Gradient descent with memontum using dynamic stochastic computing

Siting Liu and Warren J. Gross, Department of Electrical and Computer Engineering, McGill University
McGill Edge Intelligence Lab (MEIL)

Abstract

- A dynamic stochastic computing (DSC) circuit is proposed to perform efficient weight update for neural networks.
- The circuit provides estimates of the weights optimized by gradient descent with momentum.
- It is then deployed in training VGG16, ResNet18 and MobileNetV2 on CIFAR10 dataset and obtains a similar test accuracy compared to their floating-point counterparts.

Motivations

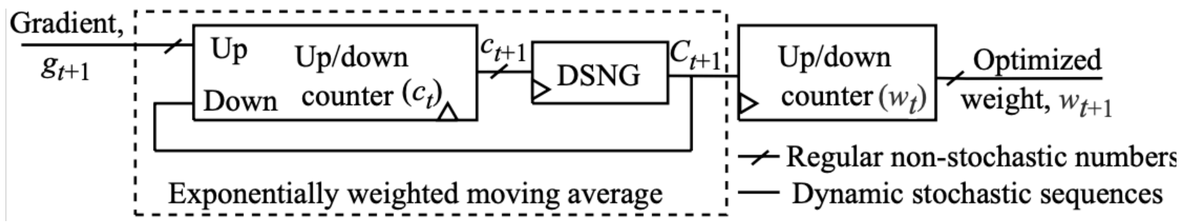
- The computation and energy for training increase as the size of machine learning models continues to grow for more complex problems.
- The demand for customizable machine learning models is emerging. However, due to the sensitivity of some personal data, learning processes that involve these data are preferred to be run locally to ensure their privacy.
- For edge devices running these models, they are generally required to be energy-efficient, low-cost and have a low communication overhead. Efficient training hardware is then essential for these types of applications.

DSC circuit for GDM

- Gradient descent with momentum (GDM)

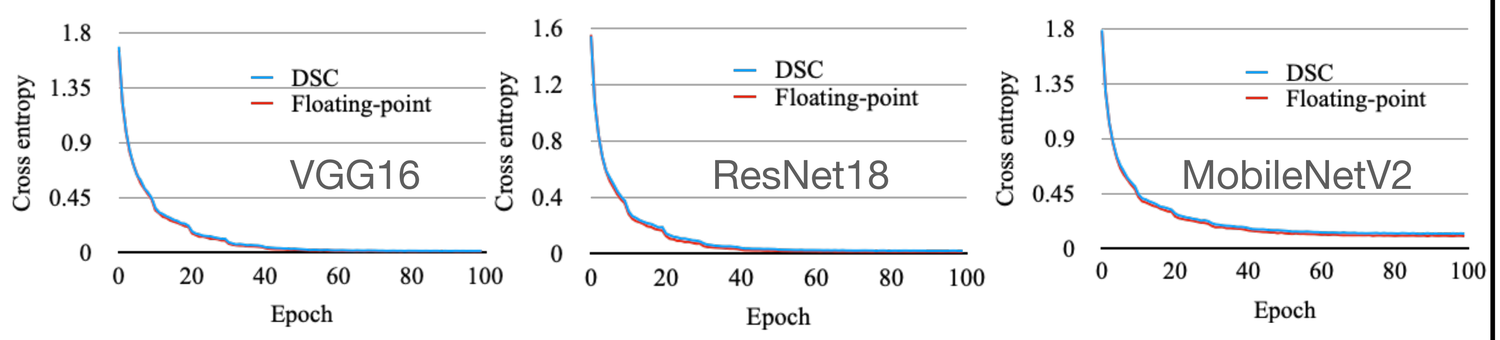
$$v_{t+1} = mv_t + g_{t+1} \qquad \theta_{t+1} = \theta_t - \mu v_{t+1}$$

Exponential moving average of g_{t+1} (gradient) — Iterative accumulation of v



- The first counter provides estimate of v_{t+1}; the DSNG generates dynamic stochastic sequence encoding $\{v_t\}$; m is selected to be $1-2^{-l}, l = 0,1,\ldots$, so that the factor m is realized by shifting.
- The second counter is used to accumulate $\{v_t\}$ by accumulating the stochastic sequence encoding signal $\{v_t\}$. For simple hardware implementation, μ is selected to be $2^{-n}, n = 1,2,\ldots$, performed by shifting.

Experiments and results

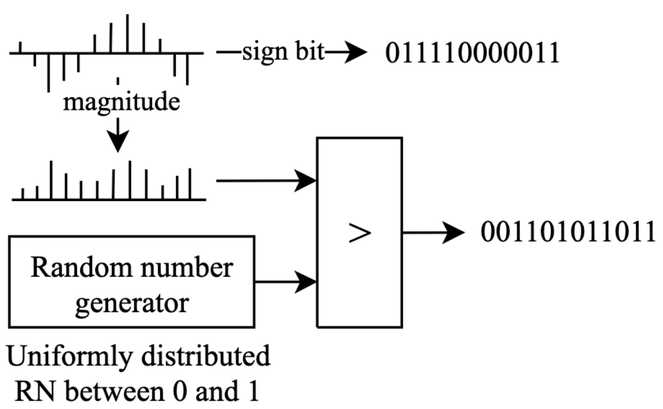


Test Accuracy (%)	VGG16	ResNet18	MobileNetV2
DSC	90.23	91.36	88.51
Floating-point	90.55	91.85	88.82

- The convergence of cross entropy, test accuracy are compared with the floating-point implementation using the same settings.

Dynamic stochastic computing (DSC)

- In DSC, changing signals are encoded by dynamic stochastic sequences consisting of '0's and '1's.
- To deal with both negative and positive numbers, the sign-magnitude representation is used with one bit stream indicates the sign bit and another bit stream for the magnitude.



For the t-th bit B_t and the t-th value in the encoded signal, s_t

$$\mathbb{E}[|B_t|] = |s_t|$$

Dynamic stochastic sequence and its generation

Summary and future work

- A simple dynamic stochastic computing circuit is proposed to perform efficient weight update in training neural networks. It implements the gradient descent with momentum.
- It produces a similar test accuracy compared to a floating-point implementation.
- The hardware efficiency will be evaluated in the future and compared with the other designs.