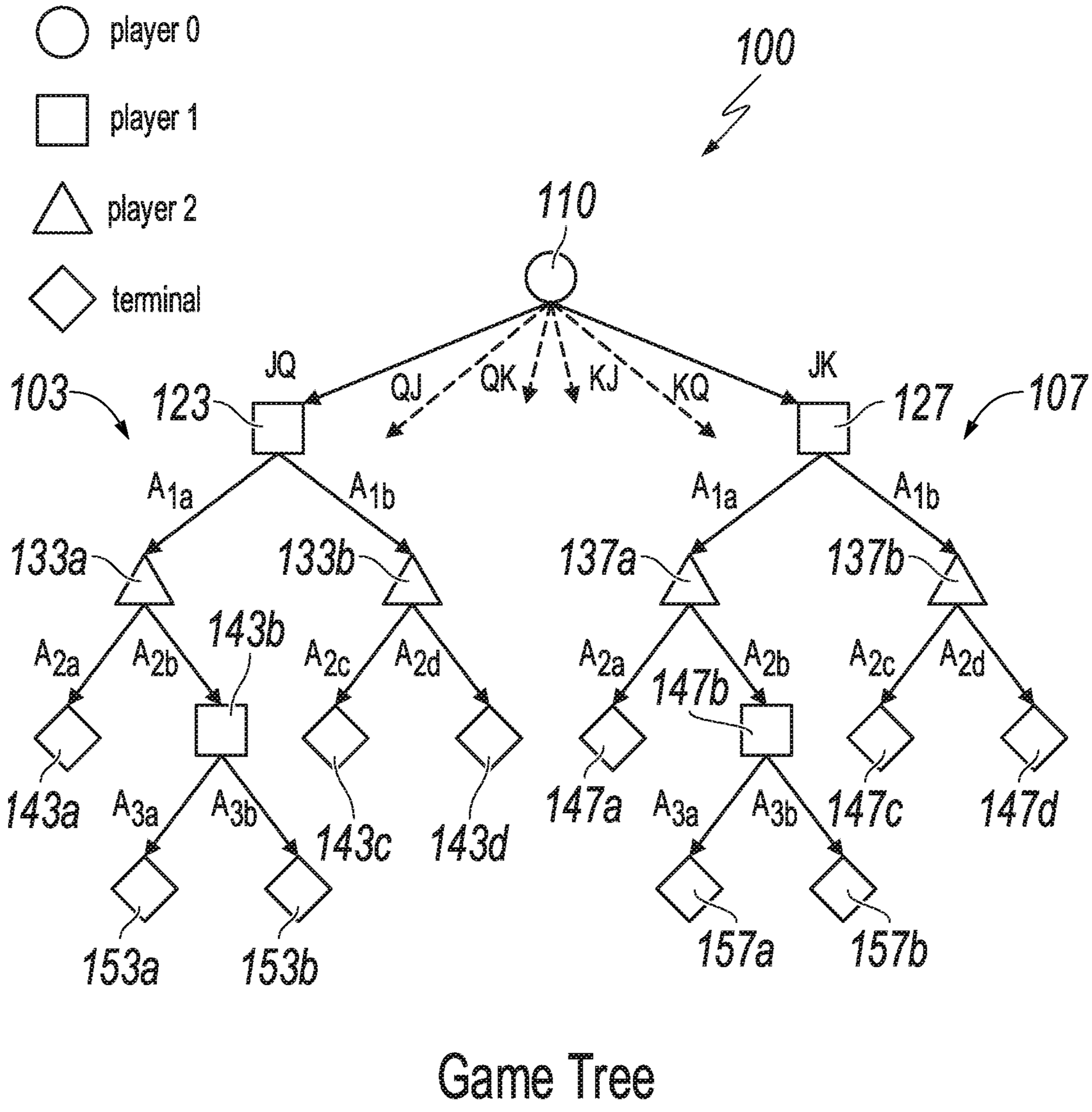
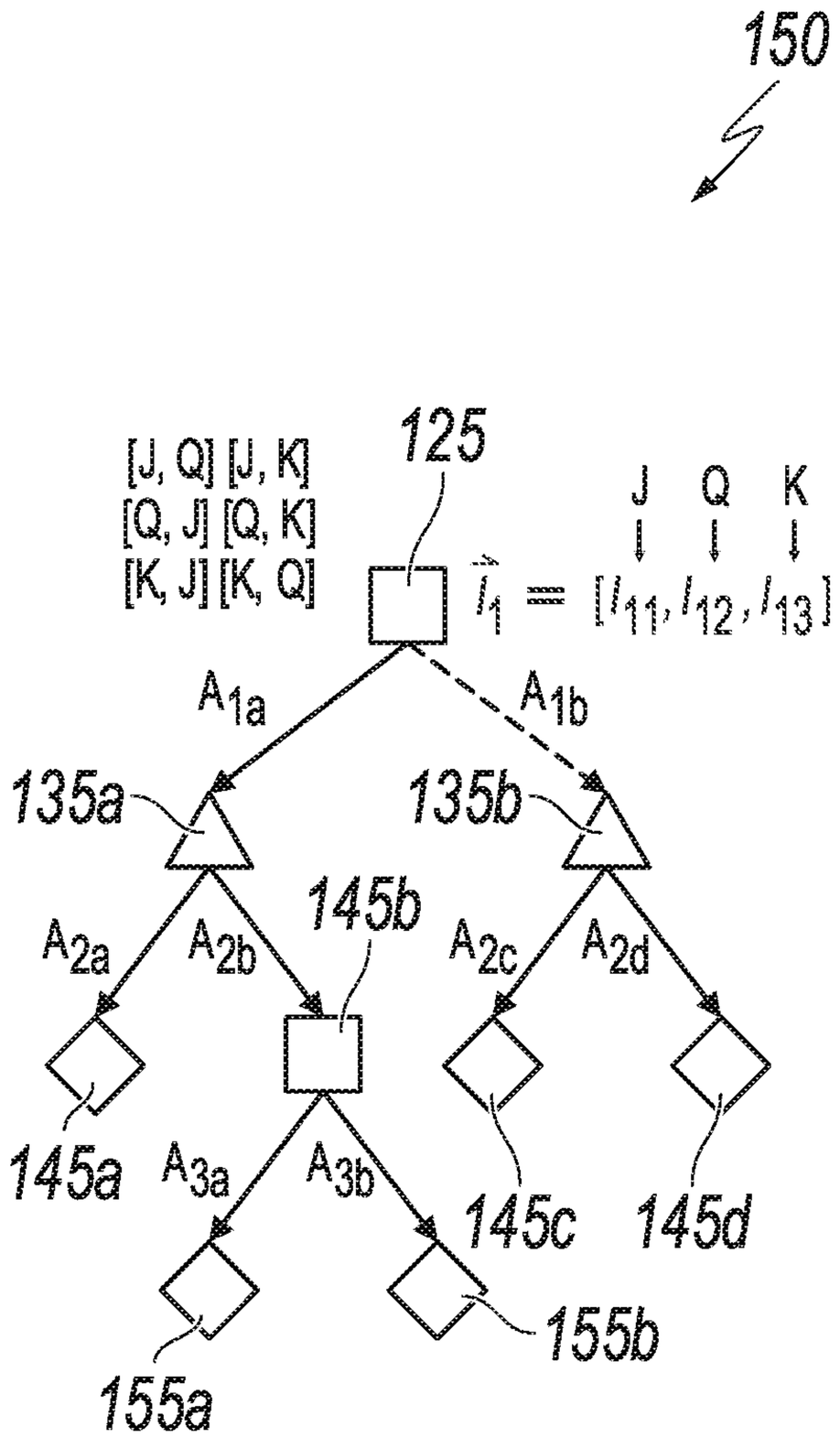


Game Tree

FIG. 1A



Public Tree

FIG. 1B

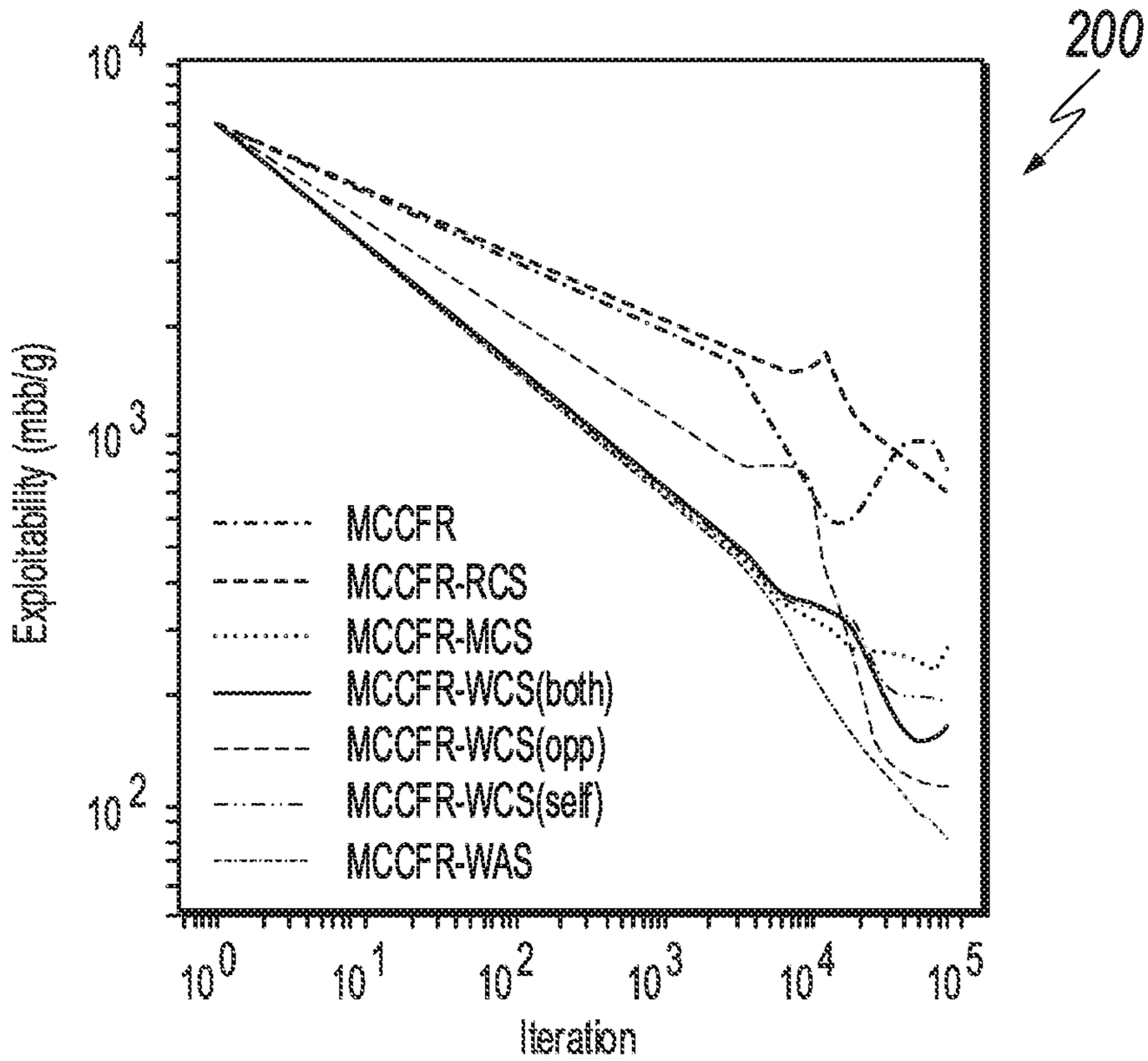

200
Exploitability (mbb/g)
10^4
10^3
10^2
MCCFR
MCCFR-RCS
MCCFR-MCS
MCCFR-WCS(both)
MCCFR-WCS(opp)
MCCFR-WCS(self)
MCCFR-WAS
10^0
10^1
10^2
10^3
10^4
10^5
Iteration

FIG. 2

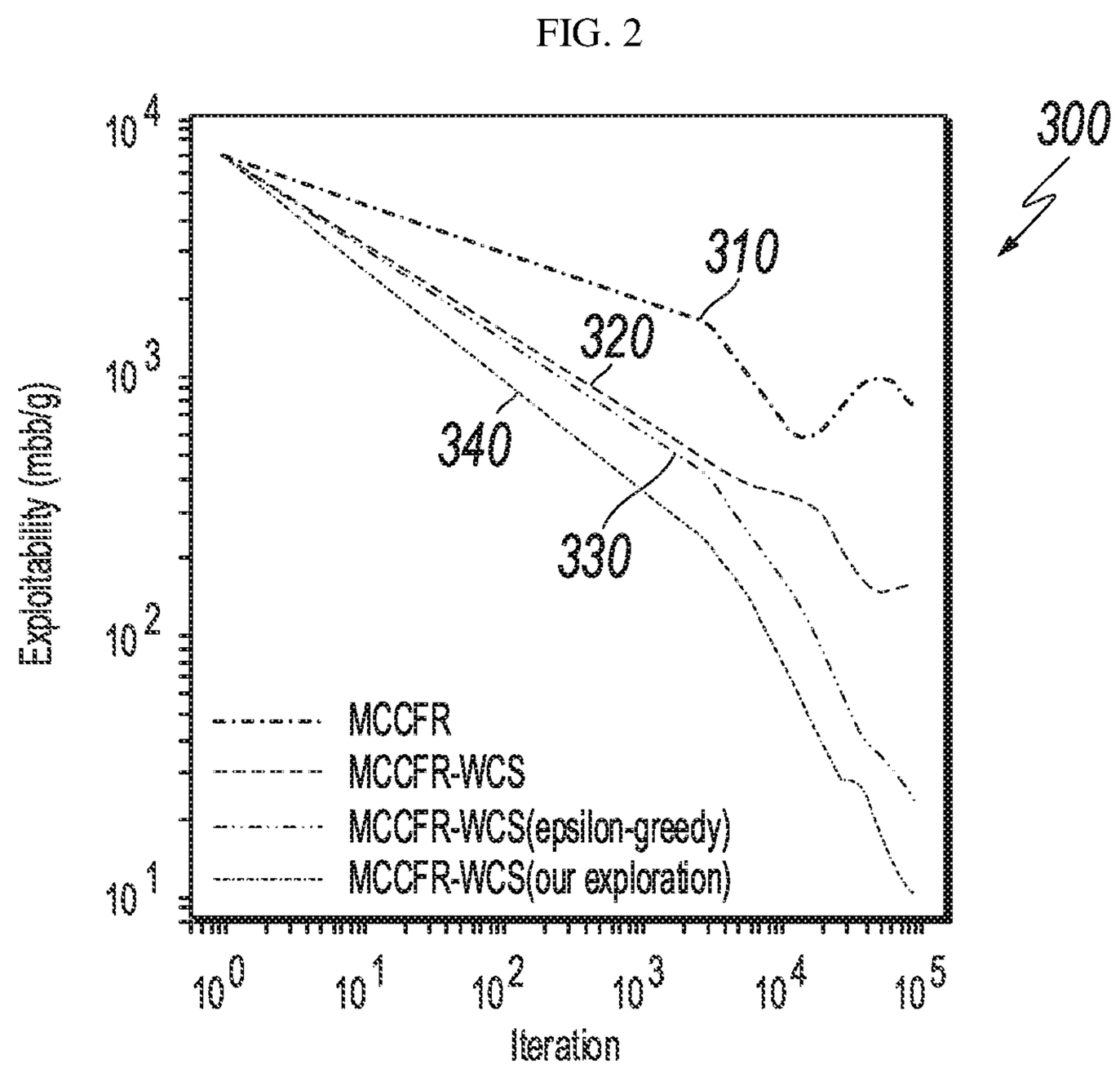

300
310
320
340
330
Exploitability (mbb/g)
10^4
10^3
10^2
10^1
MCCFR
MCCFR-WCS
MCCFR-WCS(epsilon-greedy)
MCCFR-WCS(our exploration)
10^0
10^1
10^2
10^3
10^4
10^5
Iteration

FIG. 3

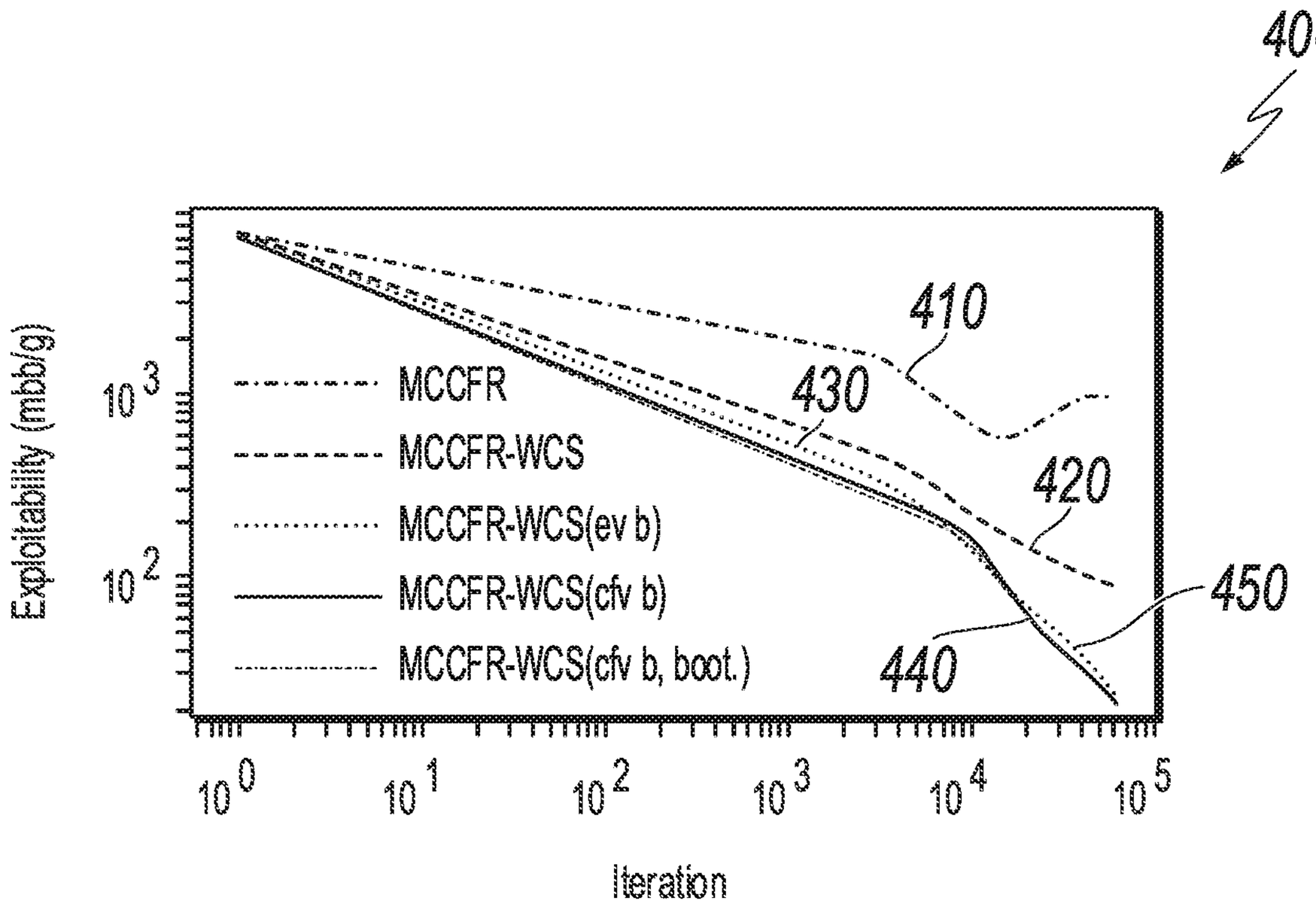

400
Exploitability (mbb/g)
10^3
10^2
MCCFR
MCCFR-WCS
MCCFR-WCS(ev b)
MCCFR-WCS(cfv b)
MCCFR-WCS(cfv b, boot.)
410
430
420
450
440
10^0
10^1
10^2
10^3
10^4
10^5
Iteration

FIG. 4A

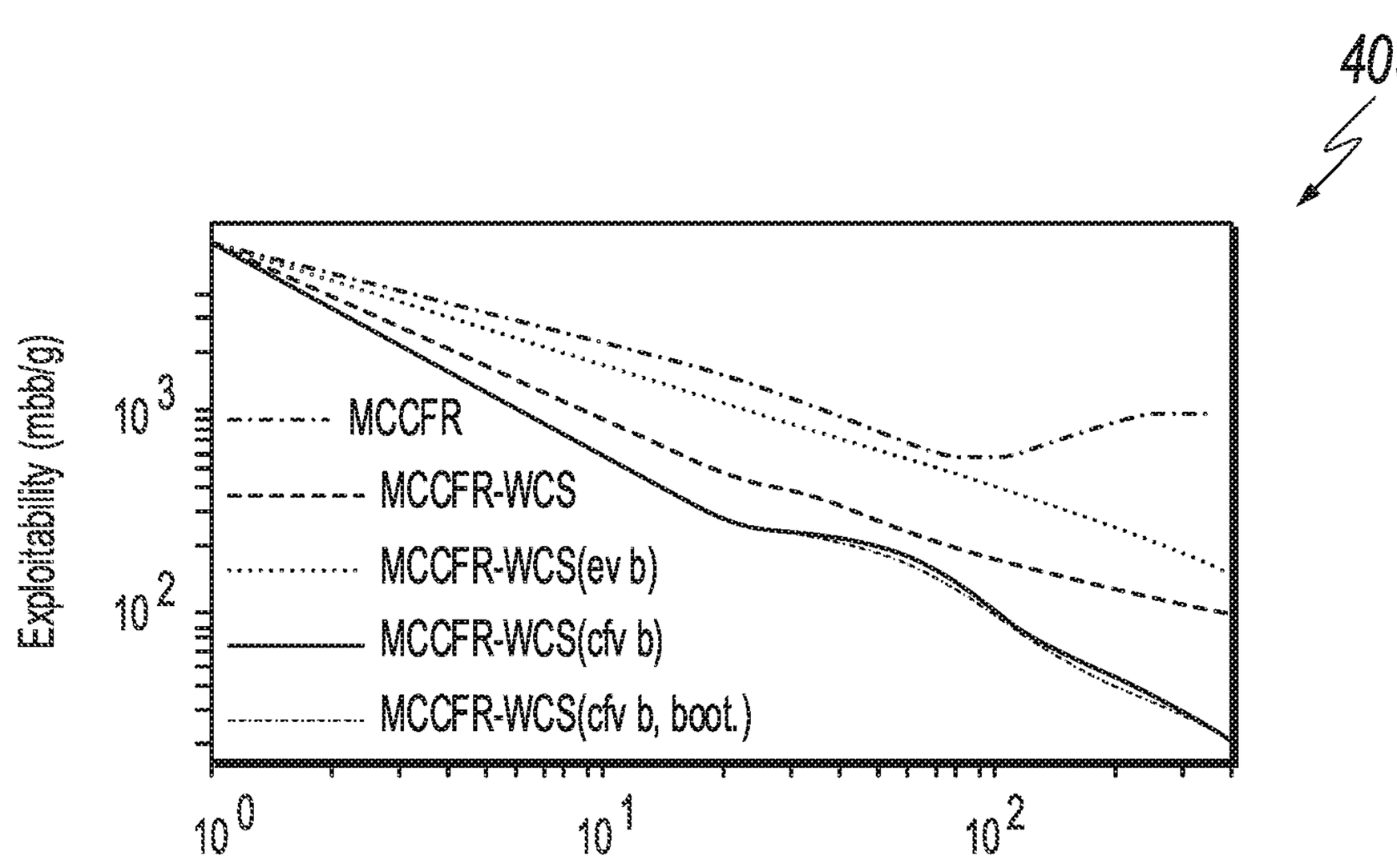

405
Exploitability (mbb/g)
10^3
10^2
MCCFR
MCCFR-WCS
MCCFR-WCS(ev b)
MCCFR-WCS(cfv b)
MCCFR-WCS(cfv b, boot.)
10^0
10^1
10^2
Time (Seconds)

FIG. 4B

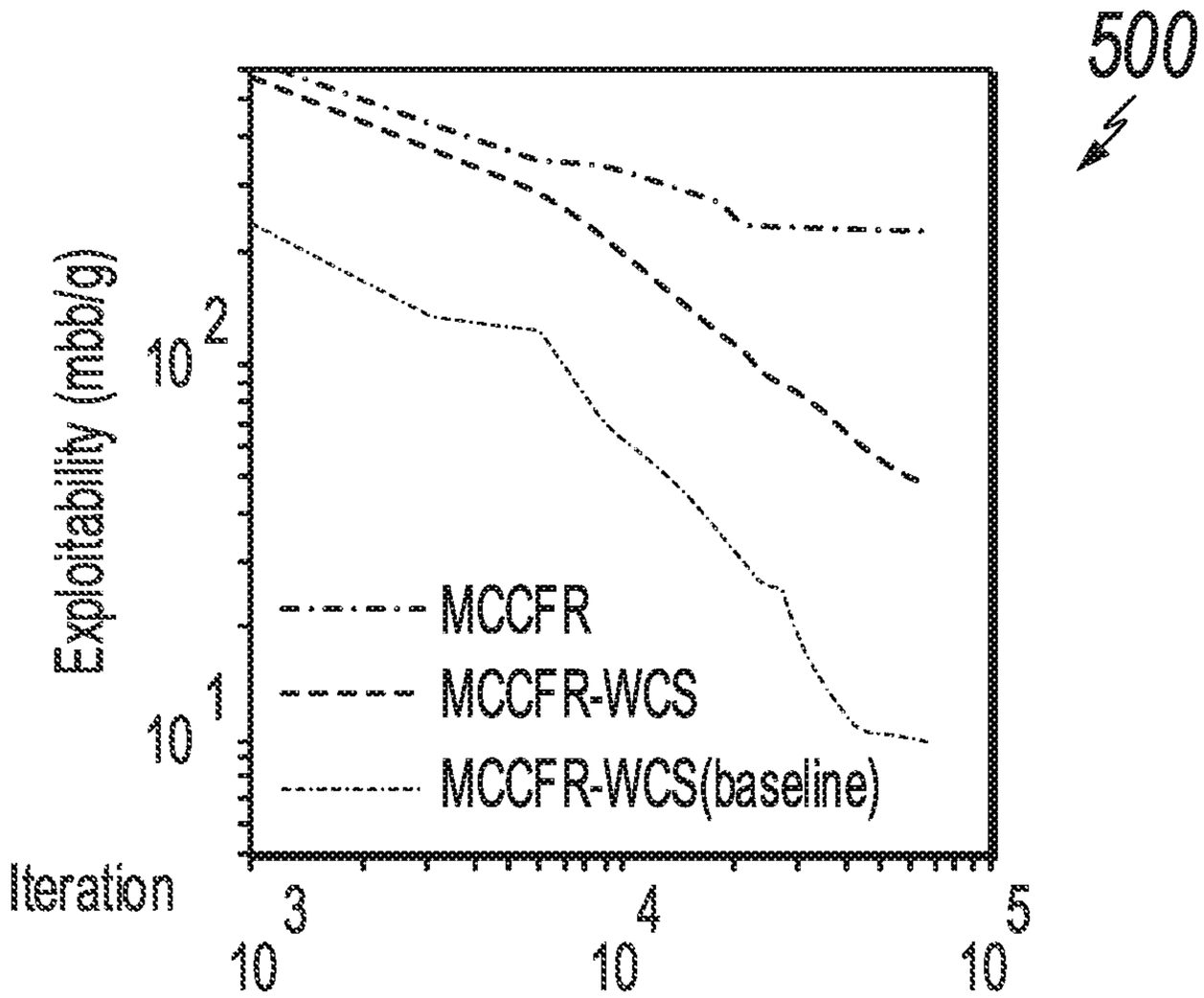

500
Exploitability (mbb/g)
10^2
10^1
MCCFR
MCCFR-WCS
MCCFR-WCS(baseline)
Iteration
10^3
10^4
10^5

FIG. 5A

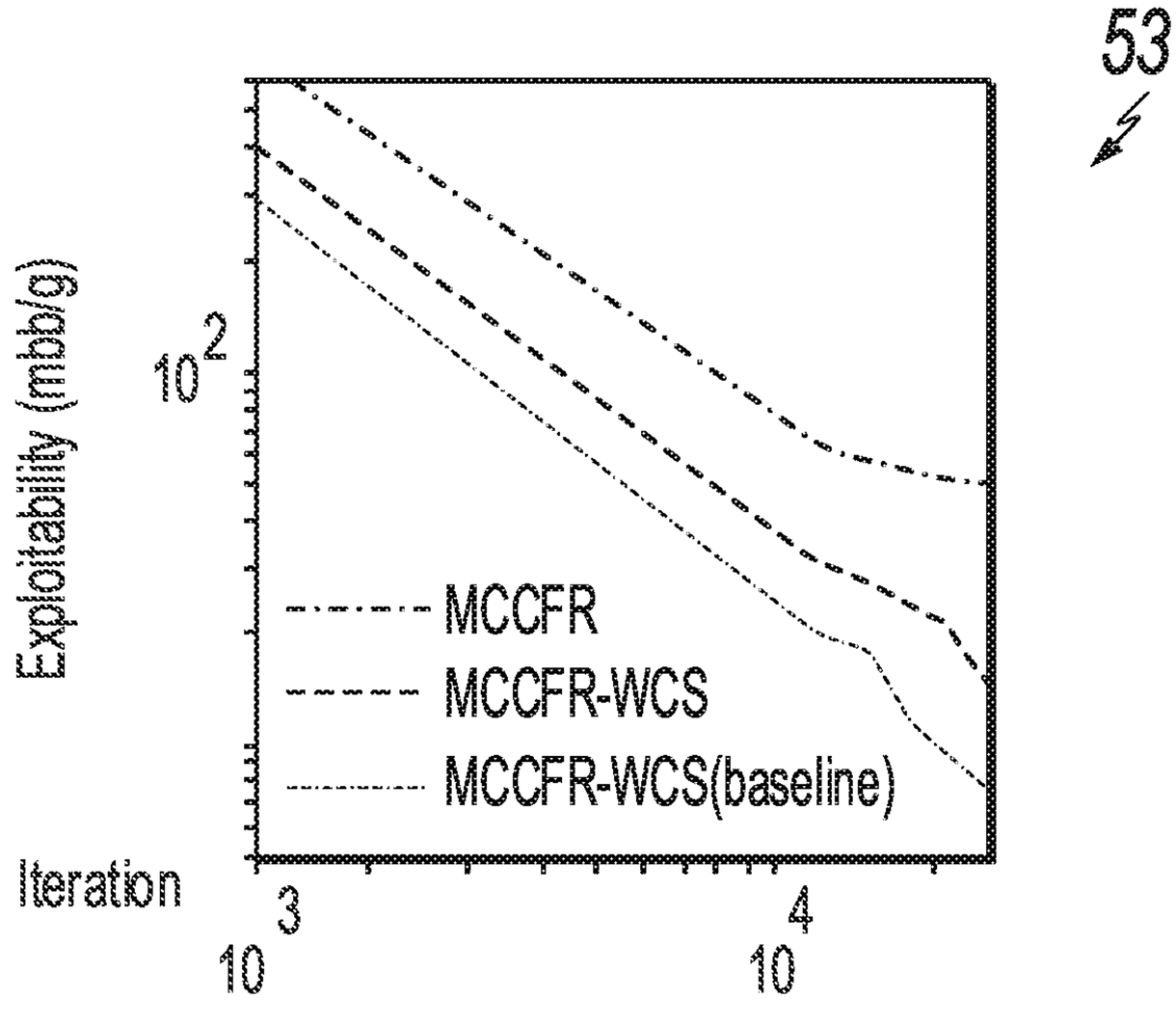

530
Exploitability (mbb/g)
10^2
MCCFR
MCCFR-WCS
MCCFR-WCS(baseline)
Iteration
10^3
10^4

FIG. 5B

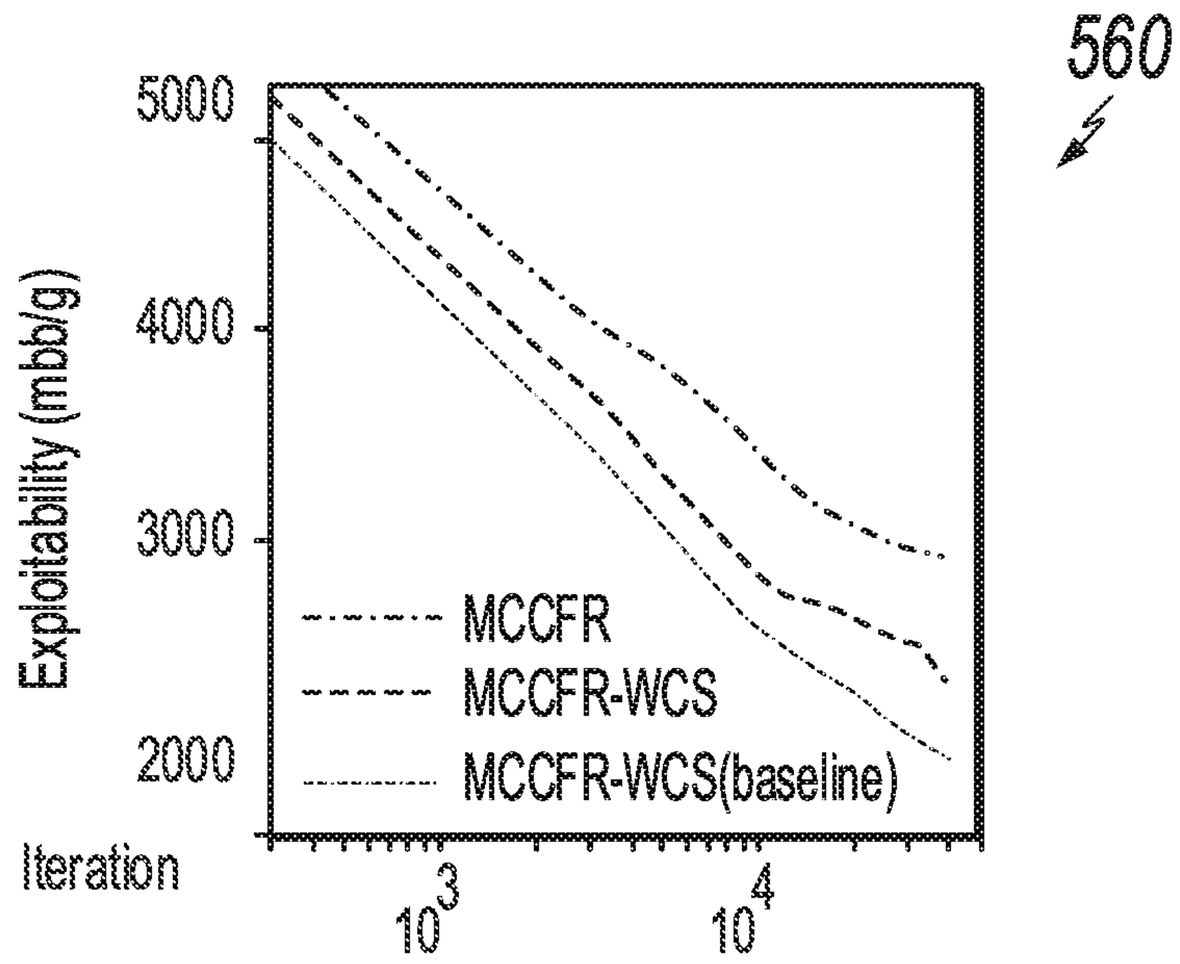

560
5000
4000
3000
2000
Exploitability (mbb/g)
Iteration
10^3
10^4
MCCFR
MCCFR-WCS
MCCFR-WCS(baseline)

FIG. 5C

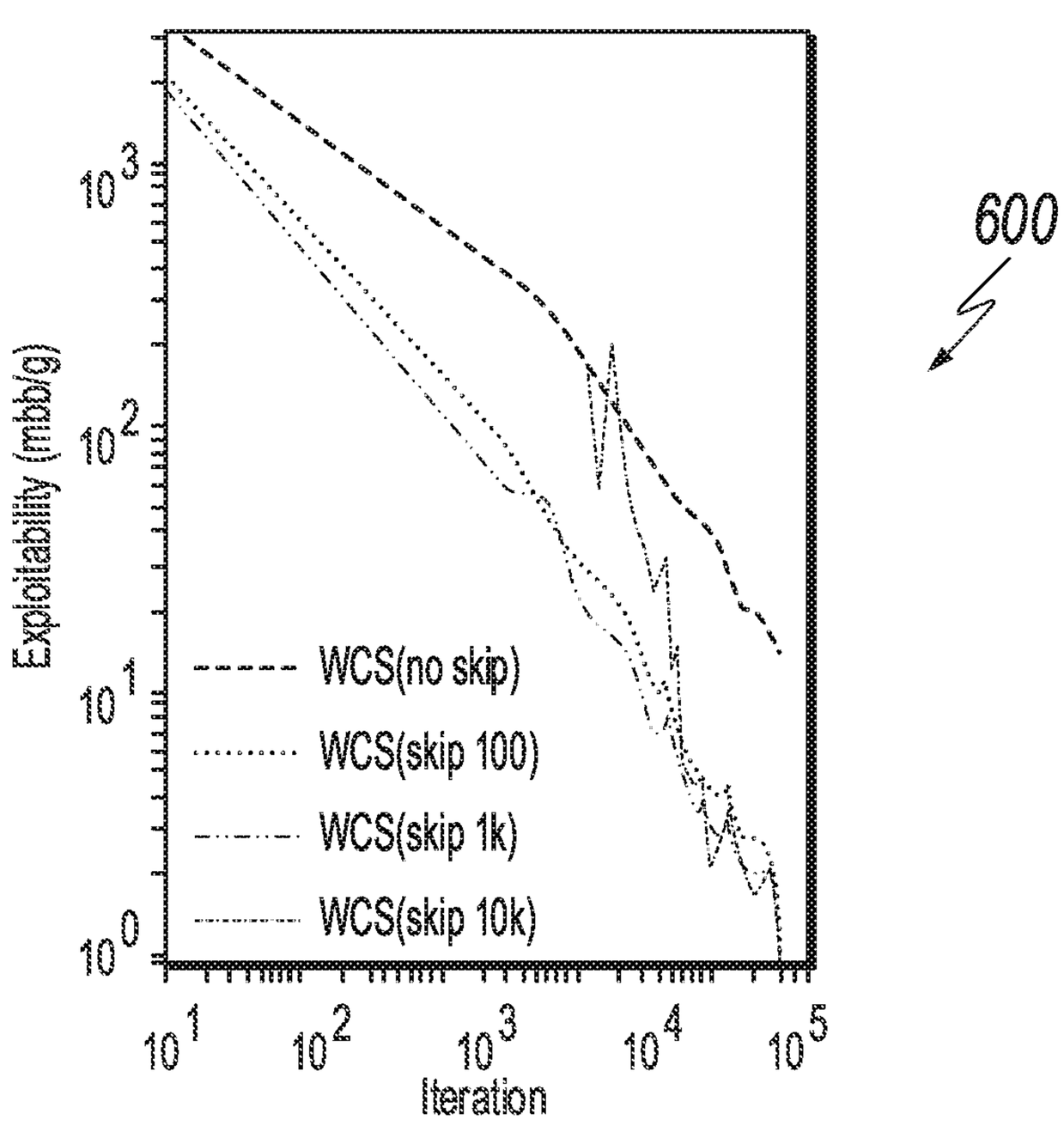

600
Exploitability (mbb/g)
10^3
10^2
10^1
10^0
WCS(no skip)
WCS(skip 100)
WCS(skip 1k)
WCS(skip 10k)
10^1
10^2
10^3
10^4
10^5
Iteration

FIG. 6A

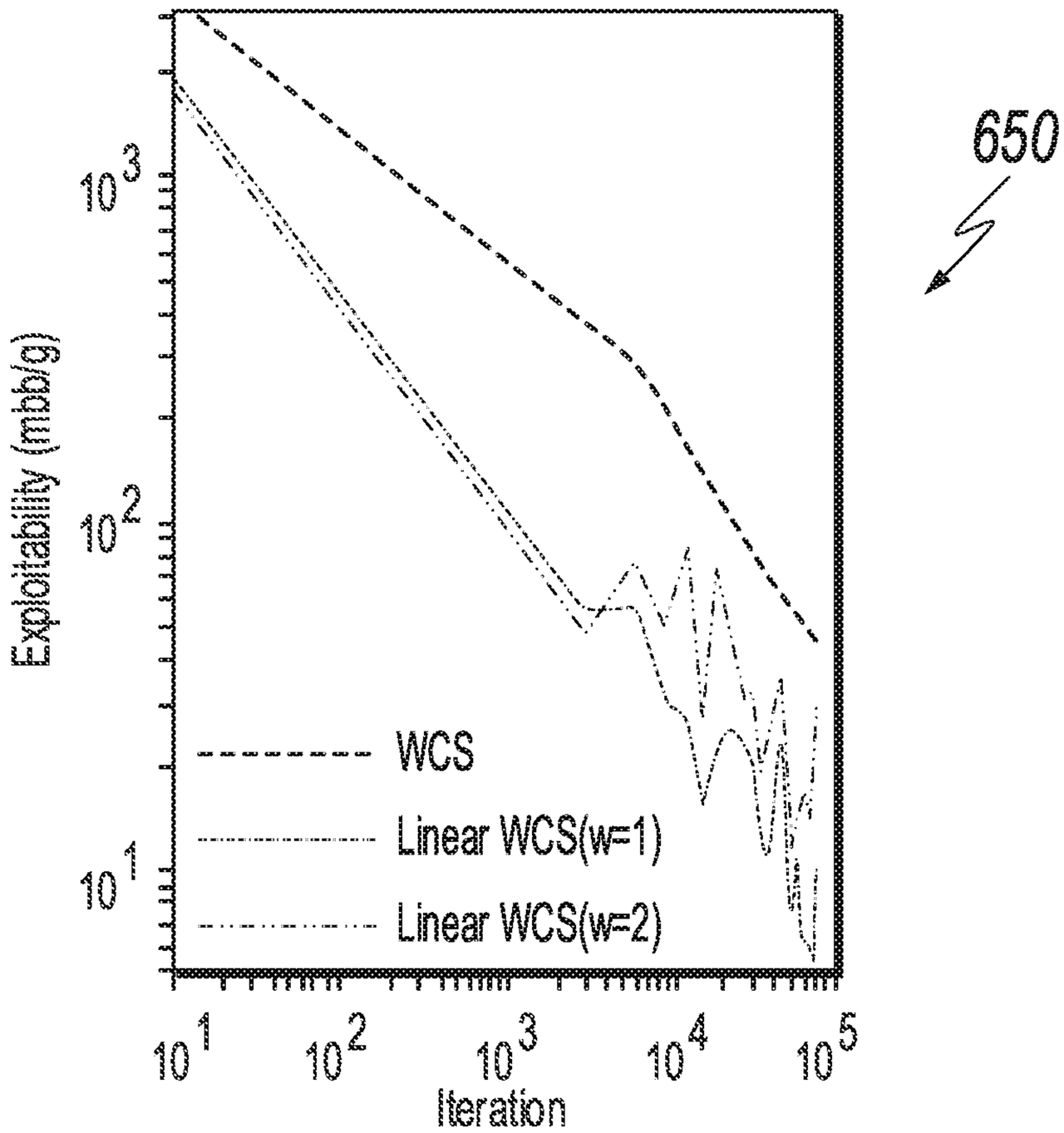

650
Exploitability (mbb/g)
10^3
10^2
10^1
WCS
Linear WCS(w=1)
Linear WCS(w=2)
10^1
10^2
10^3
10^4
10^5
Iteration

FIG. 6B

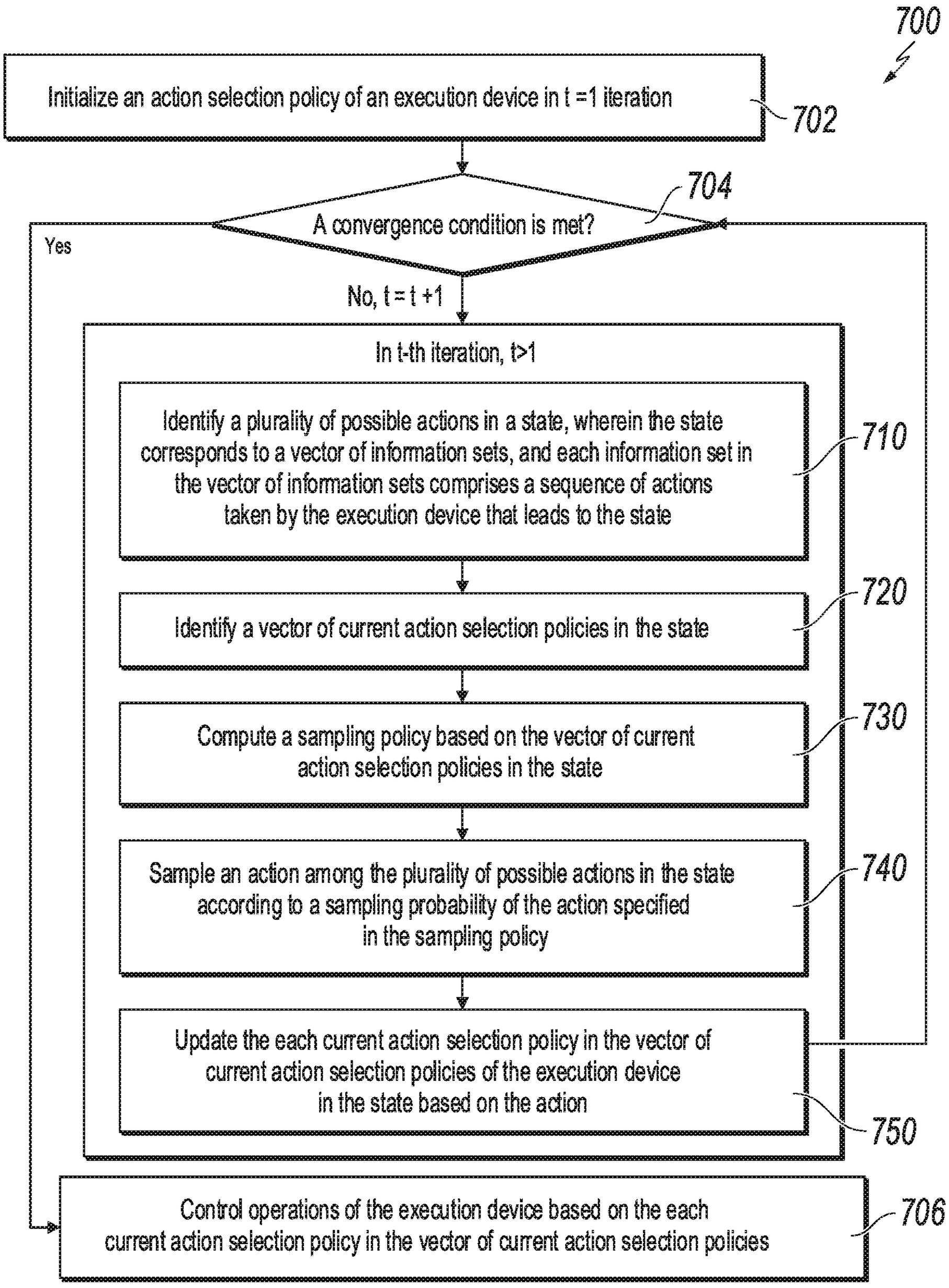

700
Initialize an action selection policy of an execution device in t =1 iteration
702
A convergence condition is met?
704
Yes
No, t = t +1
In t-th iteration, t>1
Identify a plurality of possible actions in a state, wherein the state corresponds to a vector of information sets, and each information set in the vector of information sets comprises a sequence of actions taken by the execution device that leads to the state
710
Identify a vector of current action selection policies in the state
720
Compute a sampling policy based on the vector of current action selection policies in the state
730
Sample an action among the plurality of possible actions in the state according to a sampling probability of the action specified in the sampling policy
740
Update the each current action selection policy in the vector of current action selection policies of the execution device in the state based on the action
750
Control operations of the execution device based on the each current action selection policy in the vector of current action selection policies
706

FIG. 7

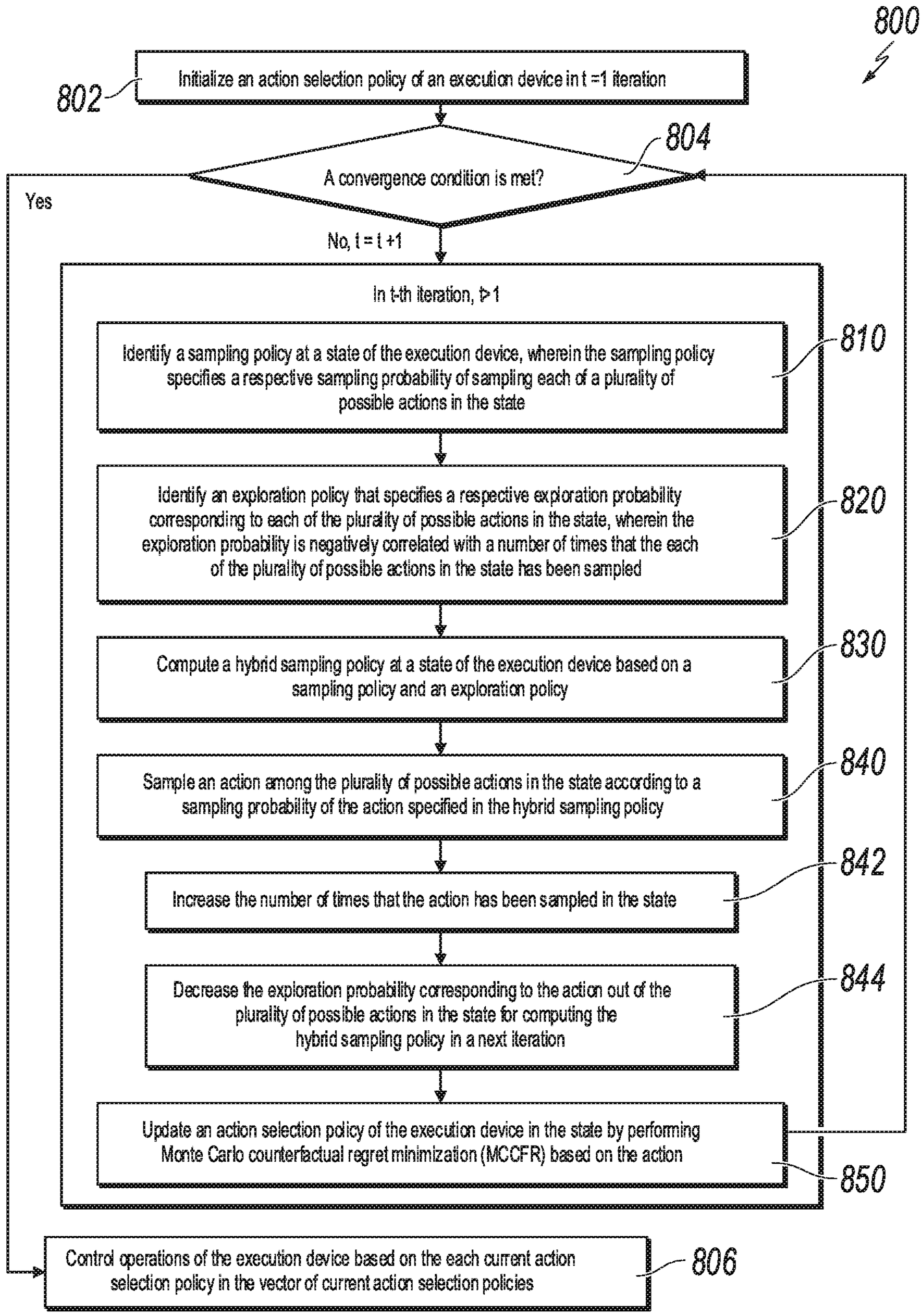

800
802
Initialize an action selection policy of an execution device in t =1 iteration
804
A convergence condition is met?
Yes
No, t = t +1
In t-th iteration, t>1
810
Identify a sampling policy at a state of the execution device, wherein the sampling policy specifies a respective sampling probability of sampling each of a plurality of possible actions in the state
820
Identify an exploration policy that specifies a respective exploration probability corresponding to each of the plurality of possible actions in the state, wherein the exploration probability is negatively correlated with a number of times that the each of the plurality of possible actions in the state has been sampled
830
Compute a hybrid sampling policy at a state of the execution device based on a sampling policy and an exploration policy
840
Sample an action among the plurality of possible actions in the state according to a sampling probability of the action specified in the hybrid sampling policy
842
Increase the number of times that the action has been sampled in the state
844
Decrease the exploration probability corresponding to the action out of the plurality of possible actions in the state for computing the hybrid sampling policy in a next iteration
850
Update an action selection policy of the execution device in the state by performing Monte Carlo counterfactual regret minimization (MCCFR) based on the action
806
Control operations of the execution device based on the each current action selection policy in the vector of current action selection policies

FIG. 8

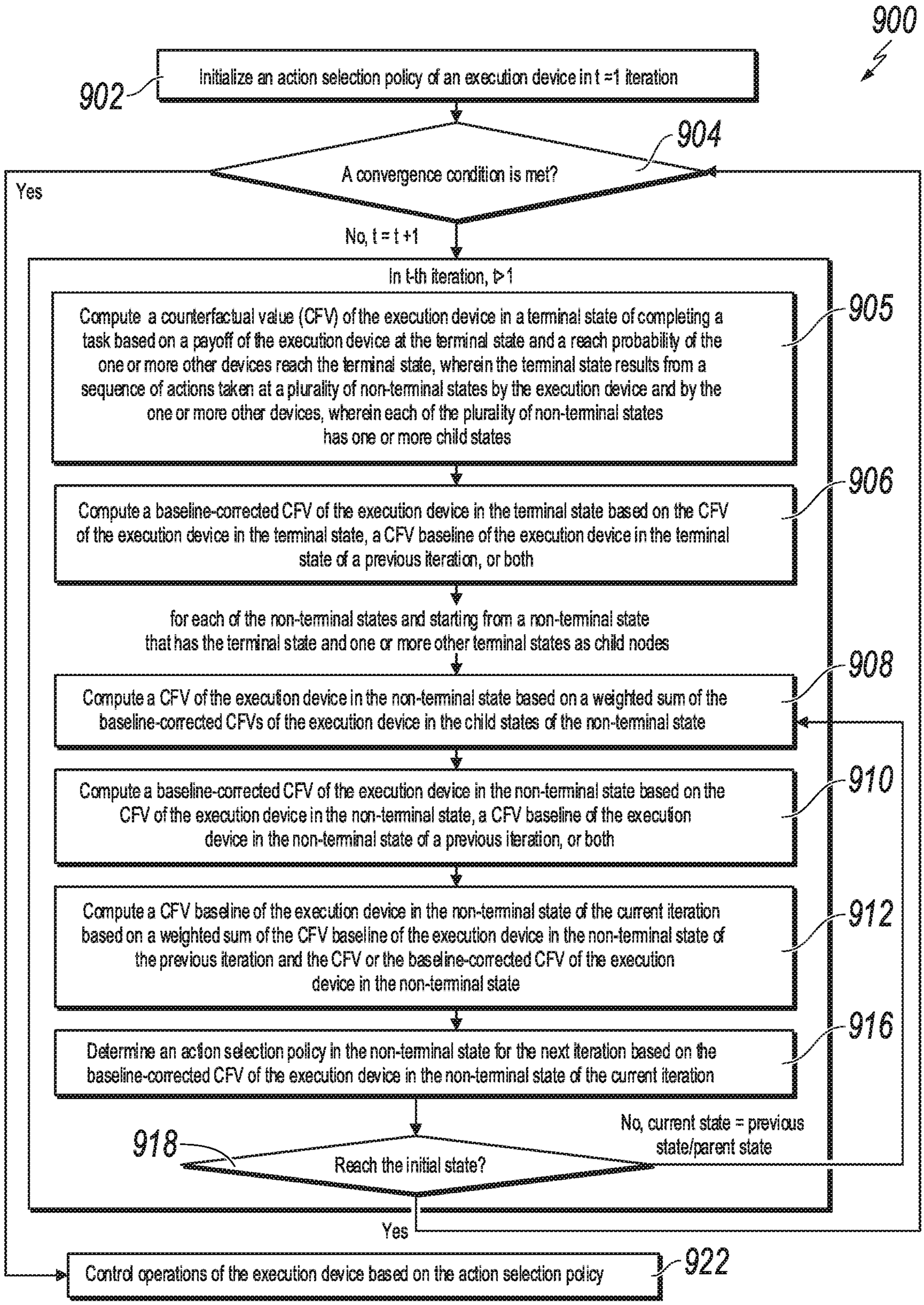

900
Initialize an action selection policy of an execution device in t =1 iteration
902
A convergence condition is met?
904
Yes
No, t = t +1
In t-th iteration, t>1
Compute a counterfactual value (CFV) of the execution device in a terminal state of completing a task based on a payoff of the execution device at the terminal state and a reach probability of the one or more other devices reach the terminal state, wherein the terminal state results from a sequence of actions taken at a plurality of non-terminal states by the execution device and by the one or more other devices, wherein each of the plurality of non-terminal states has one or more child states
905
Compute a baseline-corrected CFV of the execution device in the terminal state based on the CFV of the execution device in the terminal state, a CFV baseline of the execution device in the terminal state of a previous iteration, or both
906
for each of the non-terminal states and starting from a non-terminal state that has the terminal state and one or more other terminal states as child nodes
Compute a CFV of the execution device in the non-terminal state based on a weighted sum of the baseline-corrected CFVs of the execution device in the child states of the non-terminal state
908
Compute a baseline-corrected CFV of the execution device in the non-terminal state based on the CFV of the execution device in the non-terminal state, a CFV baseline of the execution device in the non-terminal state of a previous iteration, or both
910
Compute a CFV baseline of the execution device in the non-terminal state of the current iteration based on a weighted sum of the CFV baseline of the execution device in the non-terminal state of the previous iteration and the CFV or the baseline-corrected CFV of the execution device in the non-terminal state
912
Determine an action selection policy in the non-terminal state for the next iteration based on the baseline-corrected CFV of the execution device in the non-terminal state of the current iteration
916
Reach the initial state?
918
No, current state = previous state/parent state
Yes
Control operations of the execution device based on the action selection policy
922

FIG. 9

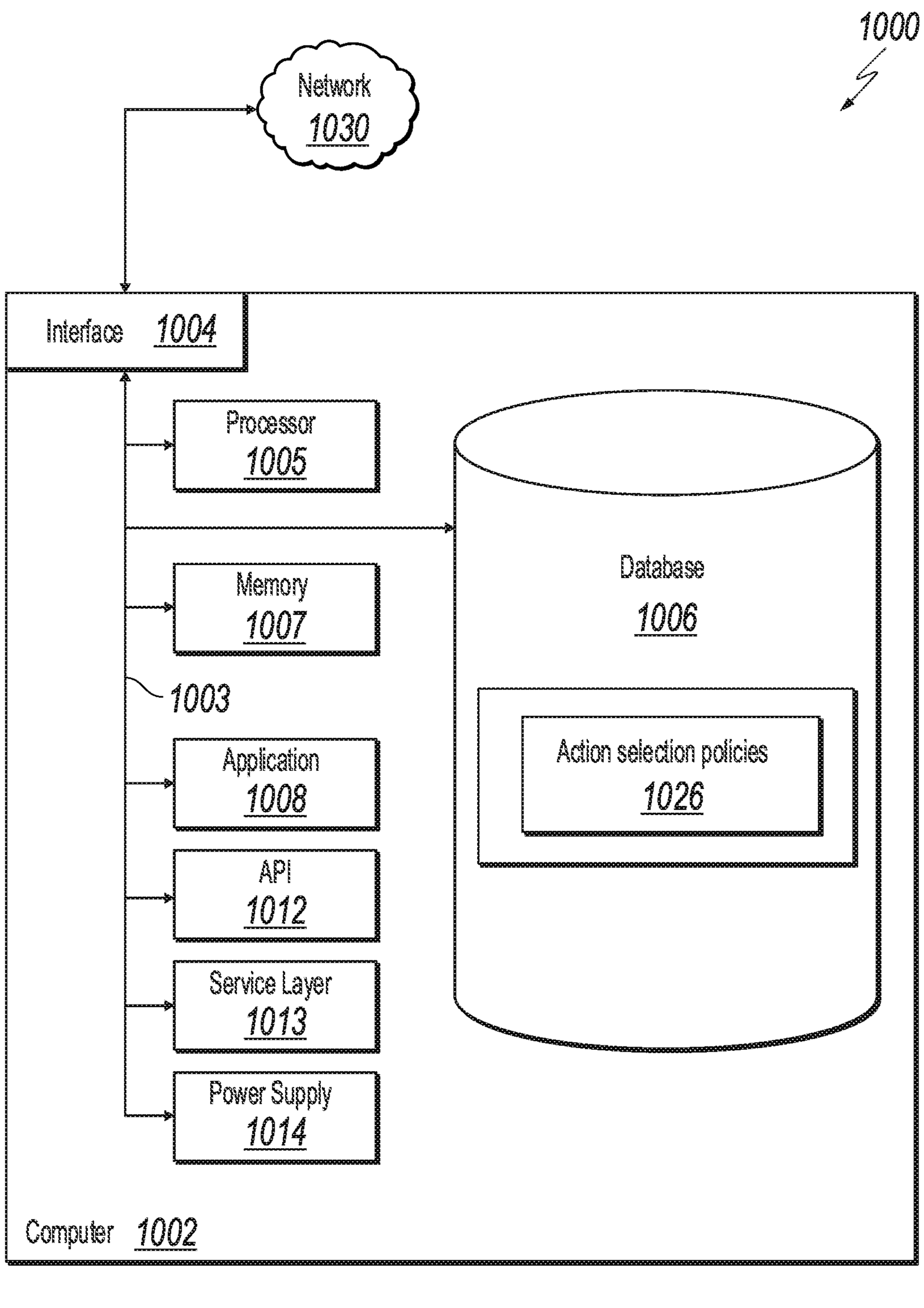

1000
Network
1030
Interface
1004
Processor
1005
Memory
1007
1003
Application
1008
API
1012
Service Layer
1013
Power Supply
1014
Database
1006
Action selection policies
1026
Computer
1002

FIG. 10

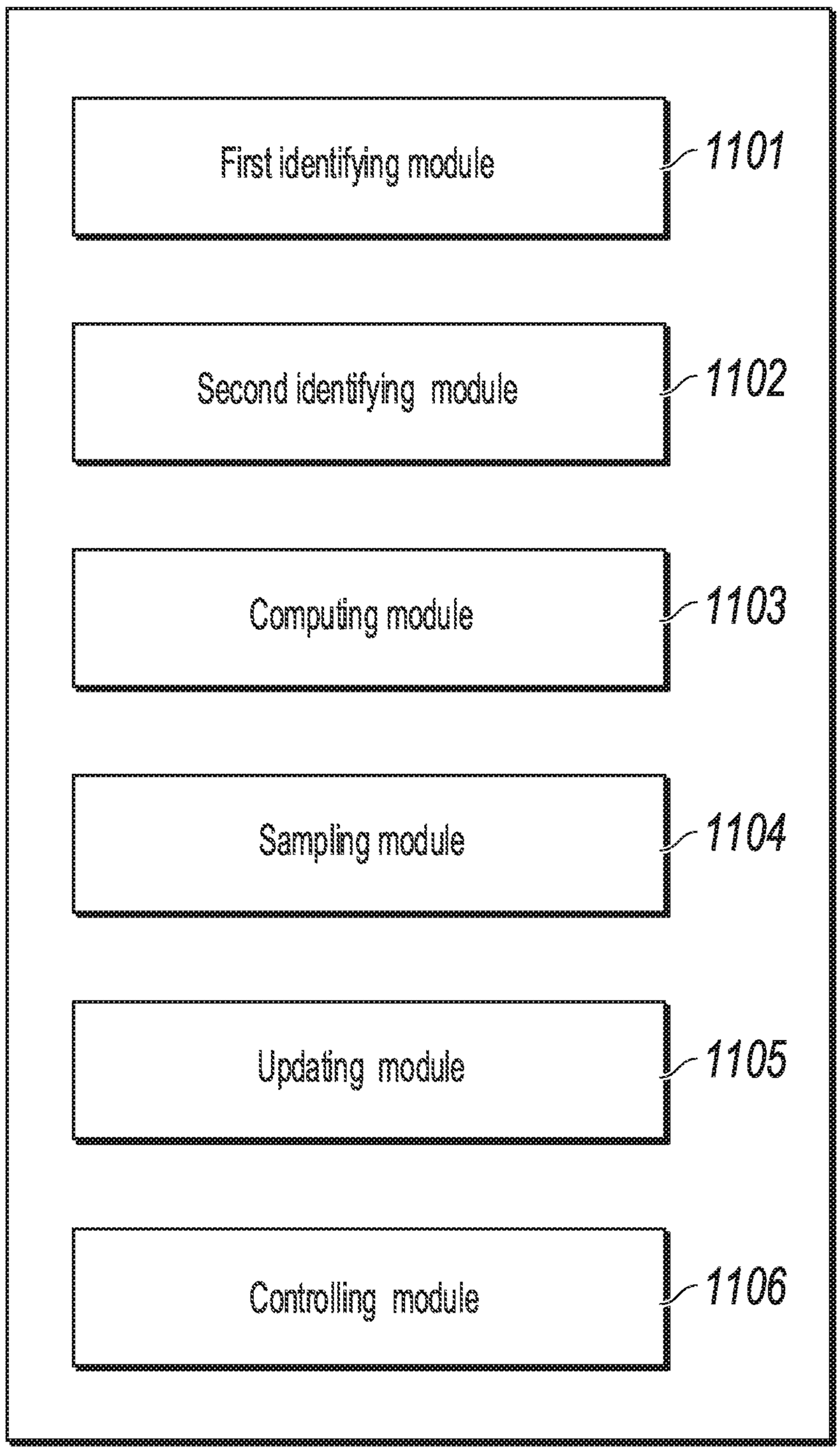

1100
First identifying module
1101
Second identifying module
1102
Computing module
1103
Sampling module
1104
Updating module
1105
Controlling module
1106

FIG. 11

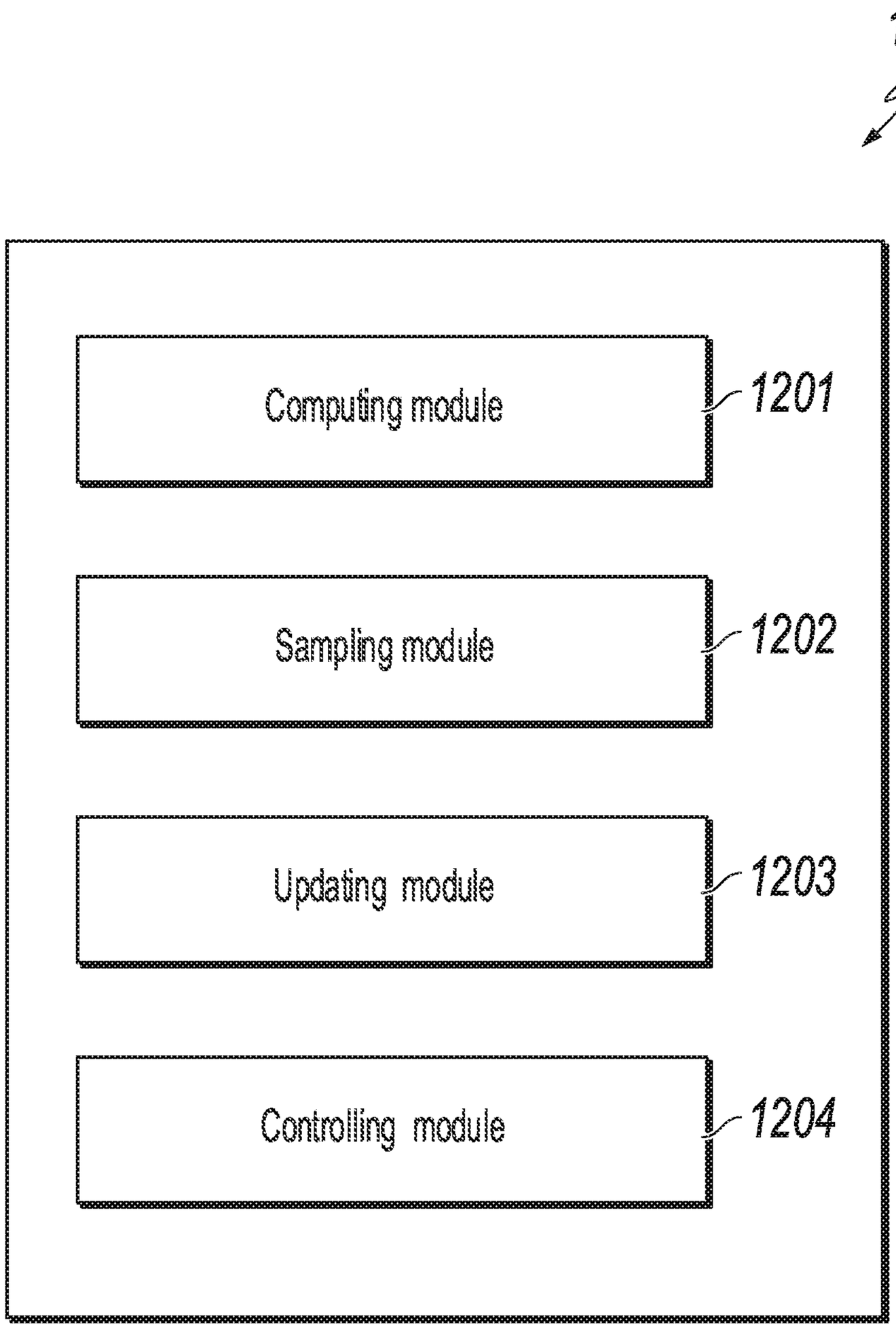

1200
Computing module
1201
Sampling module
1202
Updating module
1203
Controlling module
1204

FIG. 12

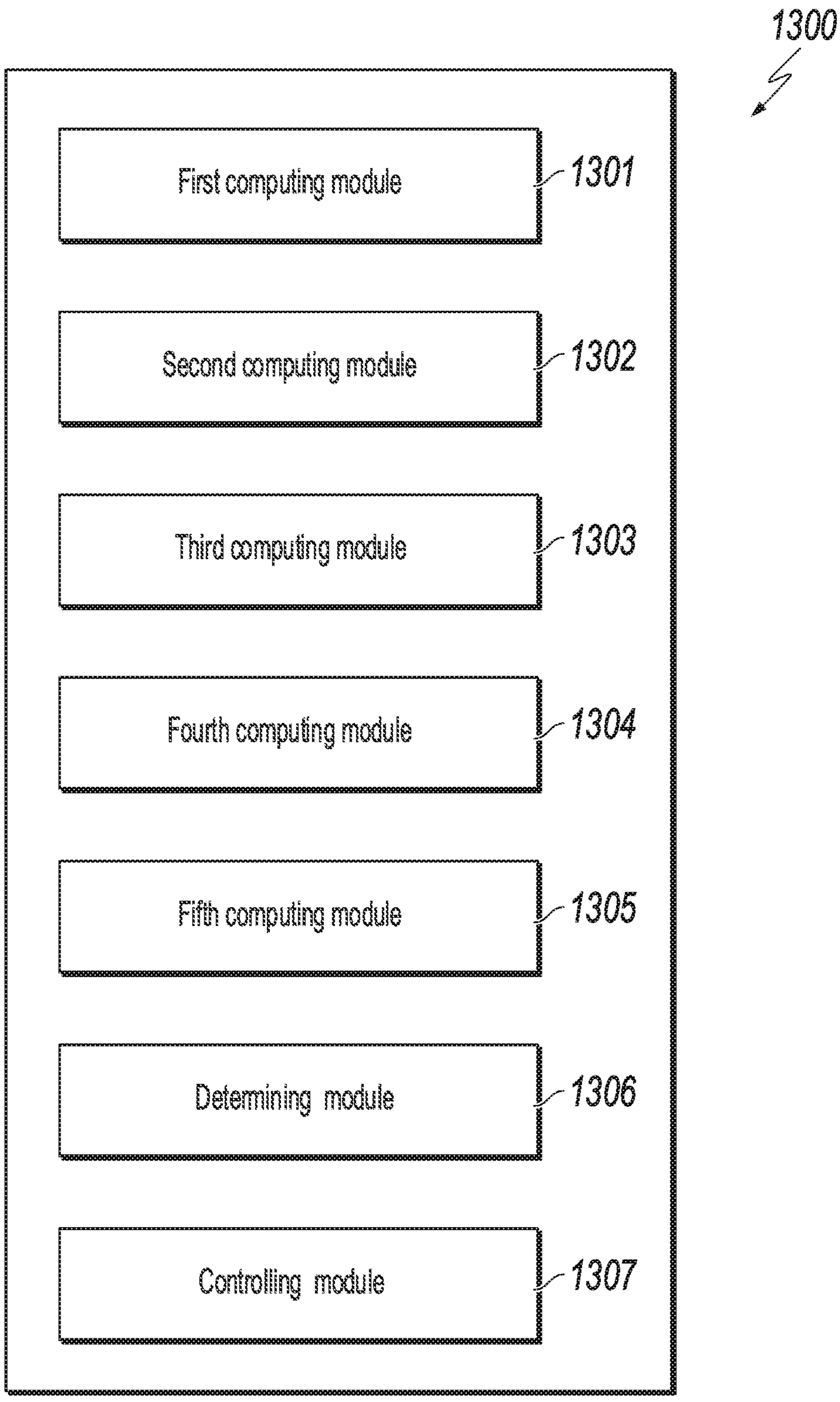

1300
First computing module
1301
Second computing module
1302
Third computing module
1303
Fourth computing module
1304
Fifth computing module
1305
Determining module
1306
Controlling module
1307

FIG. 13