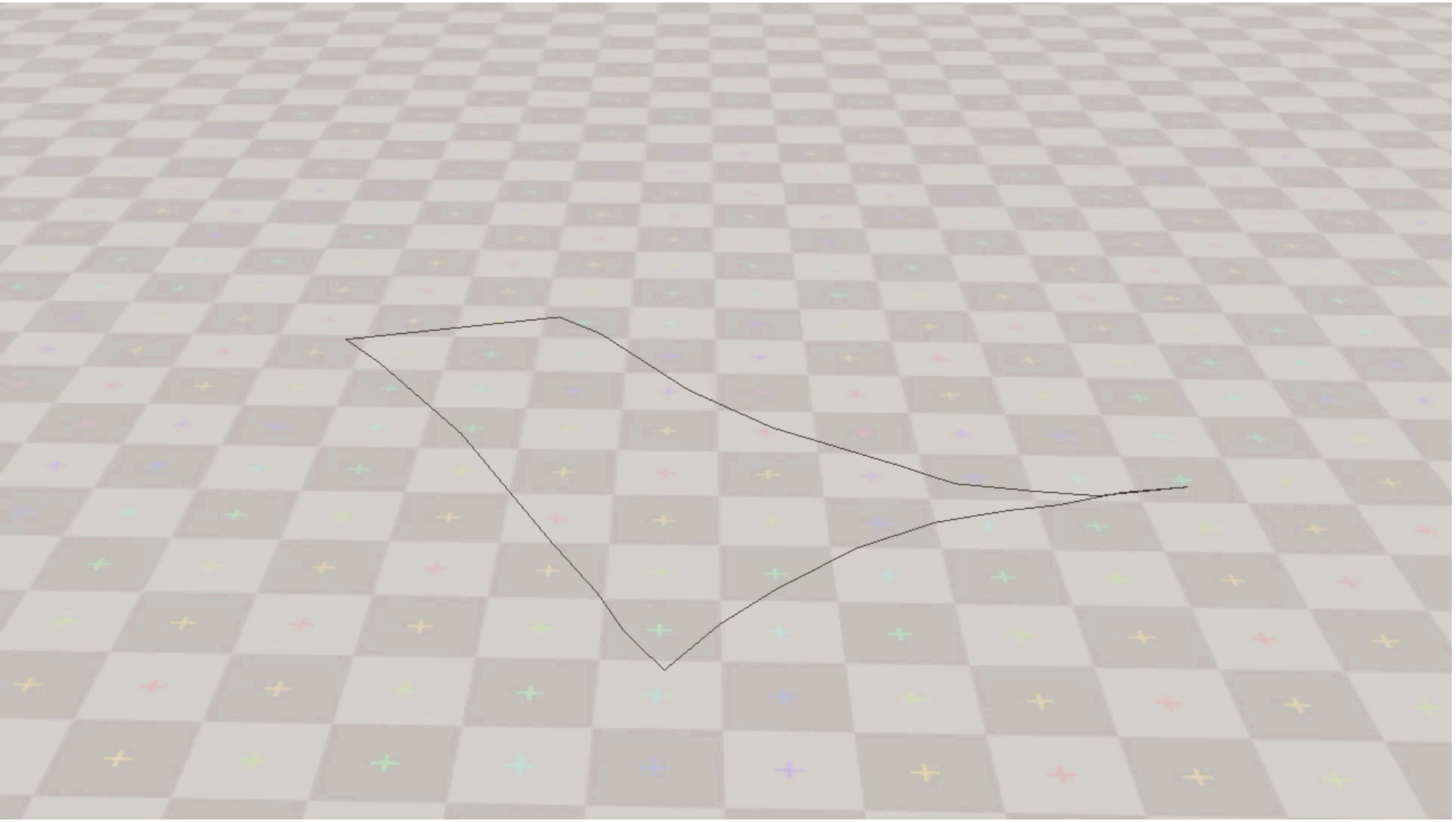
For this task, feet positions, velocities, the character root’s velocity, and a couple of snapshots from the future trajectory position and direction are the only features we need.

Feature matching





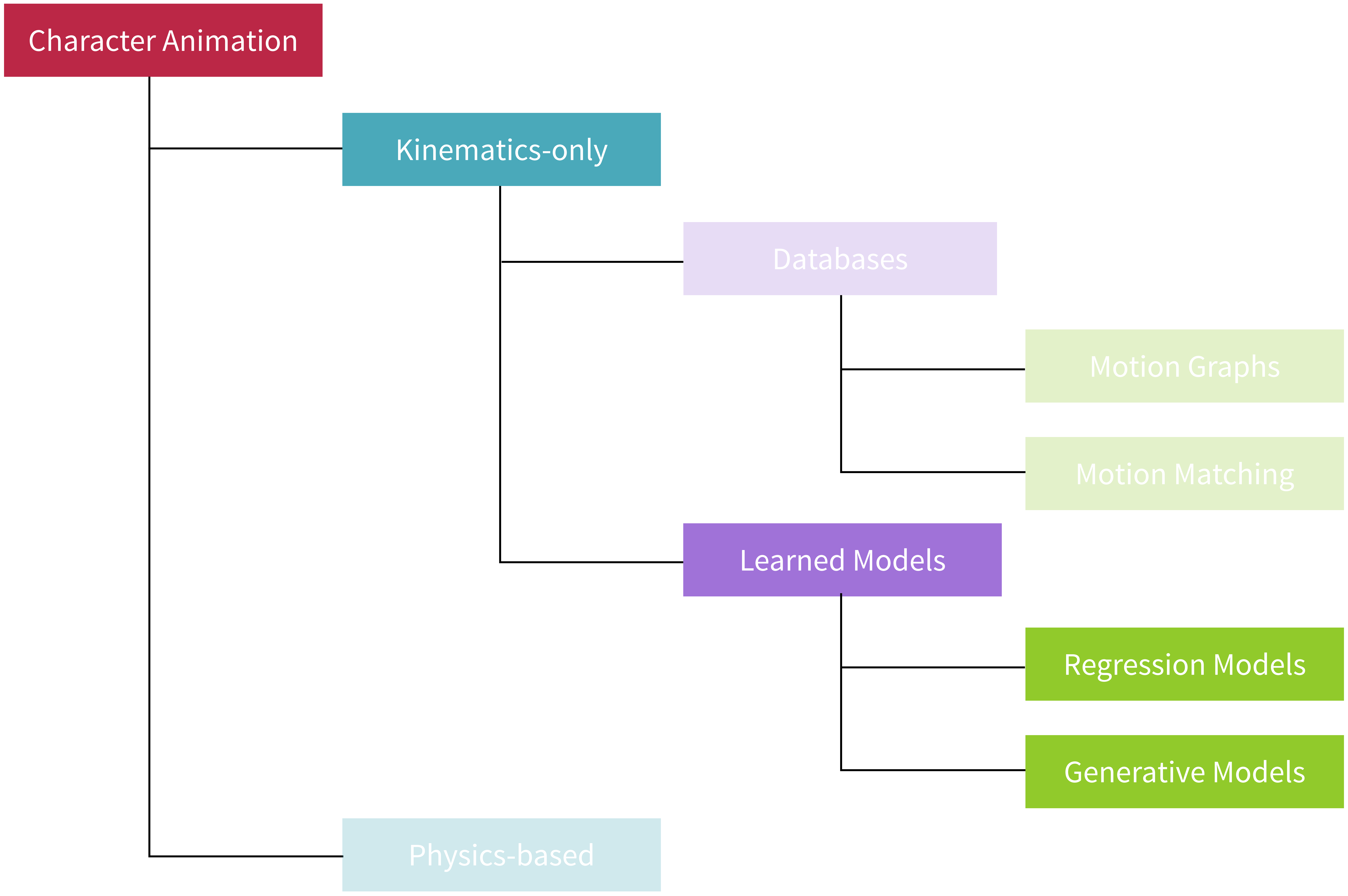
Post processing



If the change in pose is not too large, this discontinuity can be easily removed using common techniques.

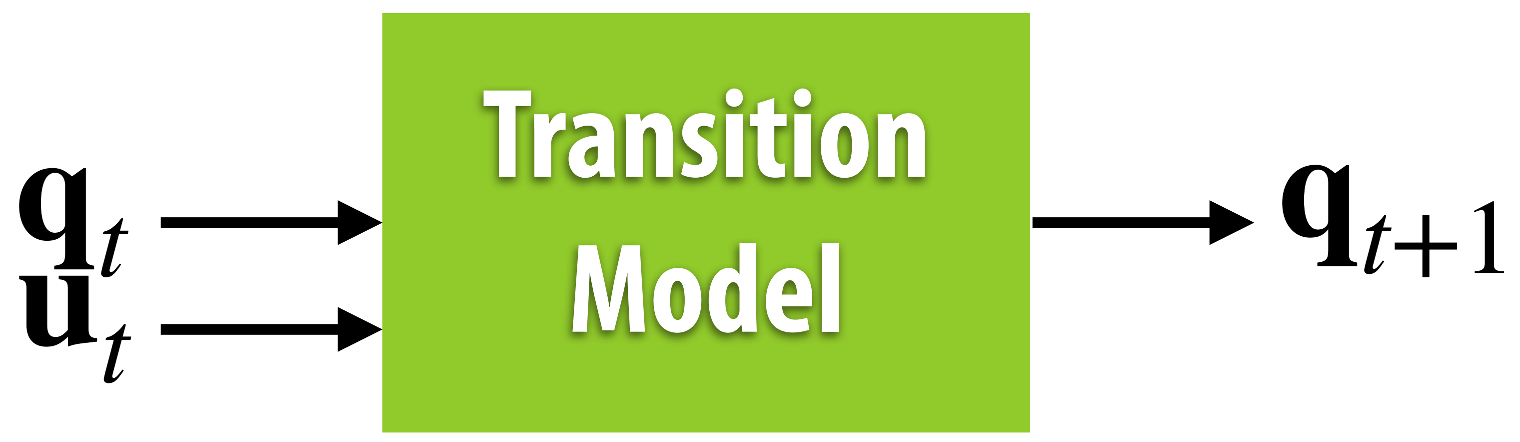

L2
R2
Y
X
B
A
L3
R3
Holden et al SIGGRAPH 2020


L2
R2
Y
X
B
A
L3
R3
Holden et al SIGGRAPH 2020


Character Animation
Kinematics-only
Databases
Motion Graphs
Motion Matching
Learned Models
Regression Models
Generative Models
Physics-based

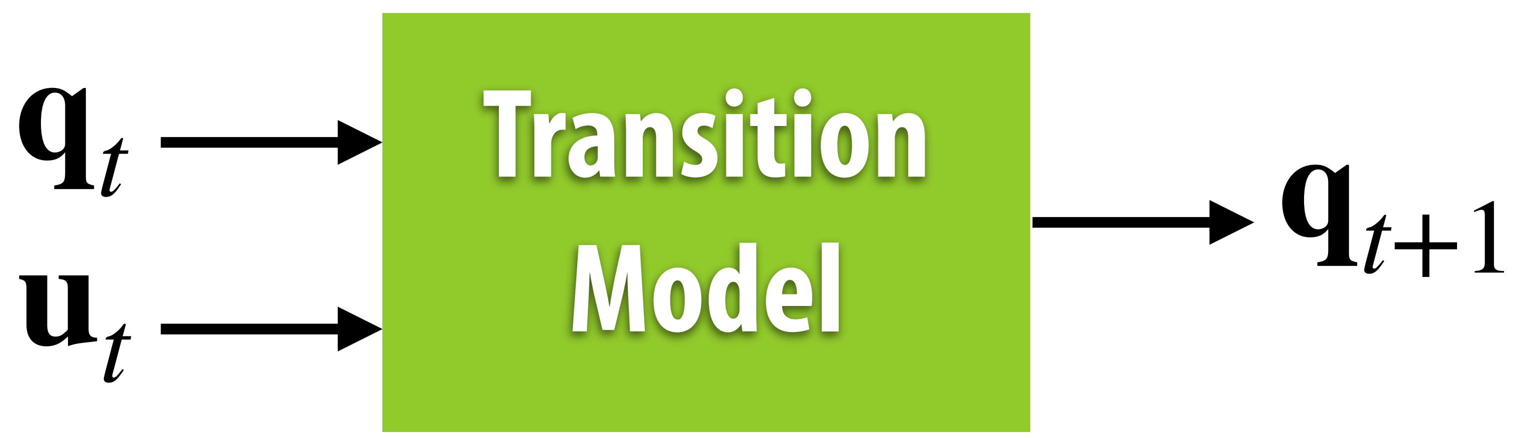
Predicting the next frame

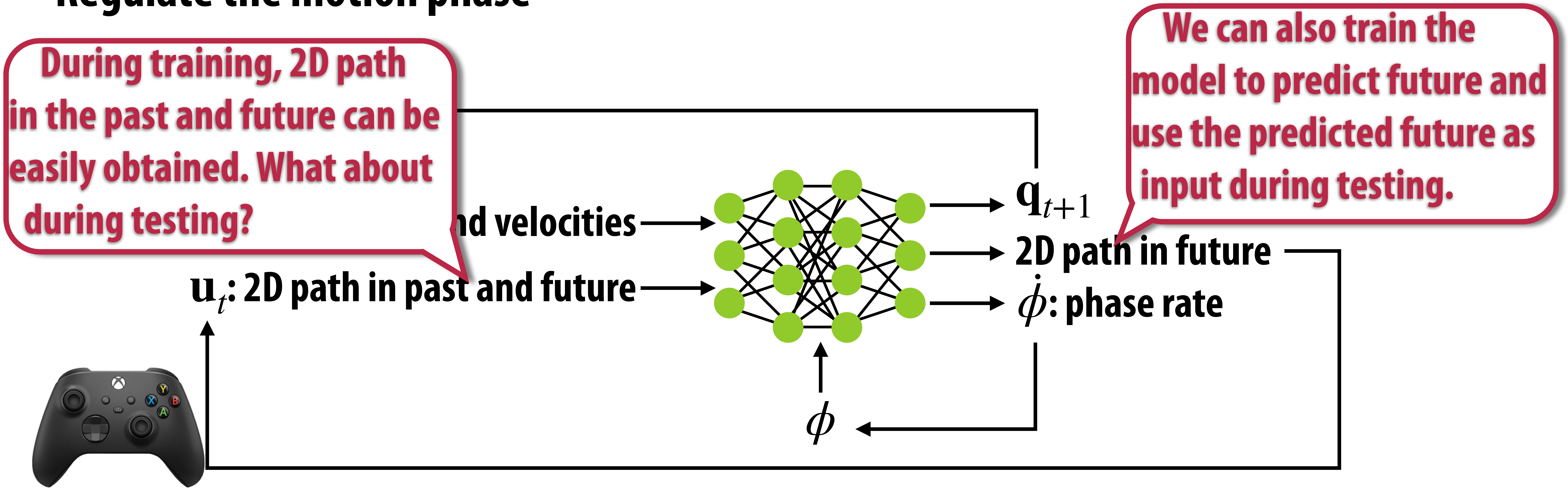
- Given sequences of human poses, $\mathbf{q}_0, \mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_T$, can we learn a transition model from data in a supervised fashion?
- Given a pose $\mathbf{q}_t$, there are many plausible next pose $\mathbf{q}_{t+1}$ — the model needs information more than just the current pose.
- Define a control/intent vector $\mathbf{u}_t$ to capture additional information necessary to disambiguate the next frame.



Predicting the next frame

- Use a neural network to represent the model
- Define $\mathbf{q}$: 3D joint positions and velocities
- Secret sauce
  - Define $\mathbf{u}$ as 2D path in past and future and learn to predict future path
  - Regulate the motion phase

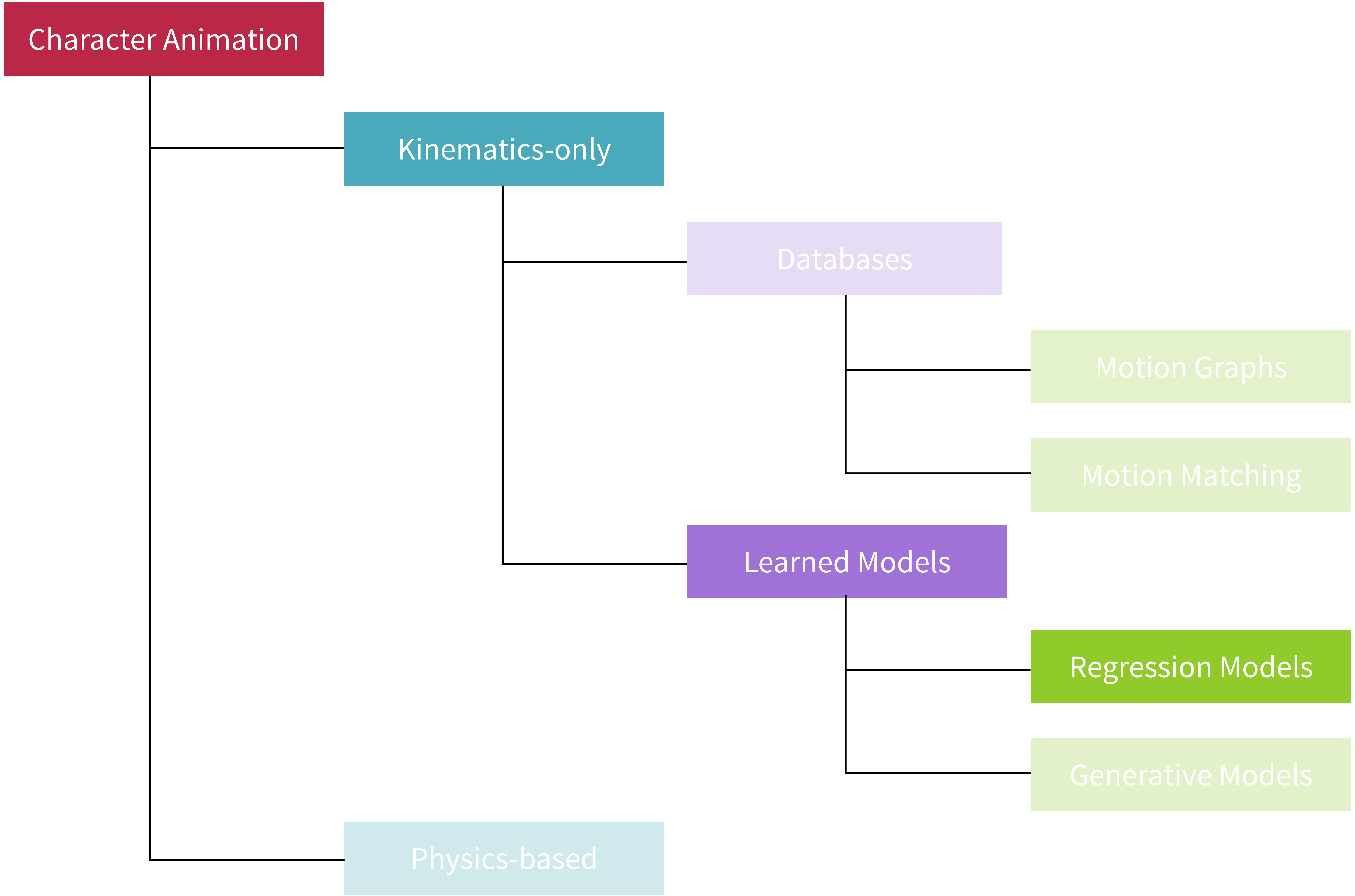




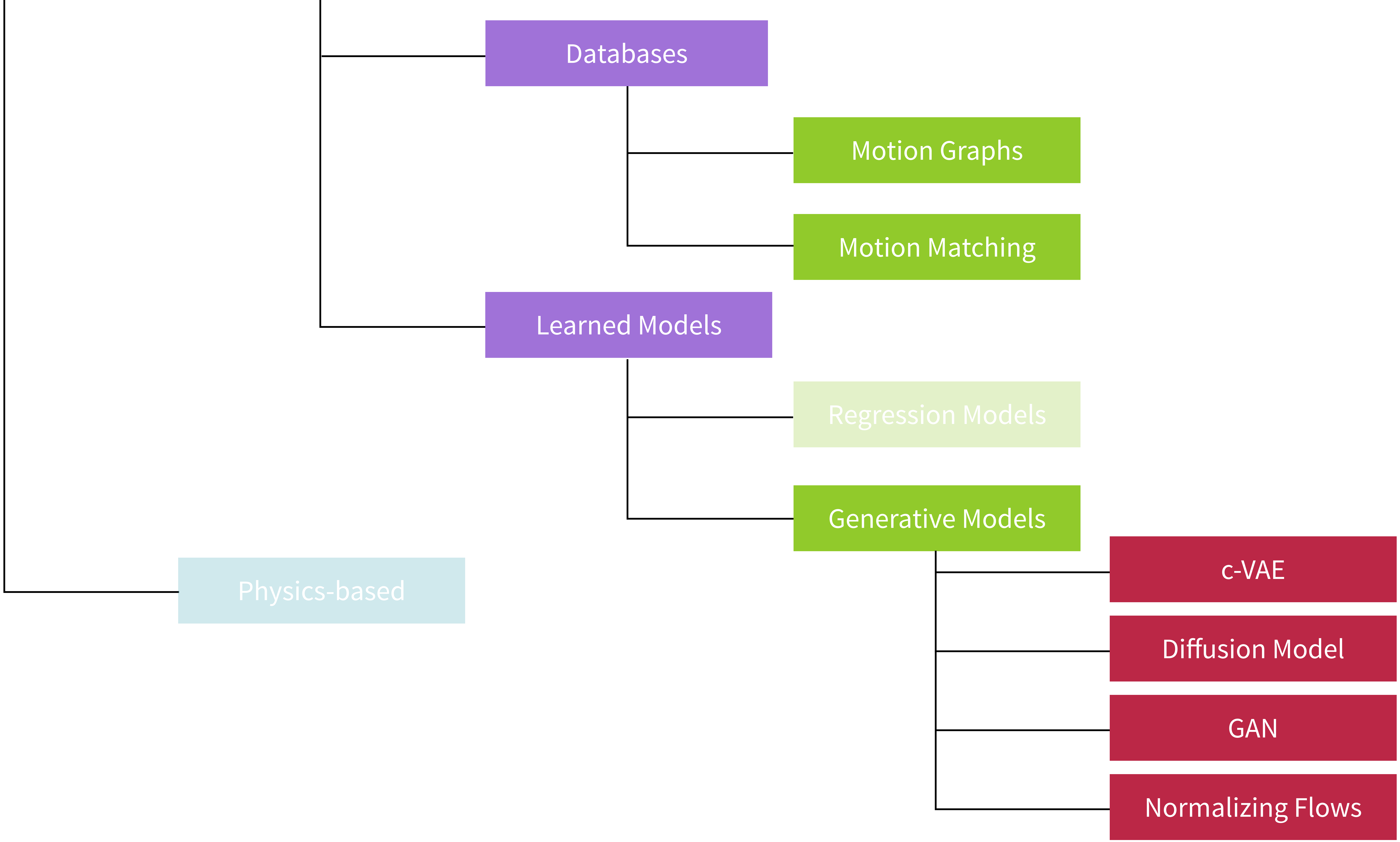
Holden et al SIGGRAPH 2019

Neural motion synthesizer

Starke et al SIGGRAPH Asia 2019

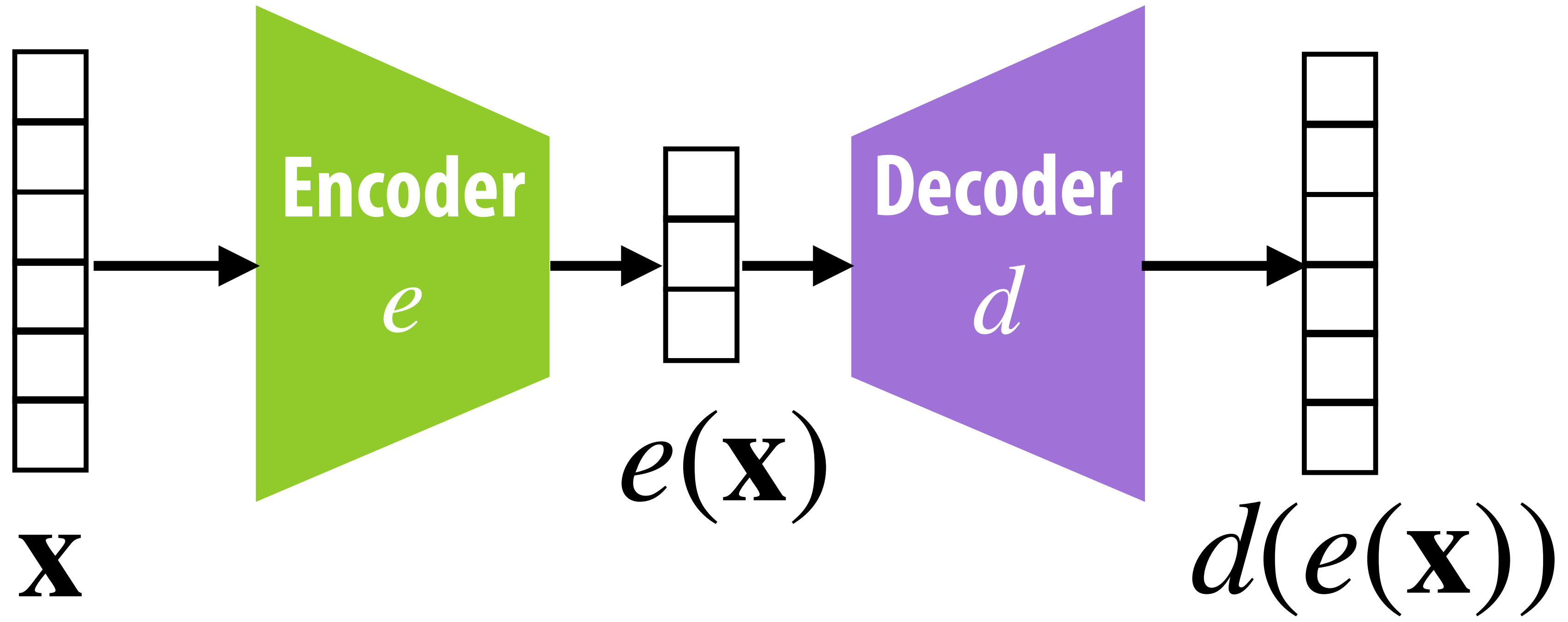


Character Animation
Kinematics-only
Databases
Motion Graphs
Motion Matching
Learned Models
Regression Models
Generative Models
Physics-based


Databases
Motion Graphs
Motion Matching
Learned Models
Regression Models
Generative Models
c-VAE
Diffusion Model
GAN
Normalizing Flows
Physics-based

Generative models for human motion

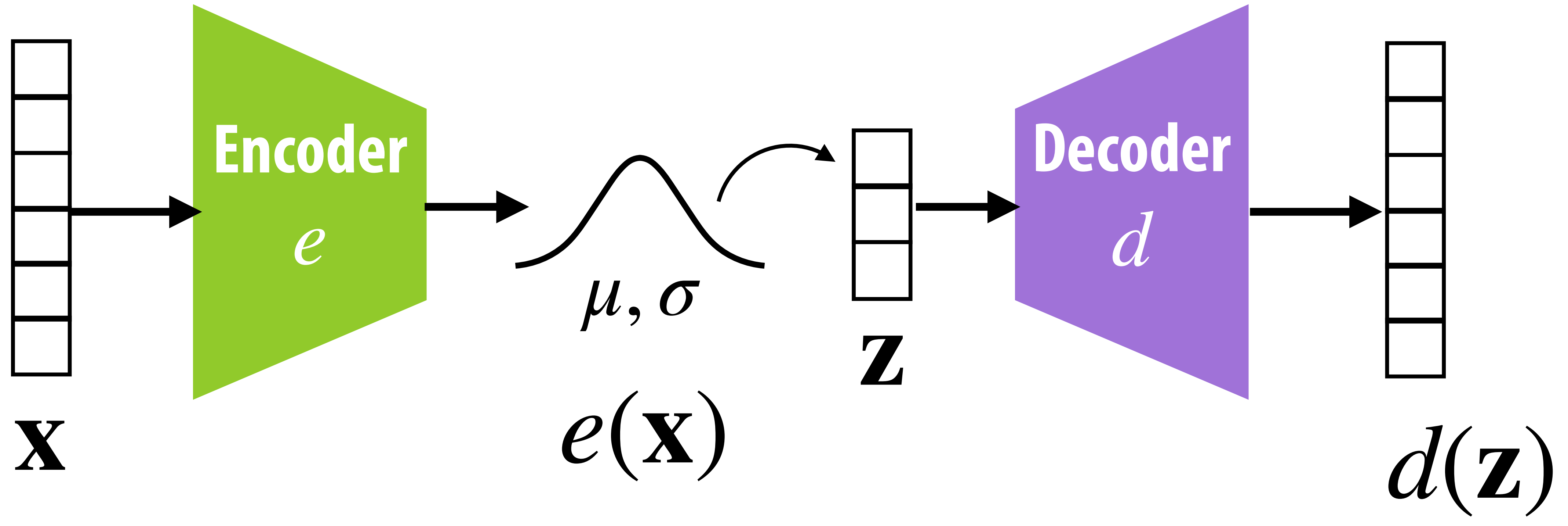
- What is an Autoencoder?



$$\text{loss} = ||\mathbf{x} - d(e(\mathbf{x}))||^2$$

Generative models for human motion

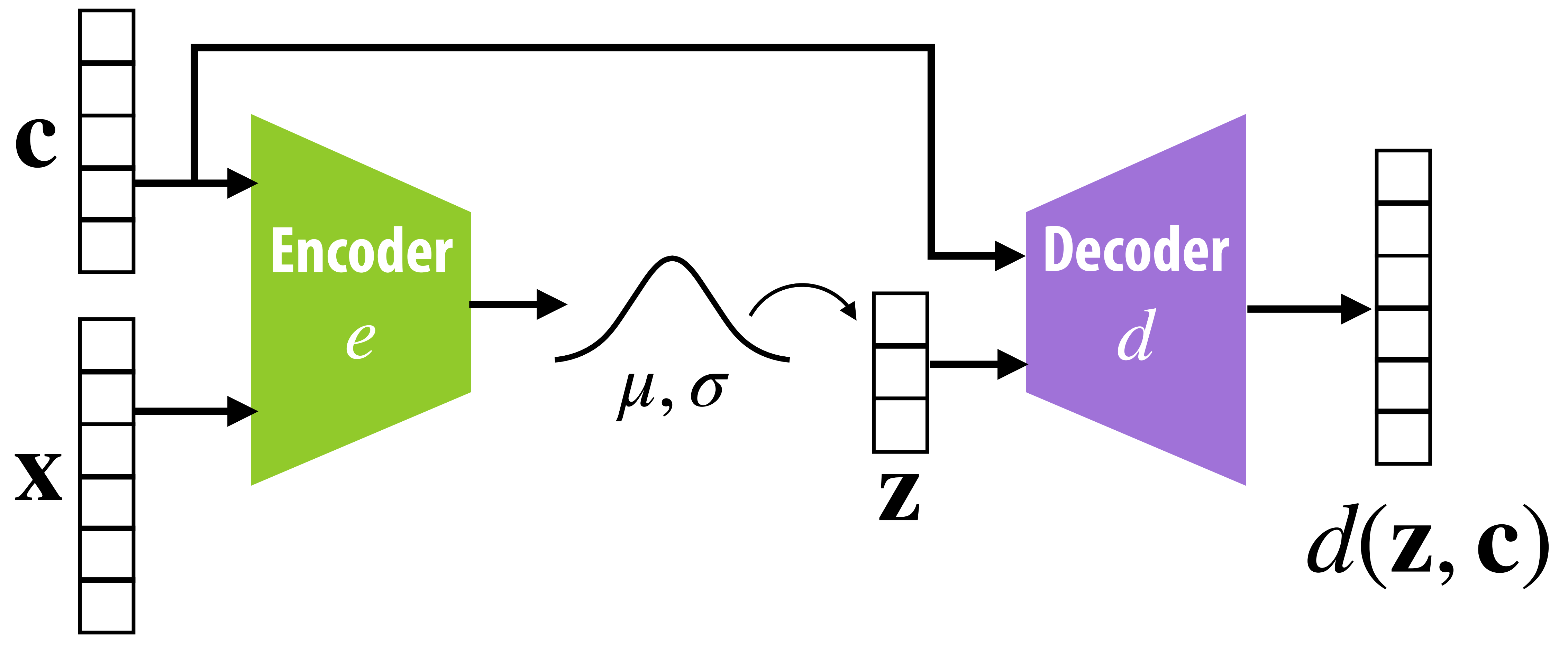
- What is an Autoencoder?
- What is a Variational Autoencoder?



loss $= \|\mathbf{x} - d(\mathbf{z})\|^2 + KL(N(\mu, \sigma), N(0, \mathbf{I}))$

Generative models for human motion

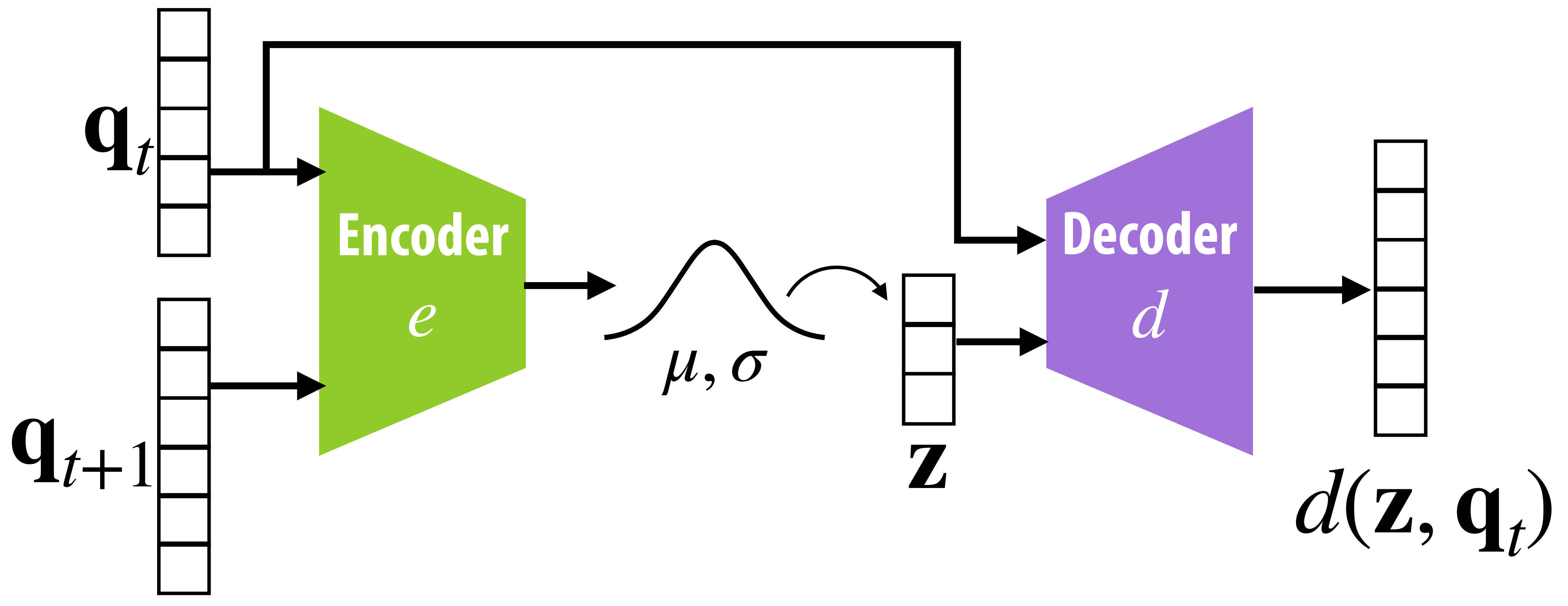
- What is an Autoencoder?
- What is a Variational Autoencoder (VAE)?
- What is a Conditional Variational Autoencoder (CVAE)?



$$\text{loss} = \|\mathbf{x} - d(\mathbf{z}, \mathbf{c})\|^2 + KL(N(\mu, \sigma), N(0,\mathbf{I}))$$

Generative models for human motion

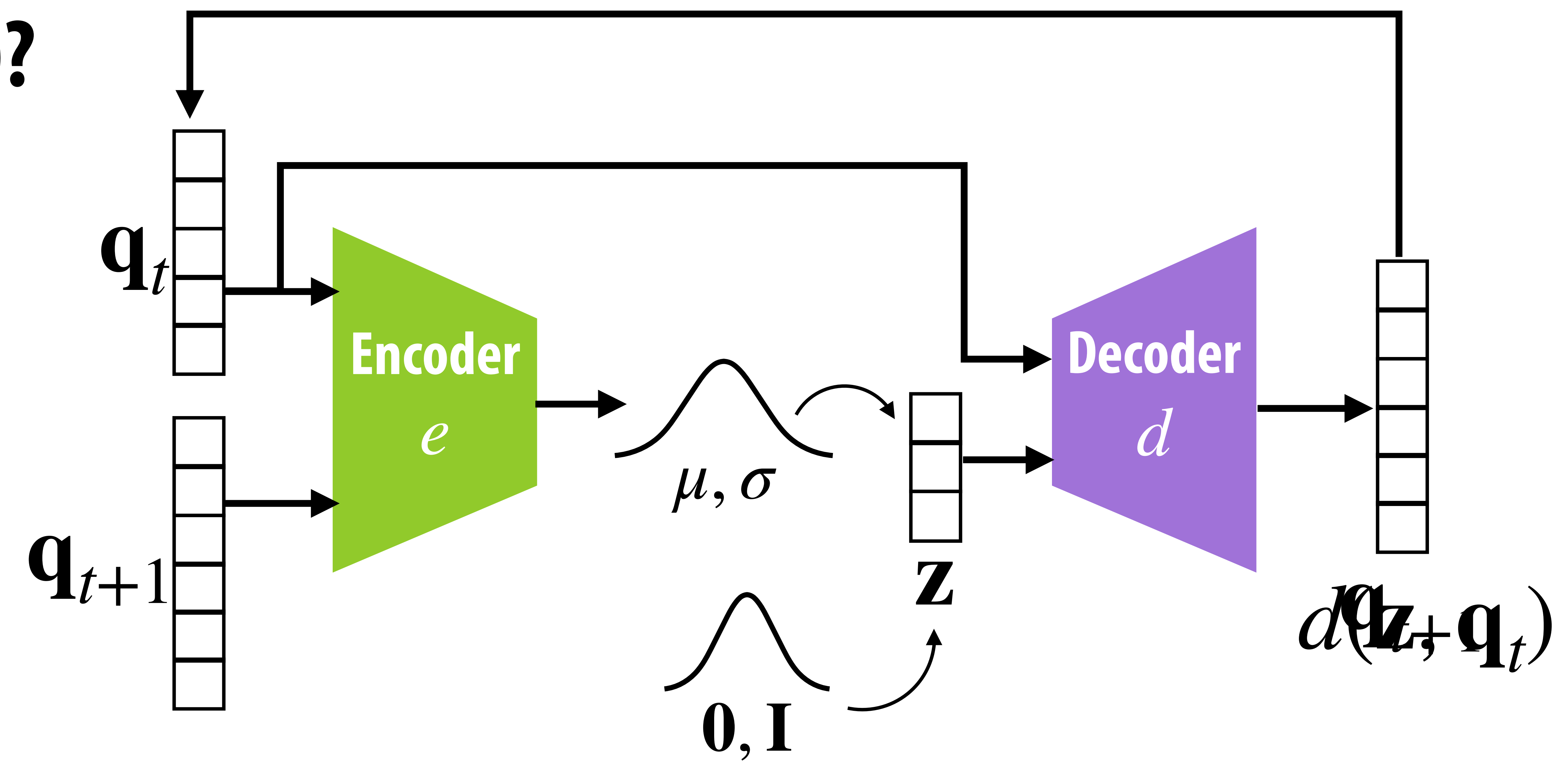
- What is an Autoencoder?
- What is a Variational Autoencoder (VAE)?
- What is a Conditional Variational Autoencoder (CVAE)?
- How to use CVAE to generate motion?
  - Define $\mathbf{x}$: next pose $\mathbf{q}_{t+1}$
  - Define $\mathbf{c}$: current pose $\mathbf{q}_t$

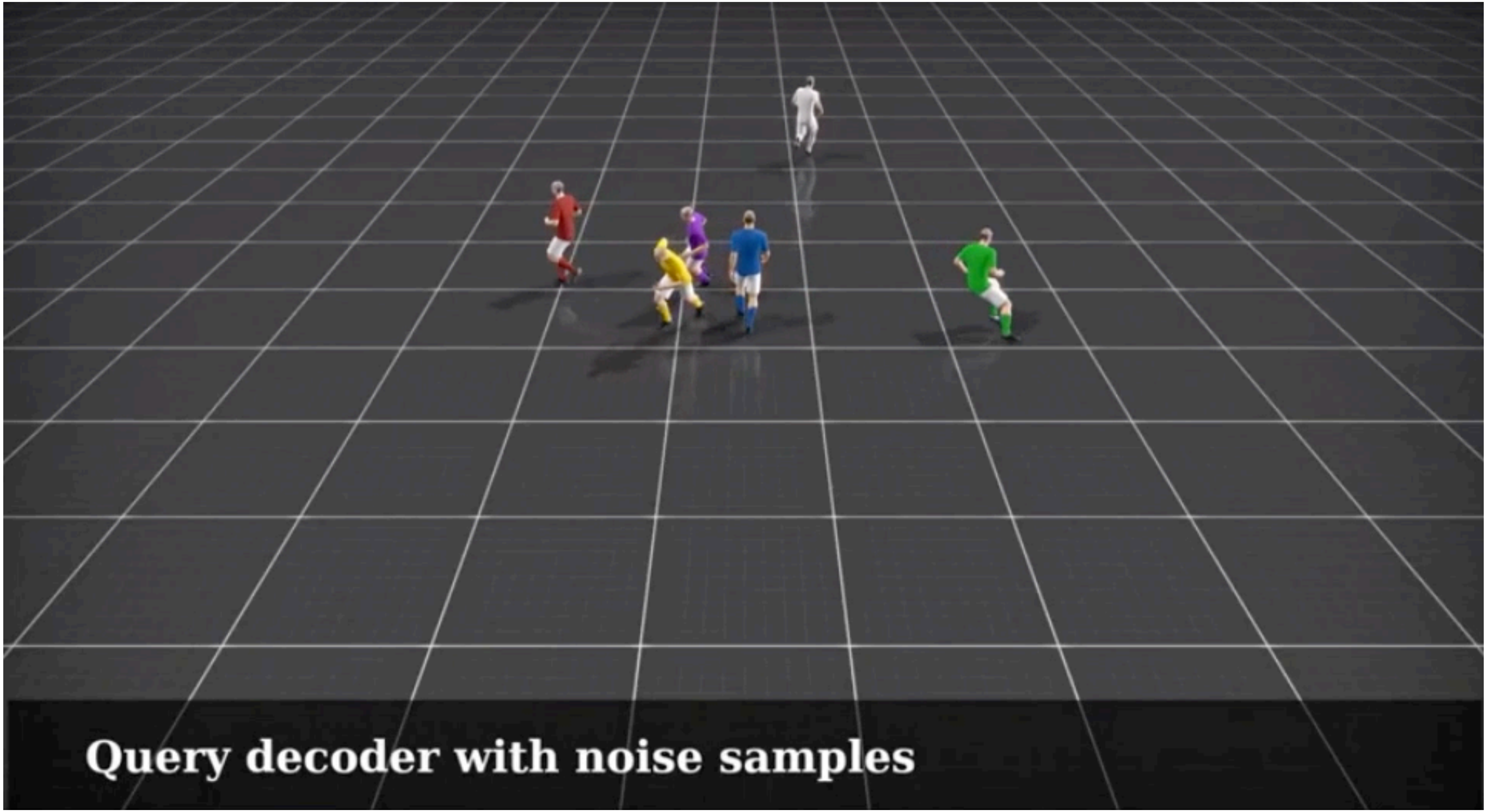


$$\text{loss} = \|\mathbf{q}_{t+1} - d(\mathbf{z}, \mathbf{q}_t)\|^2 + KL(N(\mu, \sigma), N(0, \mathbf{I}))$$

Generative models for human motion

- What is an Autoencoder?
- What is a Variational Autoencoder (VAE)?
- What is a Conditional Variational Autoencoder (CVAE)?
- How to use CVAE to generate motion?
  - Define $\mathbf{x}$: next pose $\mathbf{q}_{t+1}$
  - Define $\mathbf{c}$: current pose $\mathbf{q}_t$
- Once CVAE is trained, how to use it?
  - Discard the encoder and only use the decoder to generate next frame autoregressively
  - Sample $\mathbf{z}$ from a uniform distribution

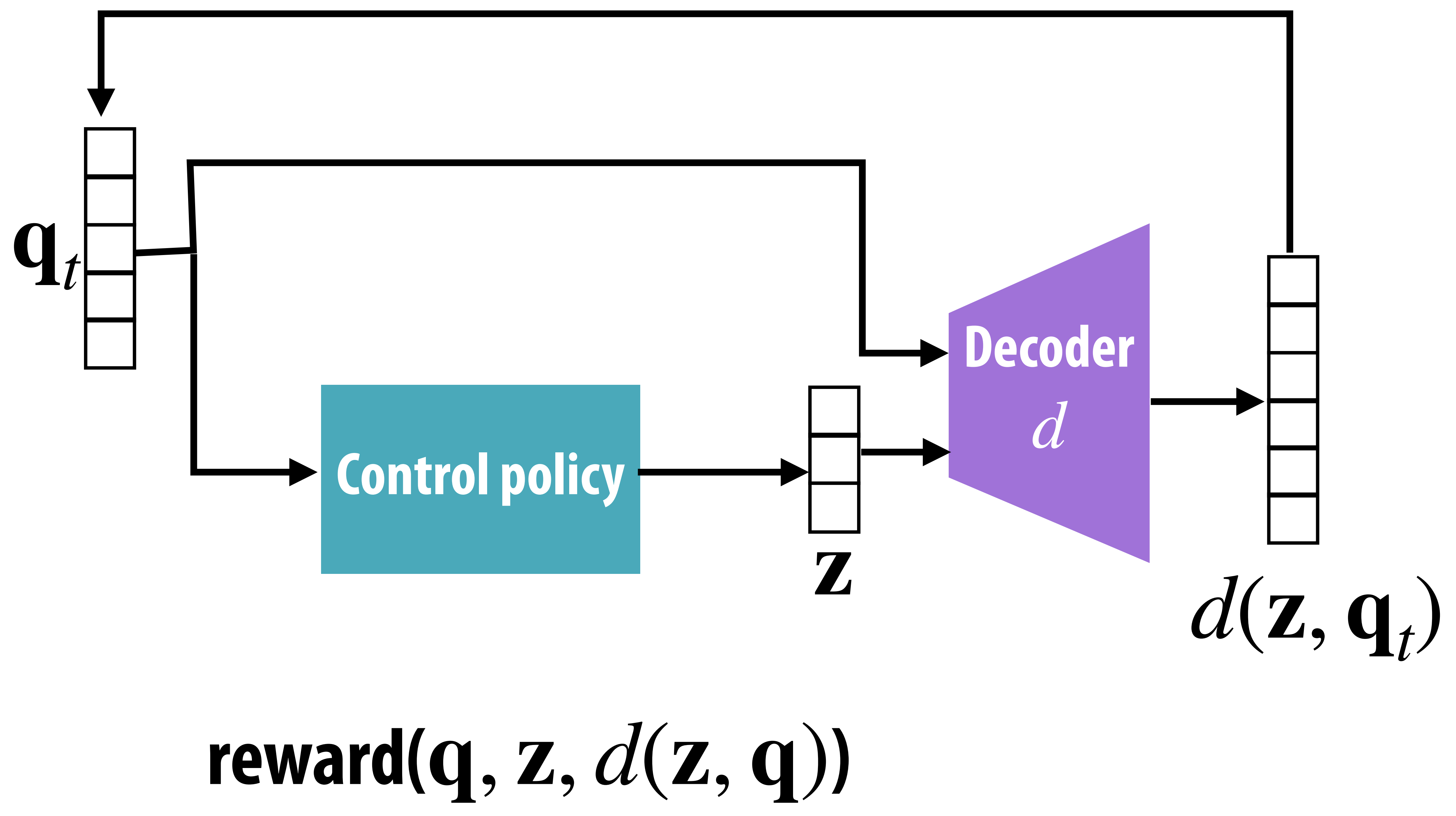


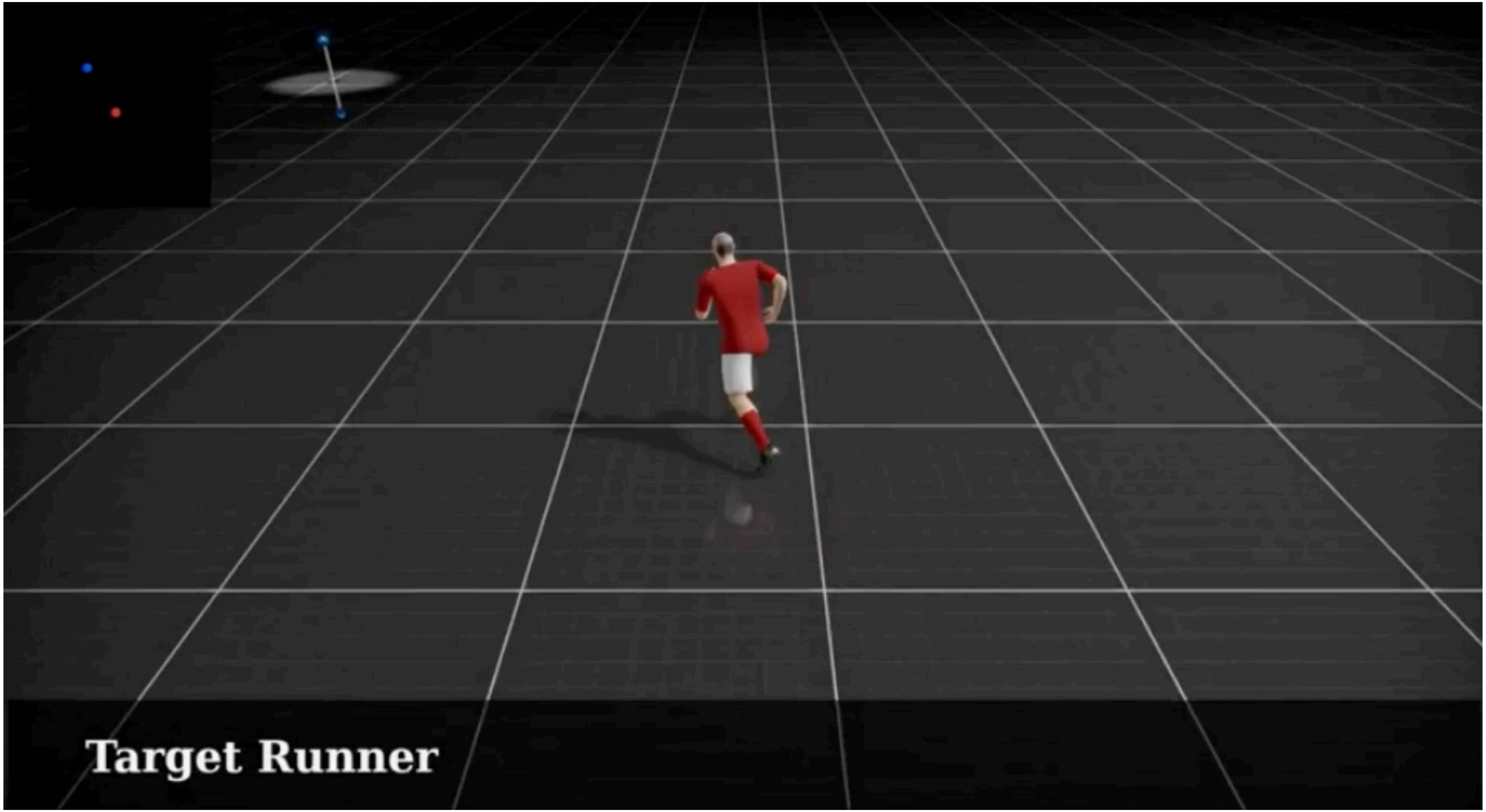


Ling et al SIGGRAPH 2021

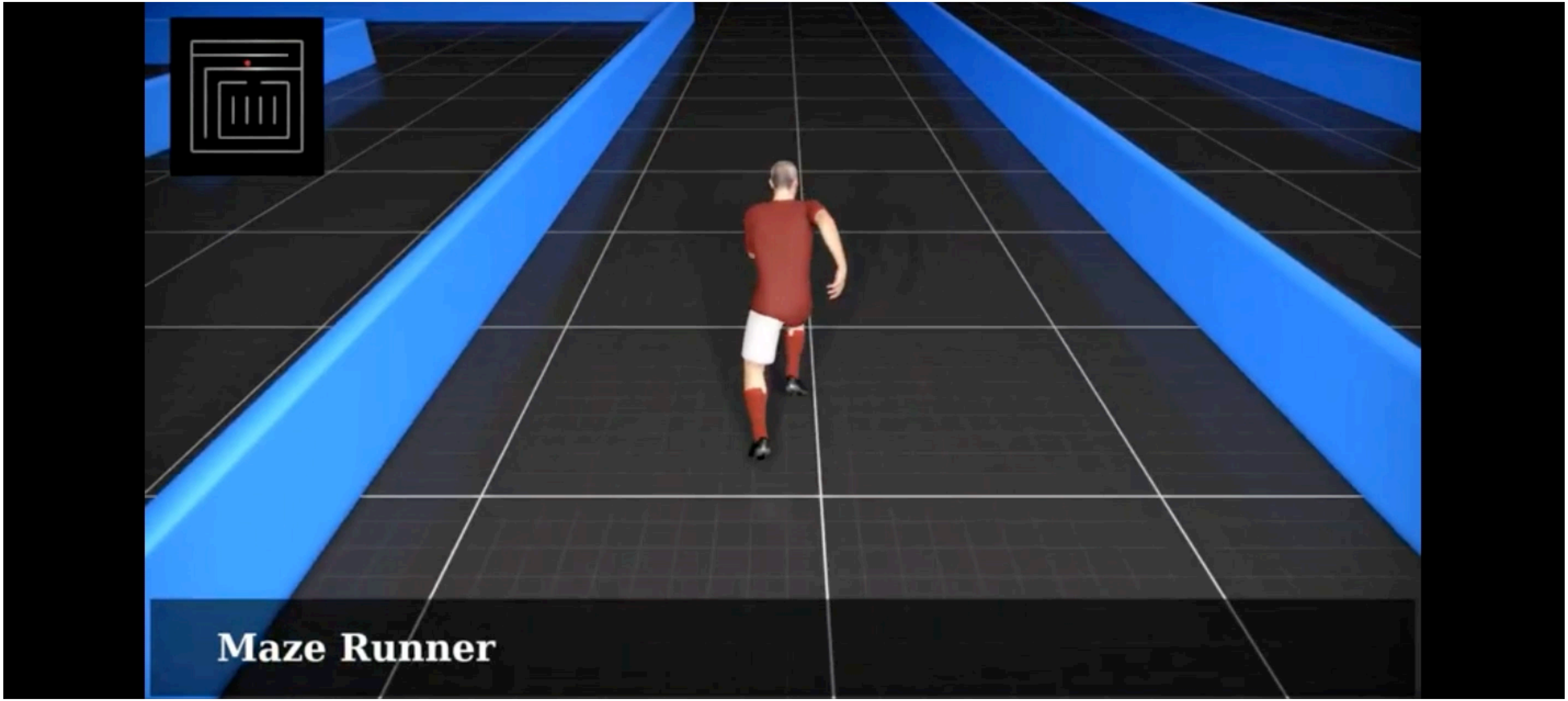
Control the decoder

- The decoder can generate plausible next frames given the current frame, but it does not achieve any specific task
- Train a control policy that controls the decoder using Reinforcement Learning
  - Input state: current pose $\mathbf{q}_t$
  - Output action: latent variable $\mathbf{z}$





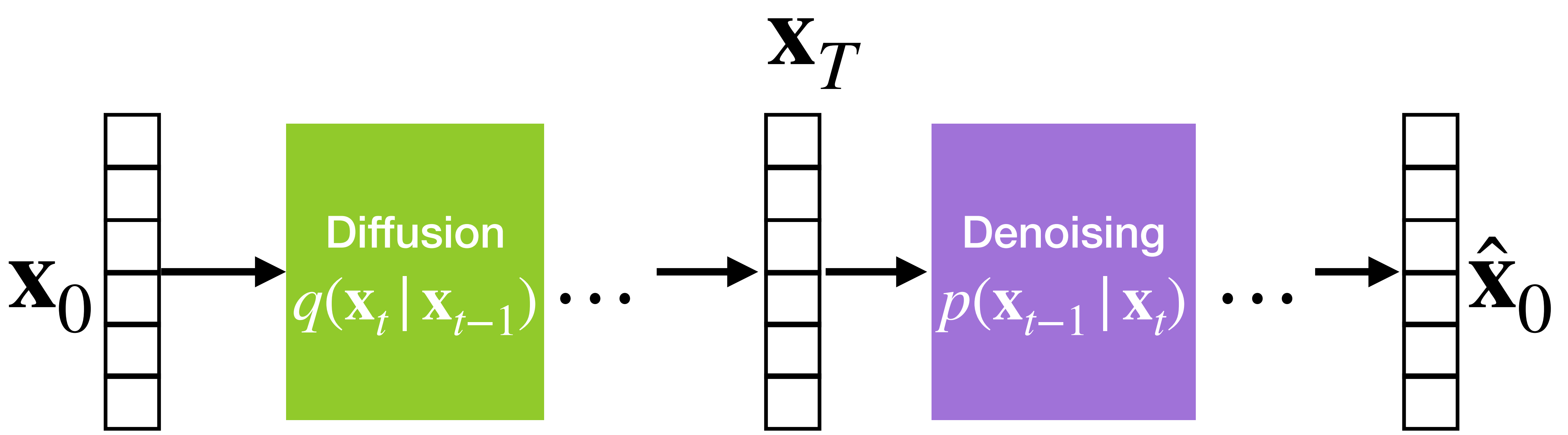
Ling et al SIGGRAPH 2021



Ling et al SIGGRAPH 2021

Diffusion Model

- **Diffusion process gradually adds gaussian noise to the original data; by approximating the reverse process, we can synthesize data from noise**



- **Diffusion process:**

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = N(\mathbf{x}_t;\ \boldsymbol{\mu}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1},\ \boldsymbol{\Sigma}_t = \beta_t \mathbf{I})$$

How to get $\mathbf{x}_t$? Repeatedly draw samples from q conditioned on the previous sample $\mathbf{x}_{t-1}$.

Using reparameterization trick: $\mathbf{x}_t = q(\mathbf{x}_0, t) = \bar{\alpha}_t \mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is drawn from $N(\mathbf{0}, \mathbf{I})$, and $\bar{\alpha}_t = \Pi_{s=0}^{t}(1-\beta_t)$

- **Denoising process:**

Approximate the reverse distribution: $p(\mathbf{x}_{t-1} \mid \mathbf{x}_t, t) = N(\mathbf{x}_{t-1};\ \boldsymbol{\mu}_\theta(\mathbf{x}_t, t),\ \Sigma_\theta(\mathbf{x}_t, t))$

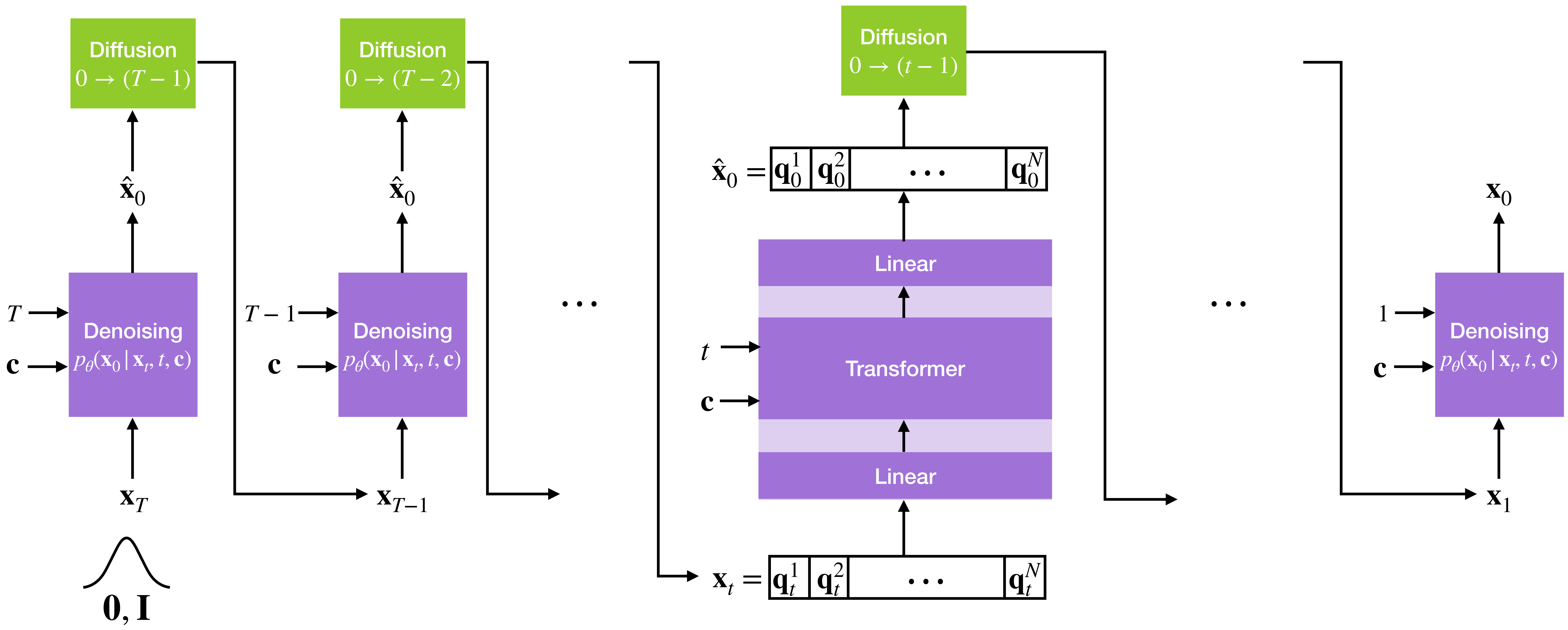
How to get $\mathbf{x}_0$? Repeatedly draw sample from p conditioned on the "next" sample $\mathbf{x}_{t+1}$

How to train p? In practice, we train a neural network that predicts $\mathbf{x}_0$ from $\mathbf{x}_t$, conditioned on t, $\hat{\mathbf{x}}_0 = p_\theta(\mathbf{x}_t, t)$.

$$L_\theta = \mathbb{E}_{\mathbf{x}_0, t}\left[\| \mathbf{x}_0 - \boxed{p_\theta\big(q(\mathbf{x}_0, t), t\big)} \| \right]$$

diffuse a data point for t steps and denoise it back to $\hat{\mathbf{x}}_0$

Diffusion Models for Human Motion



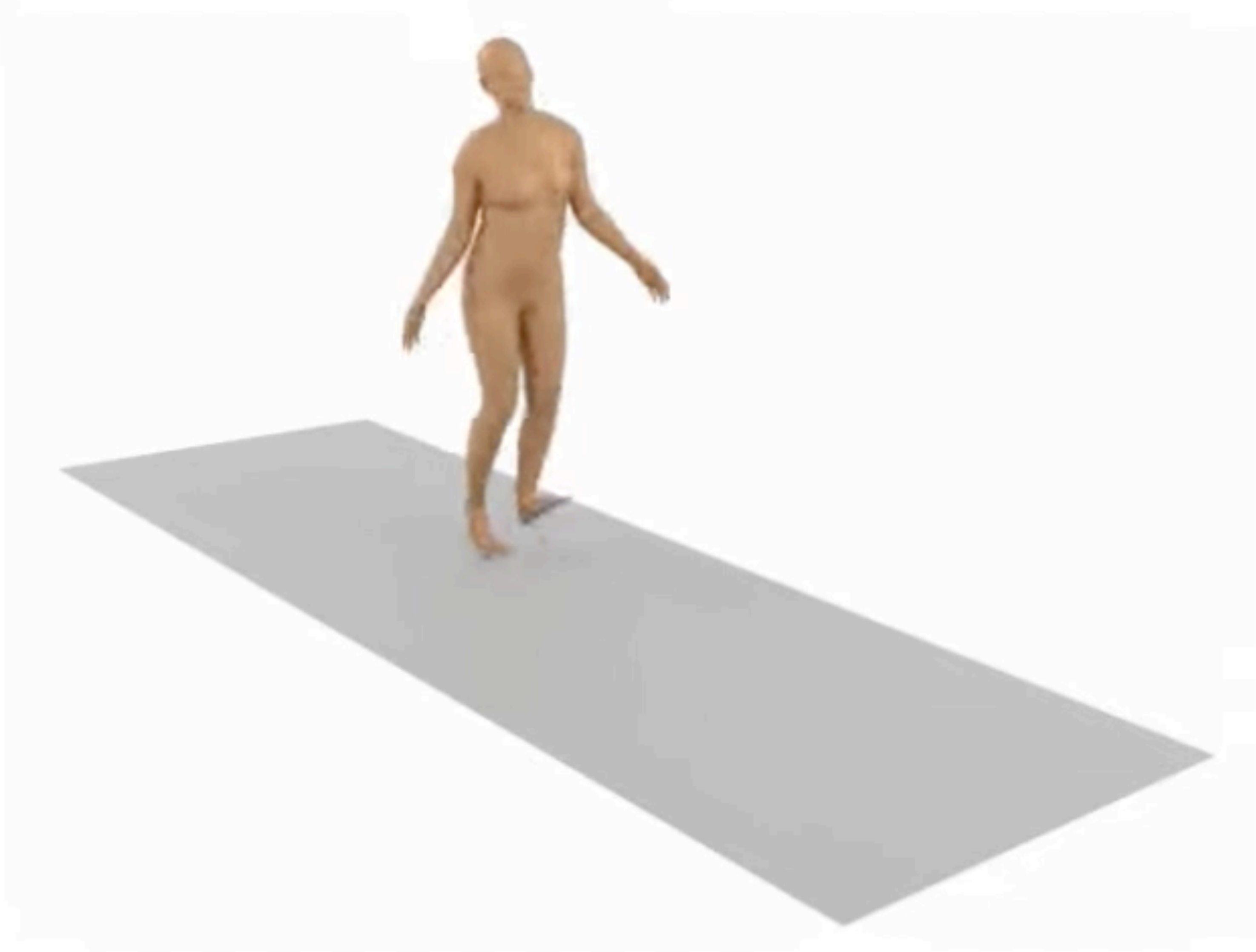
Human Motion Diffusion Model

Guy Tevet Sigal Raab Brian Gordon Yonatan Shafir
Daniel Cohen-Or Amit H. Bermano

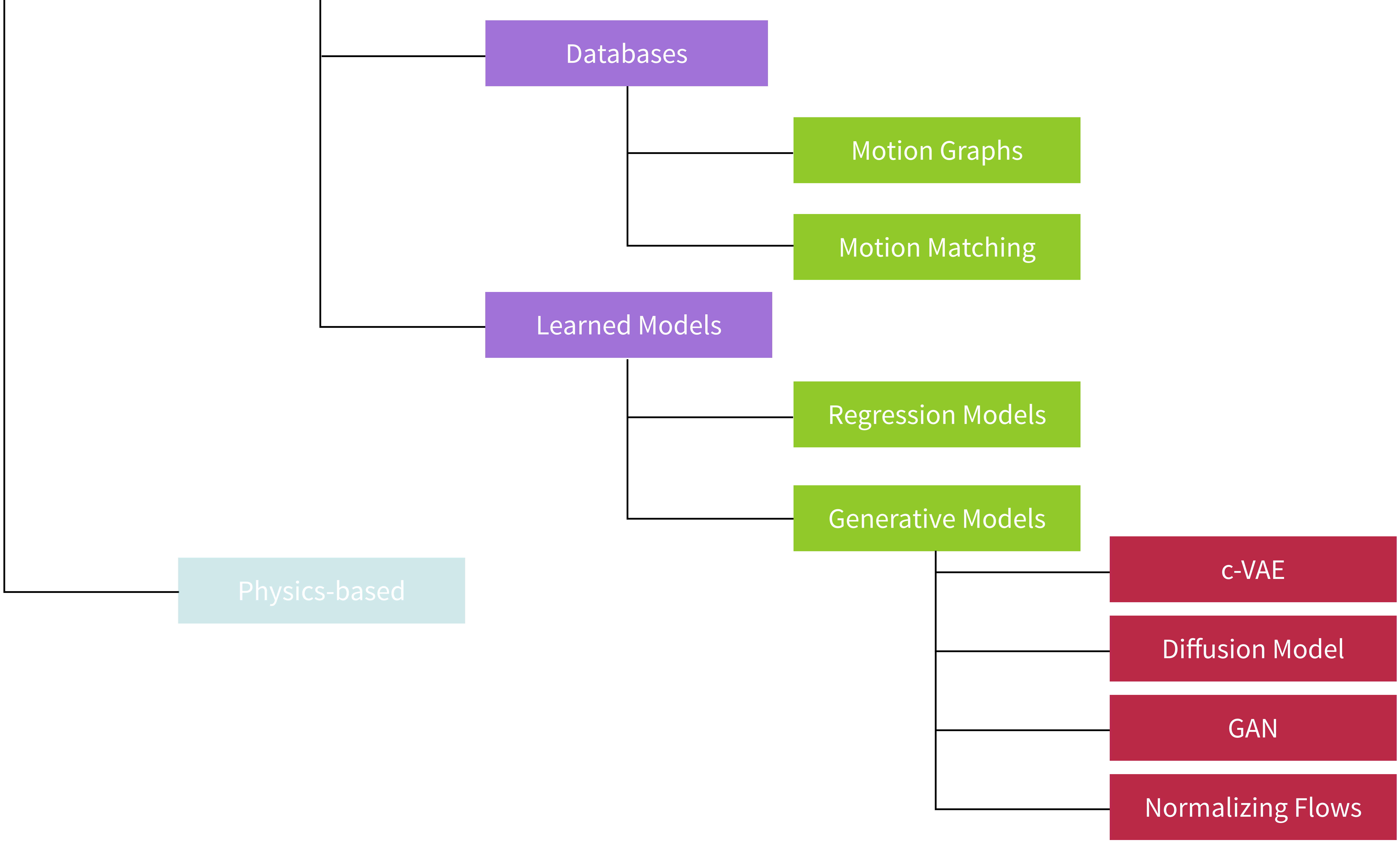
Tel Aviv University, Israel

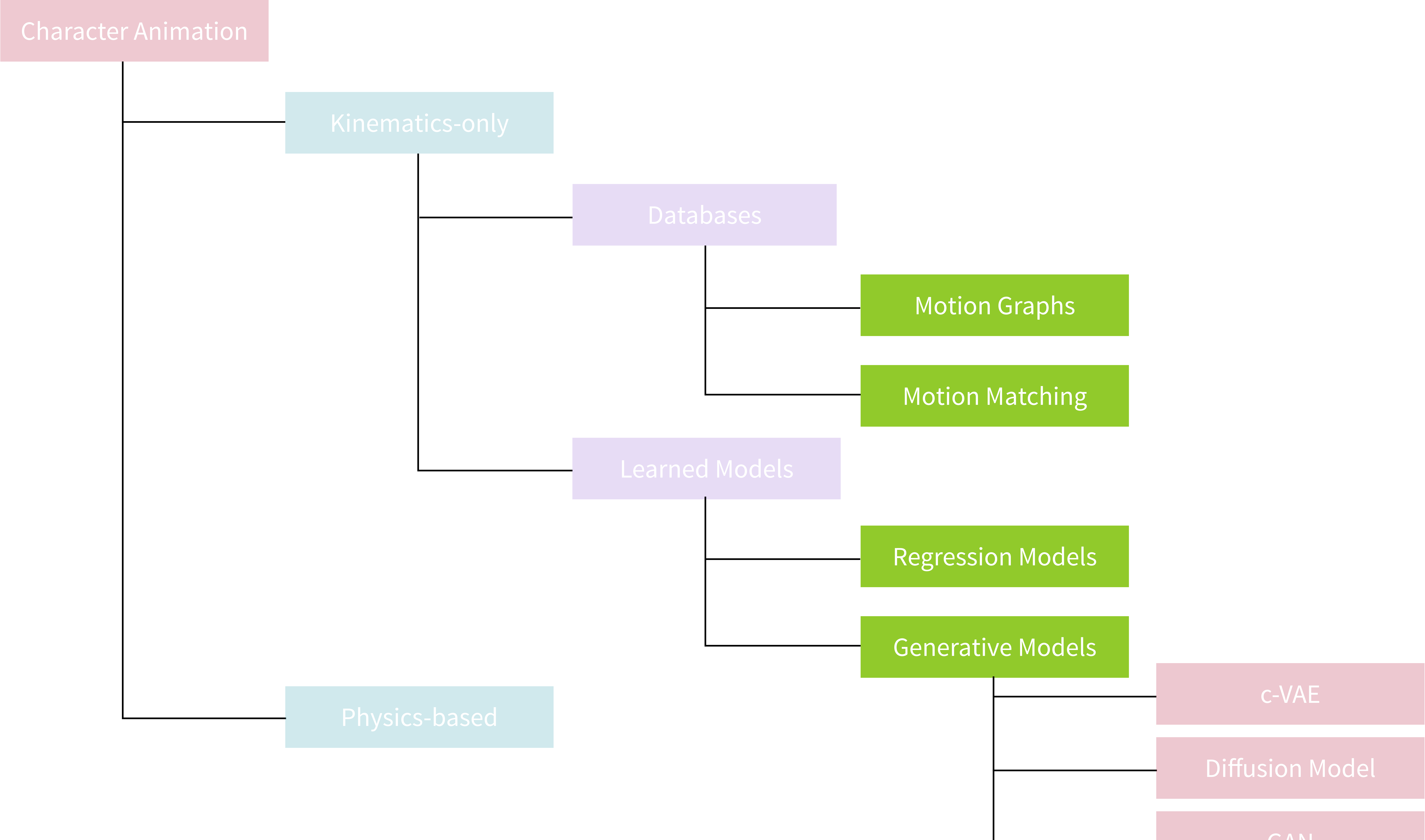
https://github.com/andreas128/RePaint

Image Diffusion

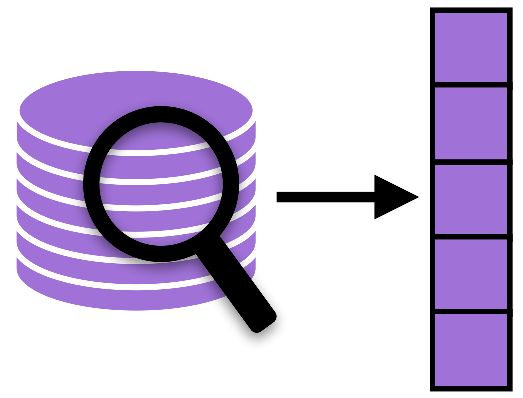


Motion Diffusion


Databases
Motion Graphs
Motion Matching
Learned Models
Regression Models
Generative Models
c-VAE
Diffusion Model
GAN
Normalizing Flows
Physics-based


Character Animation
Kinematics-only
Databases
Motion Graphs
Motion Matching
Learned Models
Regression Models
Generative Models
c-VAE
Diffusion Model
Physics-based

Comparisons

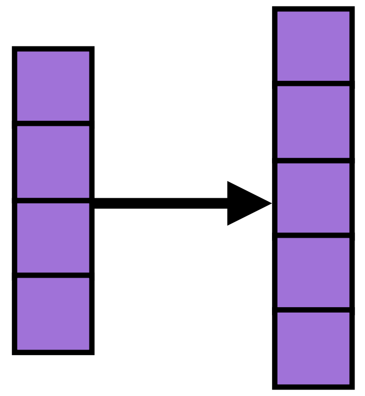


Motion Matching

Motion Matching vs Regression

- Both require dense control signals to produce best next frame
- Motion matching explicitly selects best match while Regression learns the best match

Regression Models



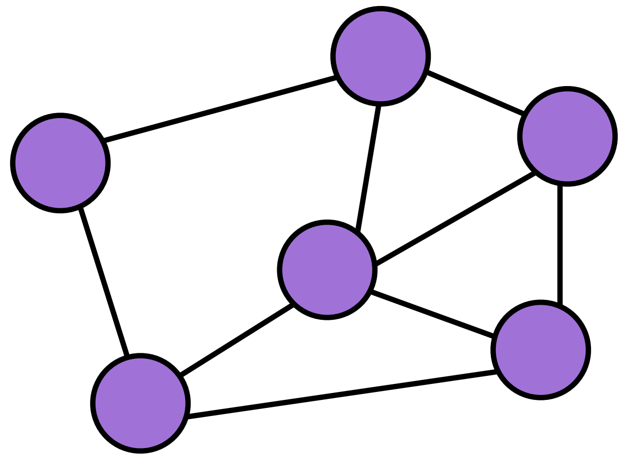
Motion Matching vs Motion Graphs

- Both require access to motion datasets
- Motion matching computes transitions at execution time while motion graphs at construction time
- Motion matching works with short-horizon control while motion graphs allows for long-term planning

Different approaches to online motion synthesis

Regression vs Generative Models

- Both train neural networks to produce next frame and do not need to access data once trained
- Regression models use an explicitly defined control space while generative models use a learned control space

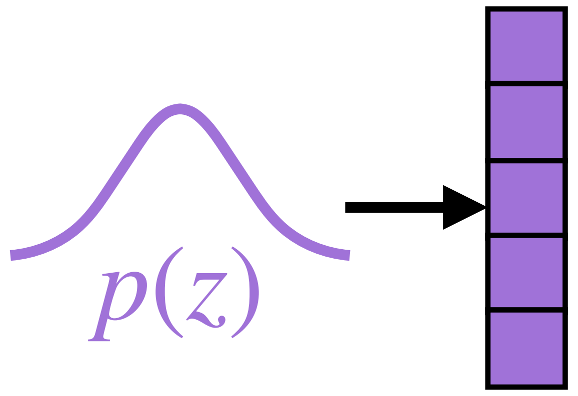


Motion Graphs

Motion Graphs vs Generative Models

- Both need an auxiliary planning/control algorithm for functional tasks
- Both predict multiple plausible next frames given current frame
- Generative models predict the distribution of next frame while motion graph uses graph edges to model next frames

Generative Models



Evaluation

- Similarity to data
  - Joint position and orientation
  - Root position and orientation
- Physical plausibility
  - Penetration with geometry
  - Unrealistic foot contact
- Task performance
- User studies