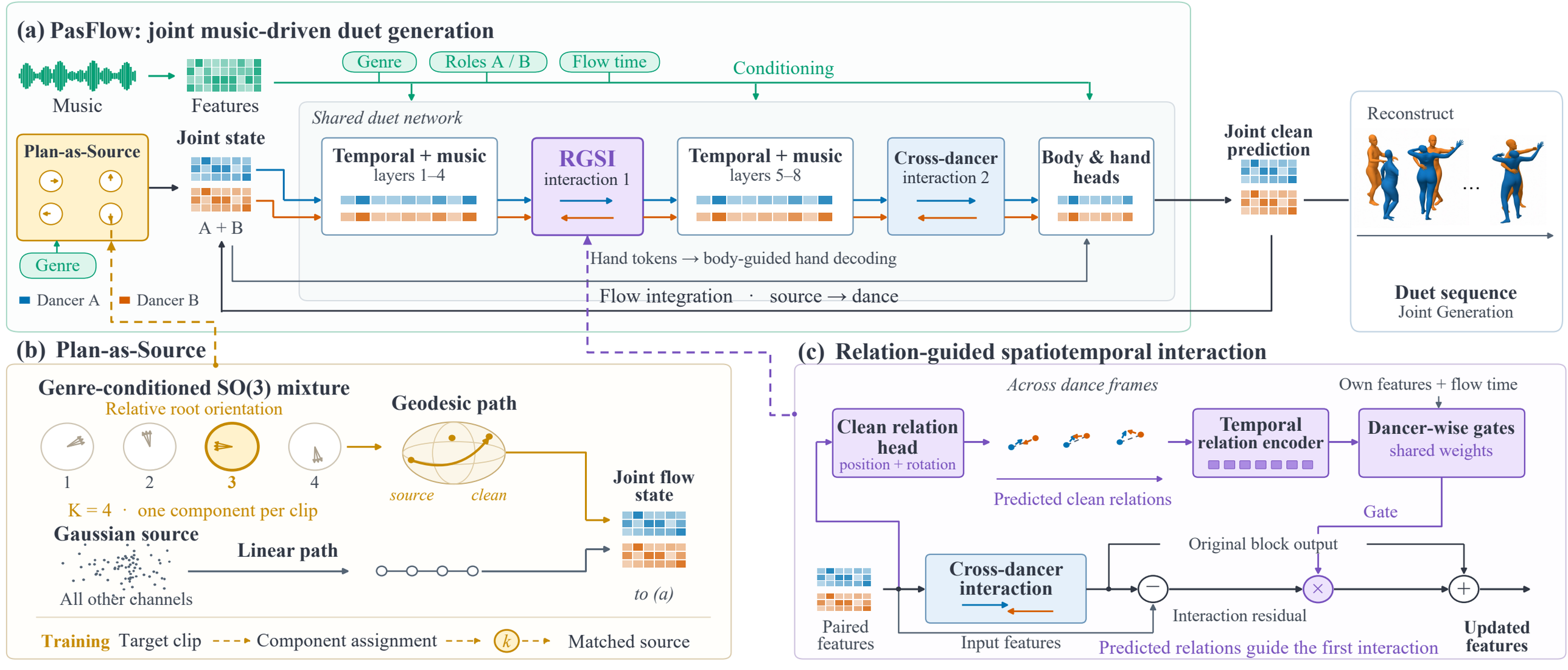

(a) PasFlow: joint music-driven duet generation
Music
Features
Genre
Roles A / B
Flow time
Conditioning
Shared duet network
Plan-as-Source
Joint state
A + B
Genre
Dancer A
Dancer B
Temporal + music
layers 1–4
RGSI
interaction 1
Temporal + music
layers 5–8
Cross-dancer
interaction 2
Body & hand
heads
Hand tokens → body-guided hand decoding
Flow integration · source → dance
Joint clean
prediction
Reconstruct
Duet sequence
Joint Generation
(b) Plan-as-Source
Genre-conditioned SO(3) mixture
Relative root orientation
1
2
3
4
K = 4 · one component per clip
Geodesic path
source
clean
Gaussian source
All other channels
Linear path
Joint flow
state
to (a)
Training
Target clip
Component assignment
k
Matched source
(c) Relation-guided spatiotemporal interaction
Across dance frames
Clean relation
head
position + rotation
Predicted clean relations
Temporal
relation encoder
Own features + flow time
Dancer-wise gates
shared weights
Gate
Original block output
Paired
features
Cross-dancer
interaction
Interaction residual
Input features
Predicted relations guide the first interaction
Updated
features