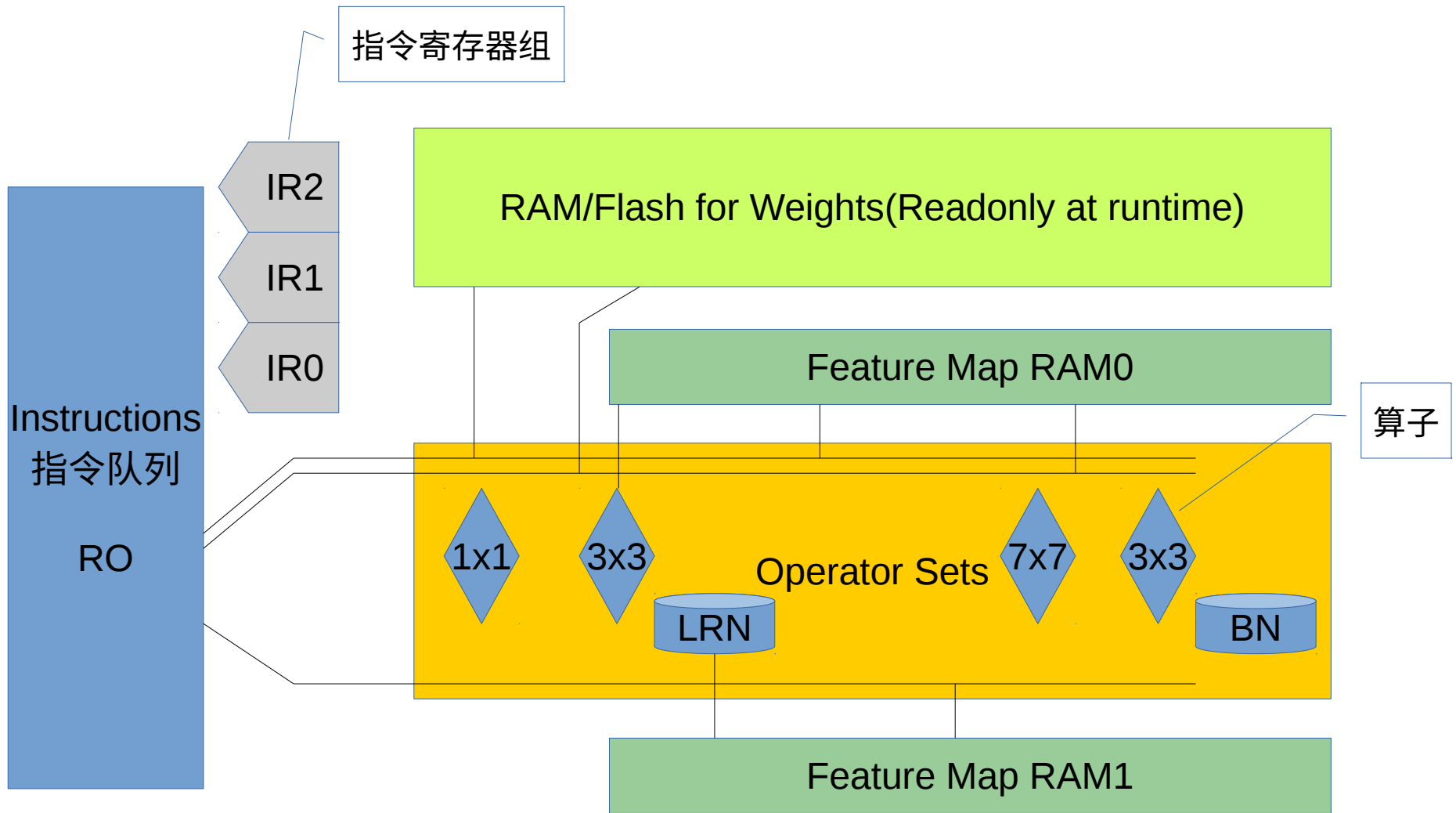
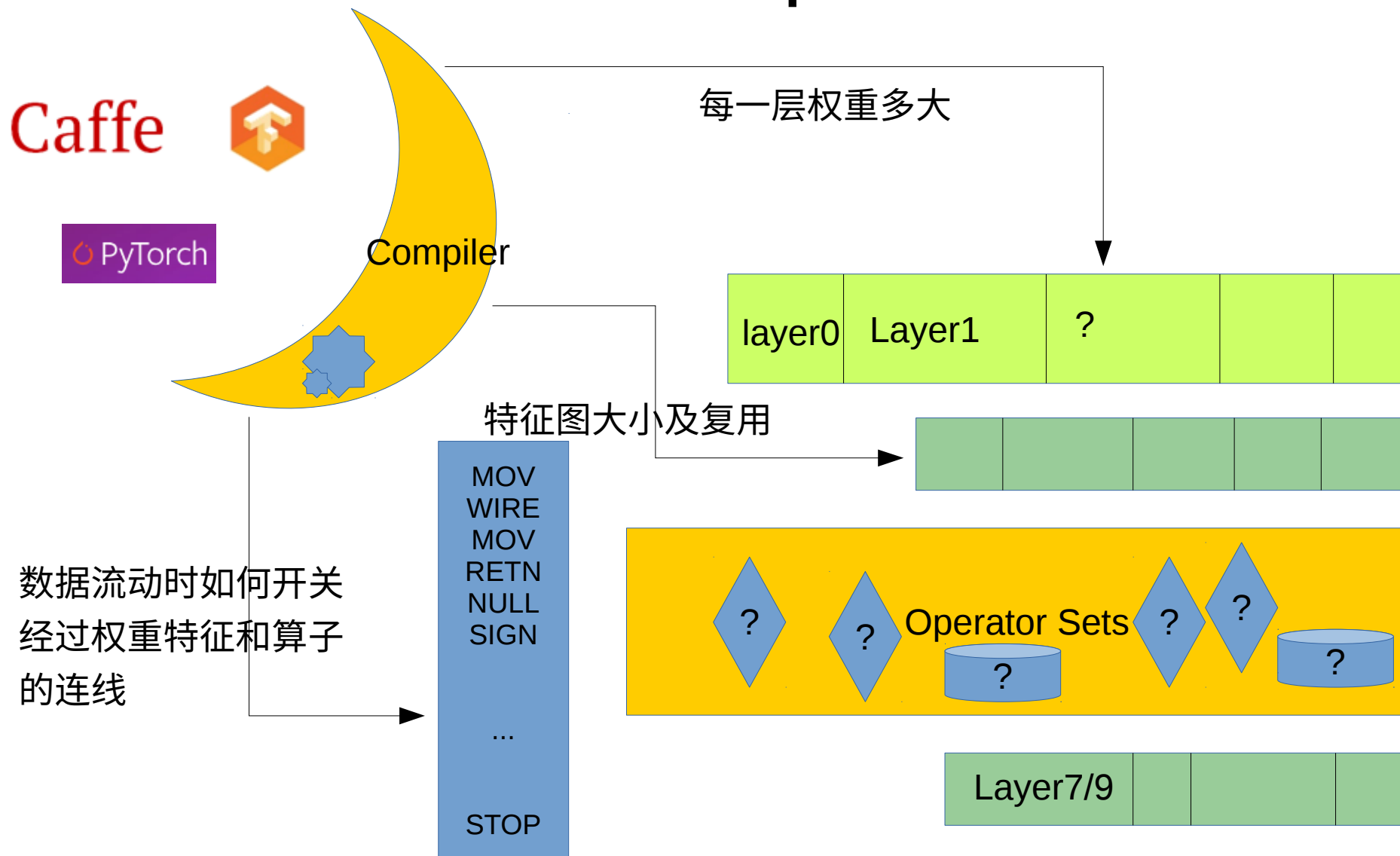


Memory Wall Free Architecture for Neural Network Computing

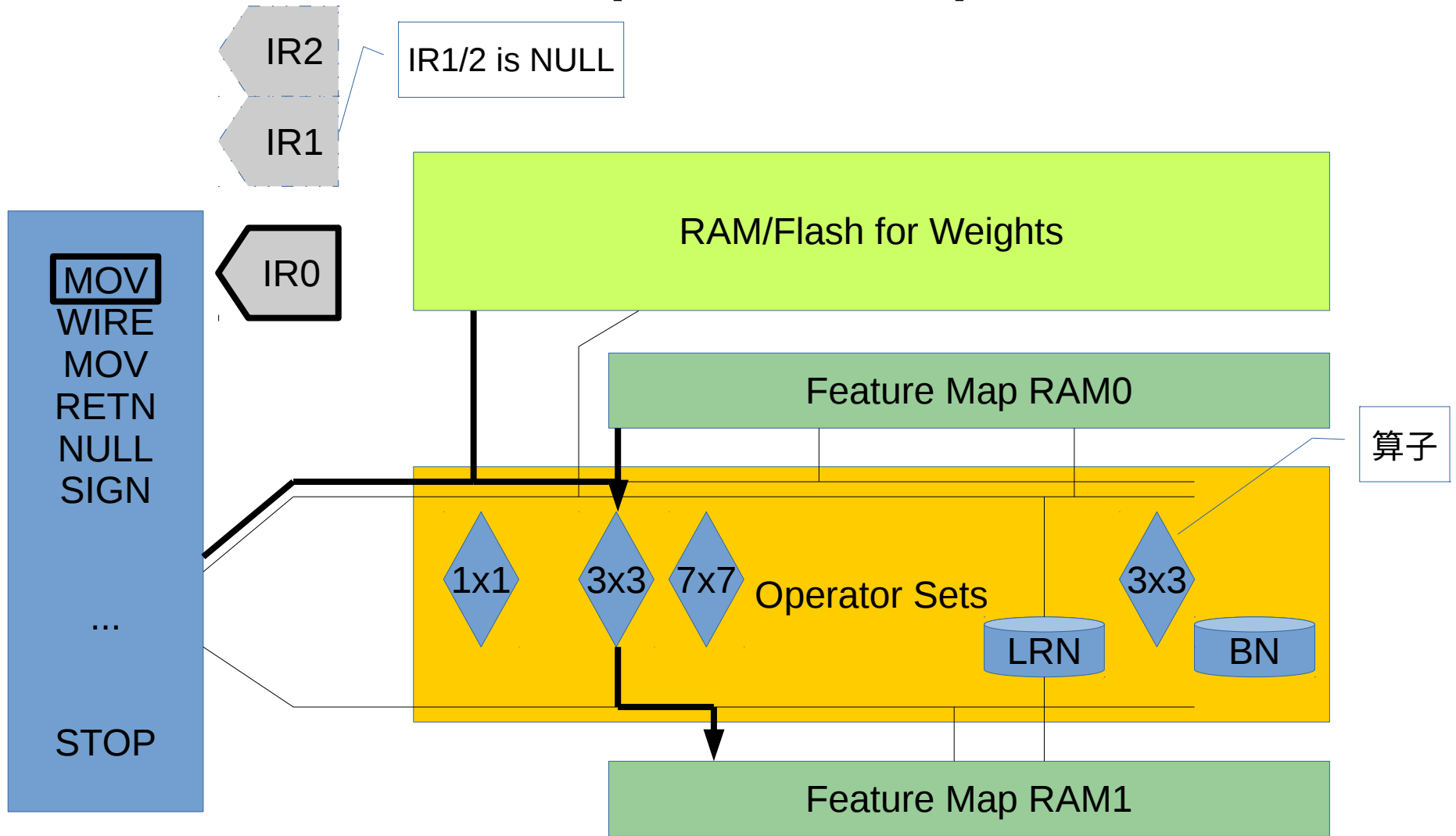


Graph Compute Architecture Compile



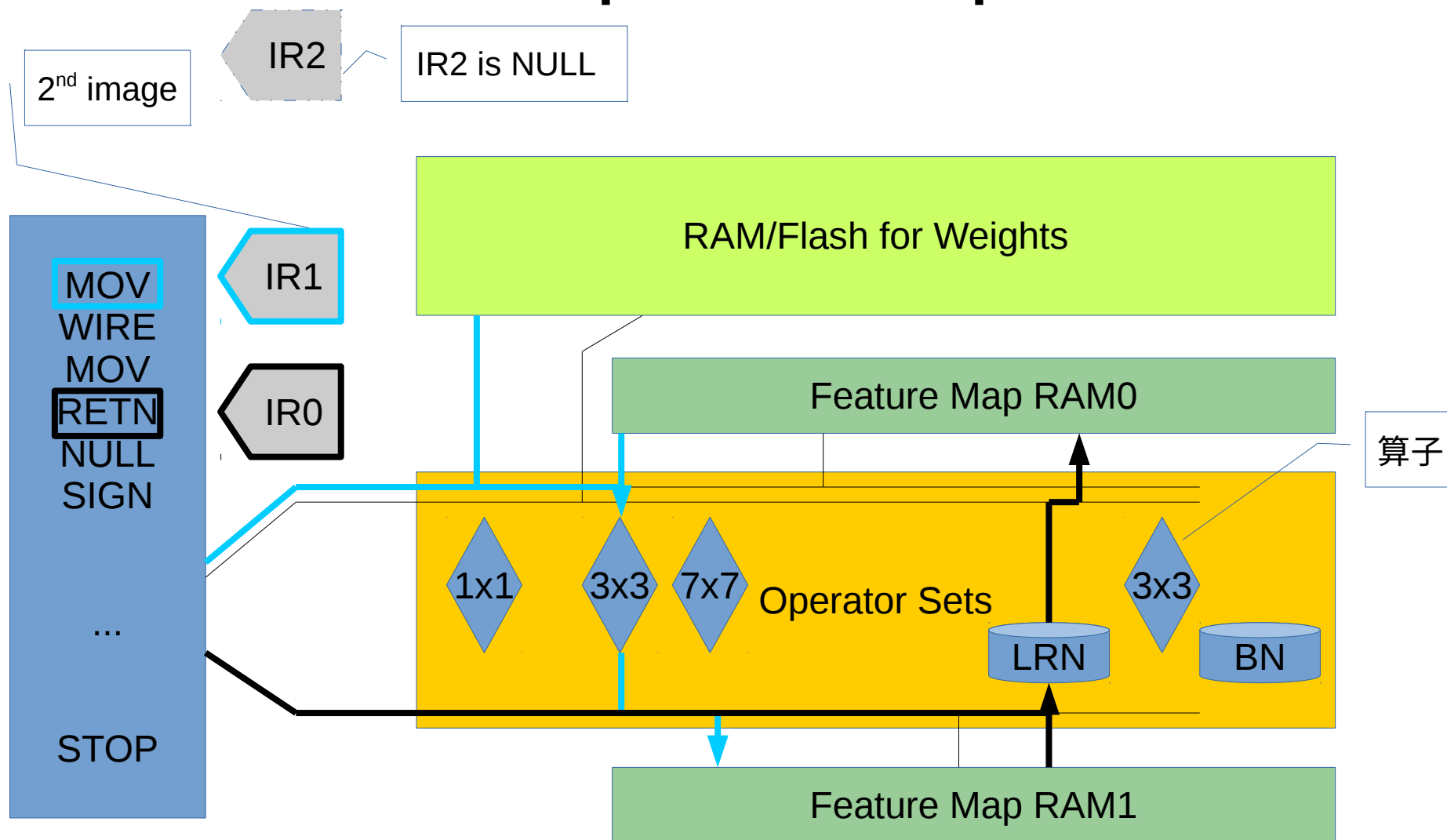
Graph Compute Architecture

Compute Step 1



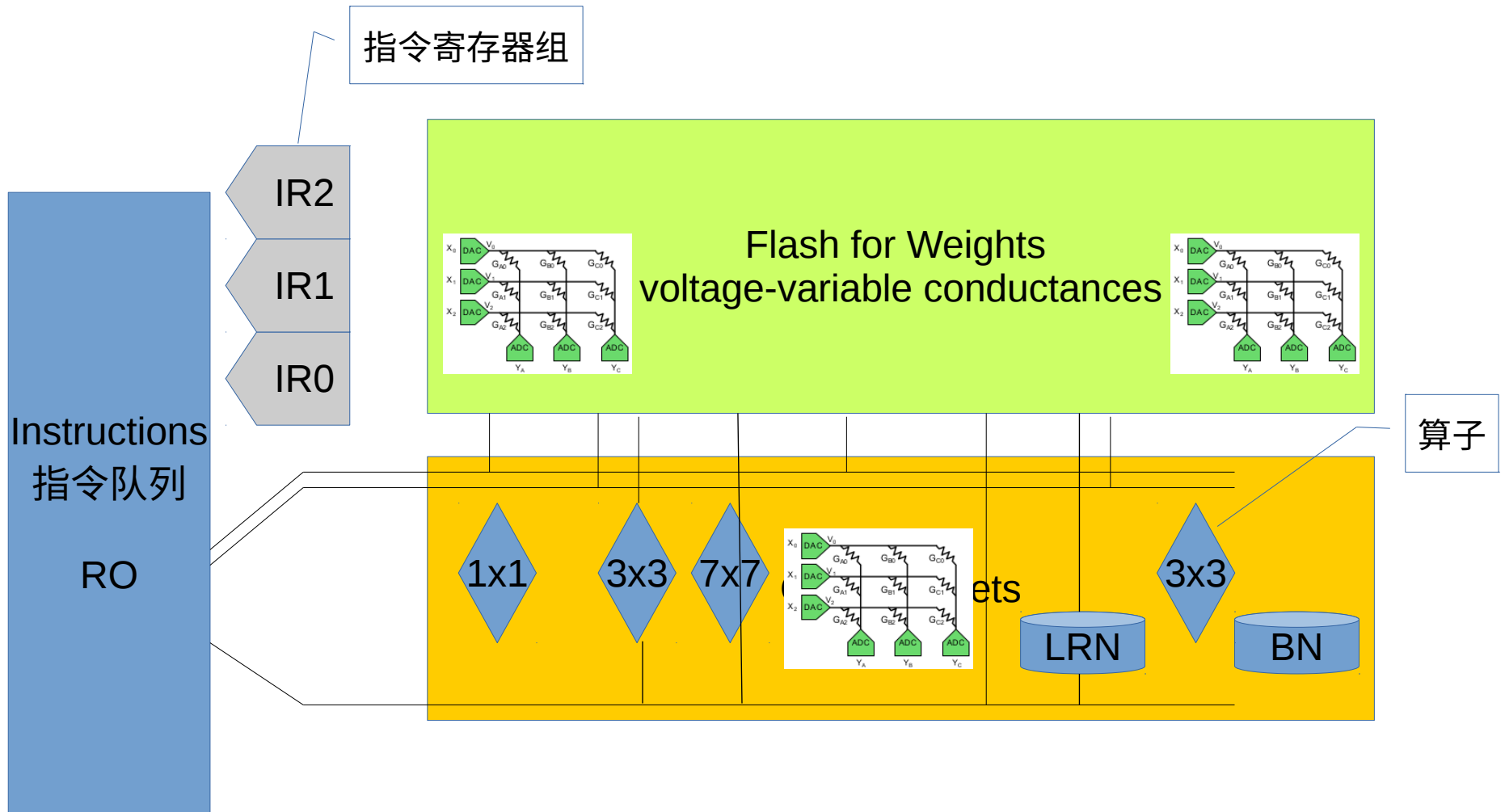
Graph Compute Architecture

Computer Step 2



Graph Compute Architecture

Low Power Consumption



We will deploy NLP CTC-RNN and Attention LSTM

For an input,
like speech



Predict a
sequence of
tokens

h e e ϵ l ϵ l l o o !

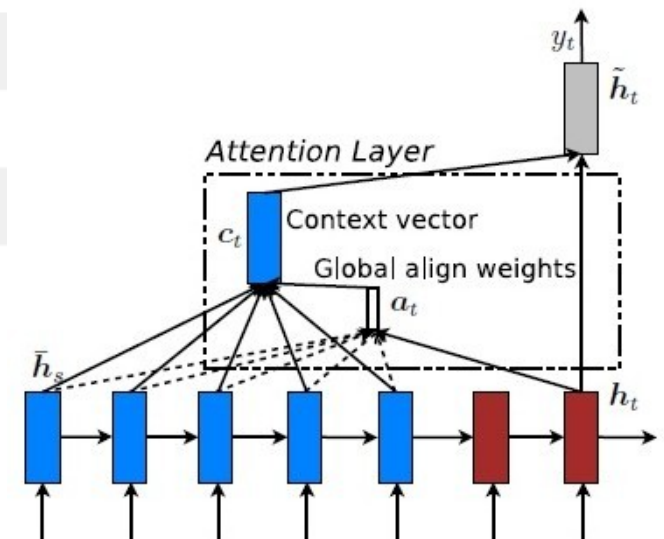
Merge repeats,
drop ϵ

h e l l o !

Final output

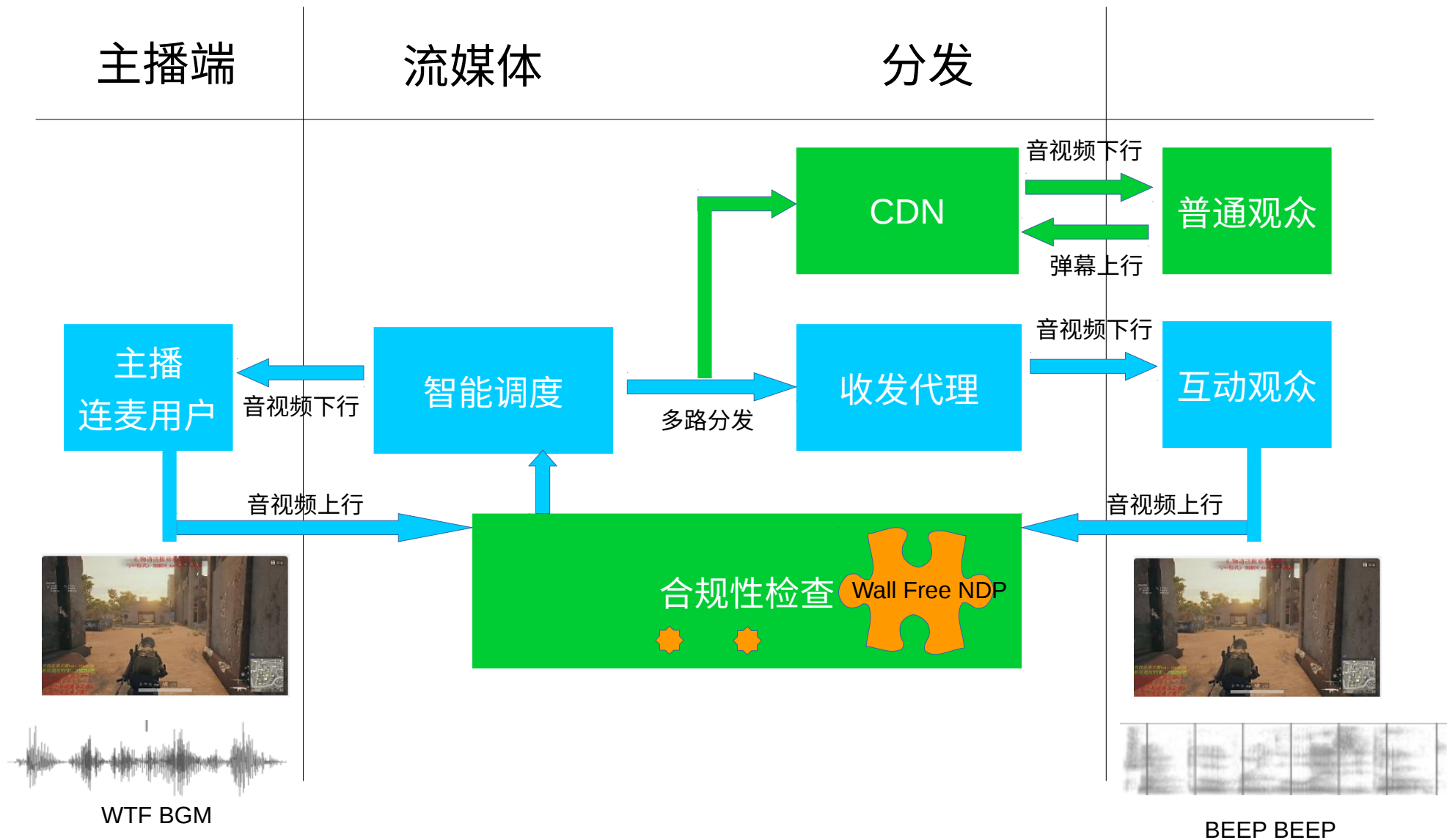
h e l l o !

Both are state-of-the-art algorithm
with sequence data



Cloud User Scenarion

Beep illegal word for live video



Edge User Scenarion 1

NLP for Office Device

Edge User Scenarion 2

NLP and Low Power for Drones

Edge User Scenarion 3

NLP for Room Device

- Any one can help

We want do more
Word for cmd and print for Priority

This is the future
Robot can help US