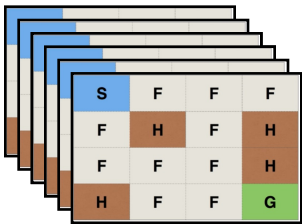
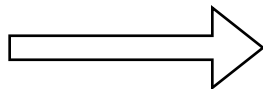


Map configurations
(train)

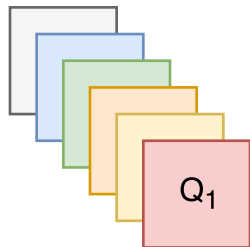


Q-learning

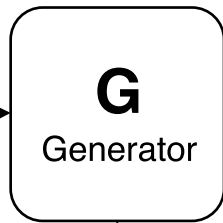


Real
Samples

Training
Q-table
policies



z
latent space
(random vector)



Generated
Fake
Q-table
policies

D
Discriminator

*real or
fake policy
?*

Fine tune training

