強誘電体トンネル接合における確率的コンダクタンス変化を用いたインメモリ強化学習ハードウェア

In-memory Reinforcement Learning Hardware with Stochastic Conductance Change of Ferroelectric Tunnel Junctions

東芝研開セ [1]，東芝メモリ [2]　○ベルダン ラドゥ [1]，丸亀孝生 [1]，太田健介 [2]，
齋藤真澄 [2]，藤井章輔 [2]，出口淳 [2]，西義史 [1]
Toshiba RDC [1], Toshiba Memory Corporation [2], ○Radu Berdan[1], Takao Marukame[1], Kensuke Ota[2], Masumi Saitoh[2], Shosuke Fujii[2], Jun Deguchi[2] and Yoshifumi Nishi[1]
E-mail: radu1.berdan@toshiba.co.jp

Building compact and efficient reinforcement learning (RL) systems for mobile deployment requires departure from the von-Neumann computing architecture and embracing novel in-memory computing, and local learning paradigms [1]. We exploit nano-scale ferroelectric tunnel junction (FTJ) memristors with inherent analogue stochastic switching arranged in selector-less crossbars to demonstrate an analogue in-memory RL system (Fig. 1). That is, via a hardware-friendly algorithm, capable of learning behavior policies. We show that commonly undesirable stochastic conductance switching is actually, in moderation, a beneficial property which promotes policy finding via a process akin to random search. We experimentally demonstrate path-finding based on reinforcement (Fig. 2), and solve a standard control problem of balancing a pole on a cart via simulation, outperforming similar deterministic RL systems [1].

[1] R. Berdan et al., VLSI Tech., 2019.

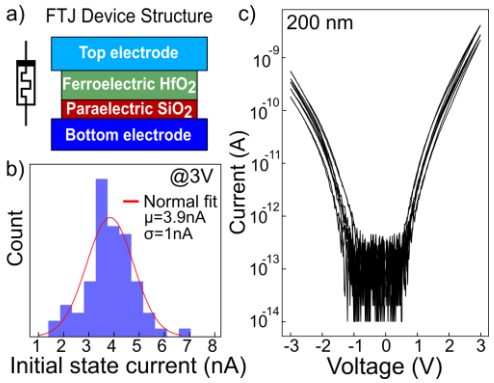

Fig. 1 a) FTJ device structure b) Initial state distributions. Non-zero distributions support random initialisations of state-action pairs for RL. c) Typical initial state FTJ device low-voltage ultra-low current *I-V* curves (5 devices). Intrinsic nonlinearity supports selector-less crossbar operation.

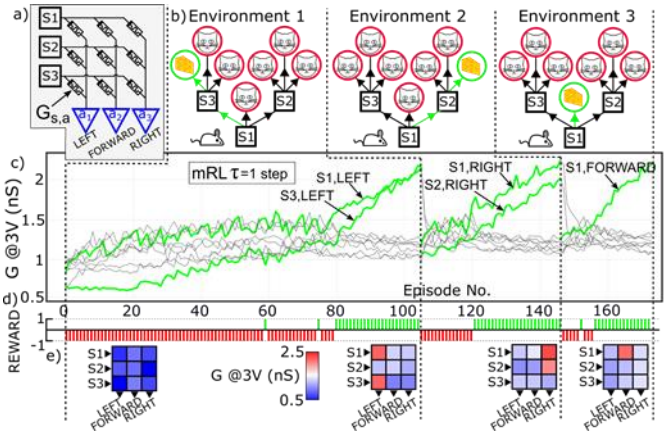

Fig. 2 Experimental path-finding demonstration in a 3x3 FTJ memristor crossbar with changing environment conditions. The mouse needs to get to the positive reward outcome and avoid the negative reward outcome. a) Crossbar implementation b) Three different environments. c) State-action pair conductance evolution. d) Reward history. e) Learned conductance map of the FTJ crossbar after each environment.

© 2019年 応用物理学会