Gradient descent with momentum using dynamic stochastic computing

Department of Electrical and Computer Engineering, McGill University

McGill Edge Intelligence Lab (MEIL)

Siting Liu and Warren J. Gross

McGill

Outline

- Introduction & background
- Dynamic stochastic computing circuit for gradient descent with momentum
- Experiments & results
- Conclusion & future work

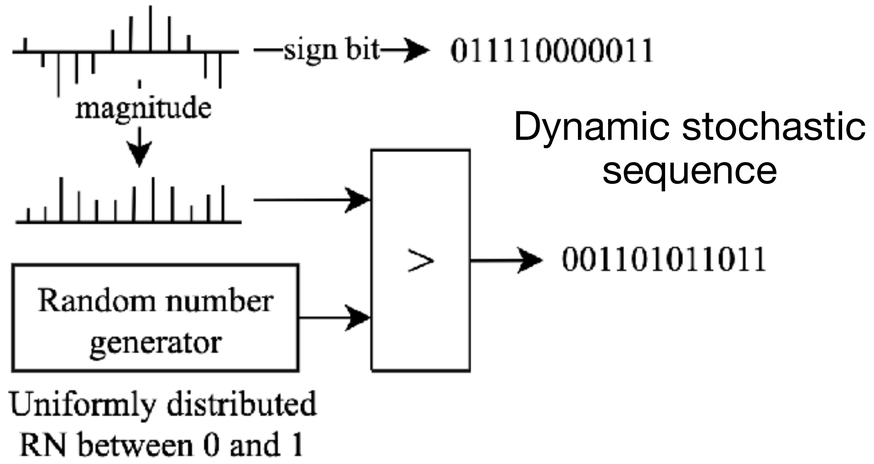
McGill

Introduction & background

- Training gets more complex as machine learning model grows.
- Edge training: personalized machine learning model, superior privacy protection and low communication overhead.
- It is required to be energy-efficient, high-performance and low-cost, especially for edge devices.
- Single-bit based dynamic stochastic computing (DSC) for efficient weight update.

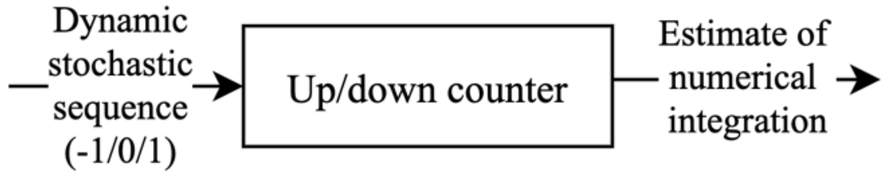
- Dynamic stochastic computing (DSC)

For the t-th stochastic magnitude bit, $\mathbb{E}[|B_t|] = |s_t|$



Sign-magnitude representation
dynamic stochastic number generator (DSNG)

- Complex computations by simple logics



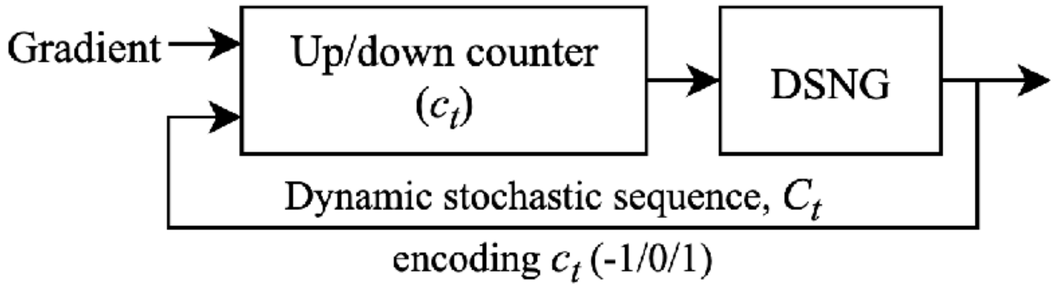
A stochastic integrator.

McGill

DSC circuit for gradient descent with momentum

$$v_{t+1} = mv_t + g_{t+1}$$

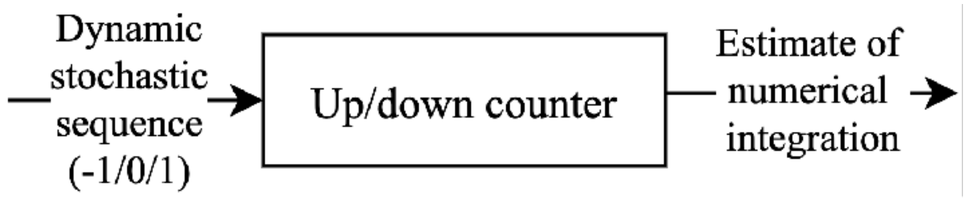
Moving average of g_{t+1} (gradient)



- The above adaptive digital element performs moving average of the input gradients.
- The 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.

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

Iterative accumulation

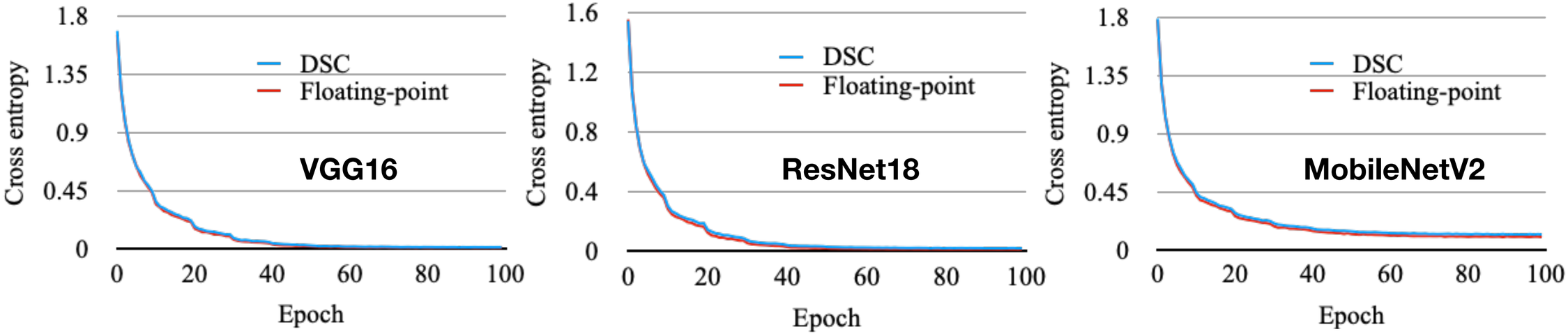


- The stochastic integrator is used to accumulate $\{v_t\}$ by accumulating the stochastic sequence encoding signal $\{v_t\}$. For simple hardware implementation, μ is selected to be $2^{-n}, n = 1,2,\ldots,$ performed by shifting.

McGill

Experiments and results

- The proposed design is tested on training VGG16, ResNet18 and MobileNetV2 on CIFAR10 dataset; learning rate decay by 1/2 after 10 epochs of training starting from 2^{-10}, momentum factor m=1-2^{-6}. The performance is compared with a floating-point implementation.



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

McGill

Conclusion and future work

- A simple dynamic stochastic computing circuit is proposed to perform the weight update in the neural network training.
- It produces a similar test accuracy compared to a floating-point implementation.
- The hardware efficiency will be evaluated in the future and compared with that of SOTA training implementations.

McGill

Thanks for your attention.

Q&A

siting.liu@mail.mcgill.ca, warren.gross@mcgill.ca

McGill