methods, as shown in Fig. 1. The model can qualitative change the grid parameter matrix for learning and designing irregular structures based on the clustering results.

The role of the clustering algorithm is to cluster the acti space, i.e., different directions of movement of grid points, to obtain several typical actions.

The role of reinforcement learning is to complete t design task, which depends on the PAAC architecture. PAAC-K is a parallel architecture based on the advantage actor-critic algorithm. Unlike in A3C, there is only one copy of the parameters in PAAC-K, and parameter updates are performed synchronously. The data are collected during the design process without calculating the gradient. Therefore, training instability is avoided and resources are wasted with asynchronous updates. The policy network (F_1) generates the design strategy of the filter, and the critic network (F_2) evaluates the strategy. Using the exploration mechanism in

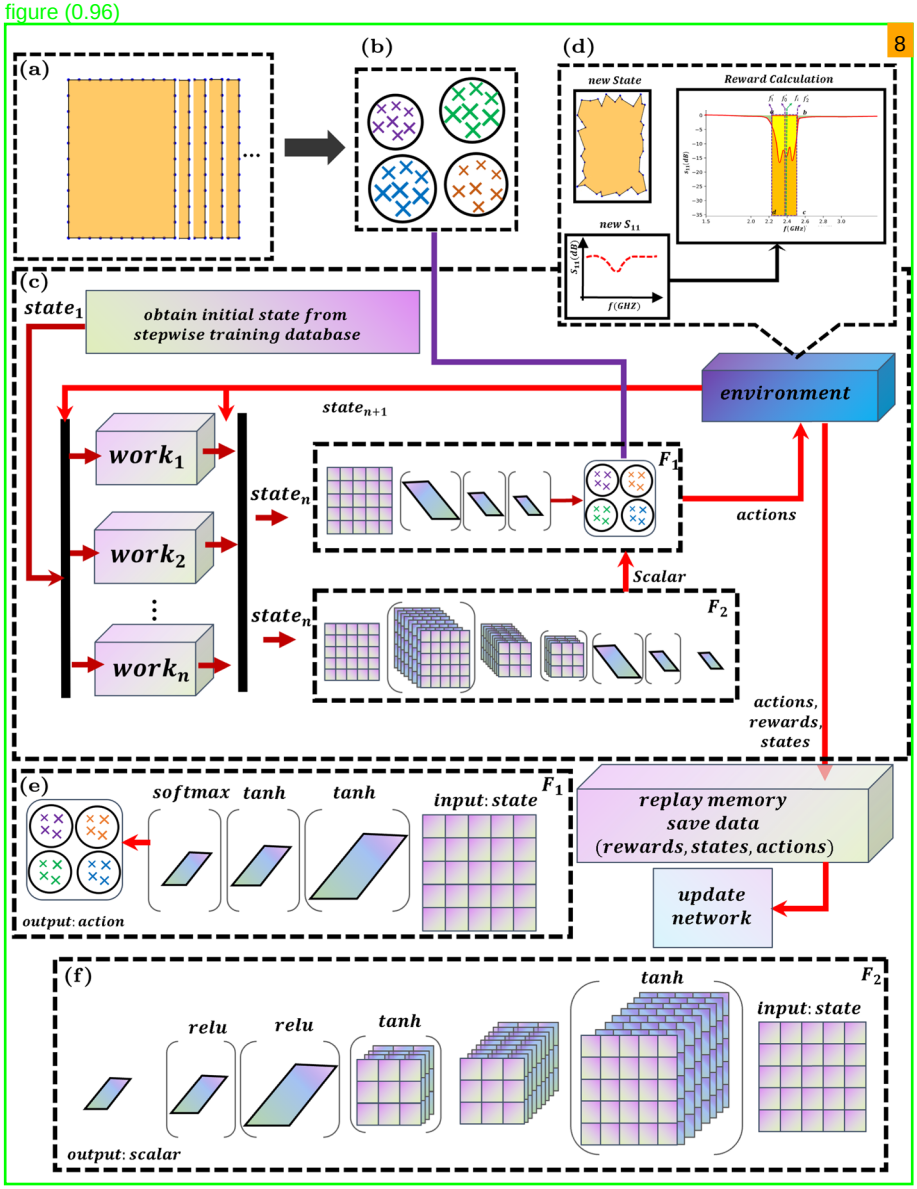
training aims to limit the probability distribution of poli $\pi(a_t \mid s_t; \theta)$, which is more rational in that different actions have a better chance of being adopted.

Stepwise training is a method that can greatly reduce t agent's exploration space. Reward functions help the agent find the circuit solution.

As shown in Fig. 1, n group environments are train simultaneously and maintained on a single device, with all work performed through interactions with the environment.

The whole environment is an electromagnetic compu ing process. It is calculated by invoking the *electromagnetic simulation-API* whenever the grid in Fig. 1d changes in any direction. The data generated by work performed through the environment are stored in replay memory, which is used to update the policy and value networks.

figure (0.96)



figure_caption (0.97)

Fig. 1 PAAC-K architecture. Dataset of the circuit grid parameter matrix. **b,** Clustering results. **c,** Environment interaction process. **d,** New state, new S_{11} curve and reward calculation. **e,** F_1 is the policy network.The circuit grid parameter matrix is input, and a new action is returned. **f,** F_2 is a value network. The circuit grid parameter matrix is input, and a value scalar is returned