Additive Powers-of-Two Quantization: An Efficient Non-Uniform Discretization for Neural Networks

Yuhang Li*, Xin Dong*, Wei Wang

National University of Singapore, Harvard University

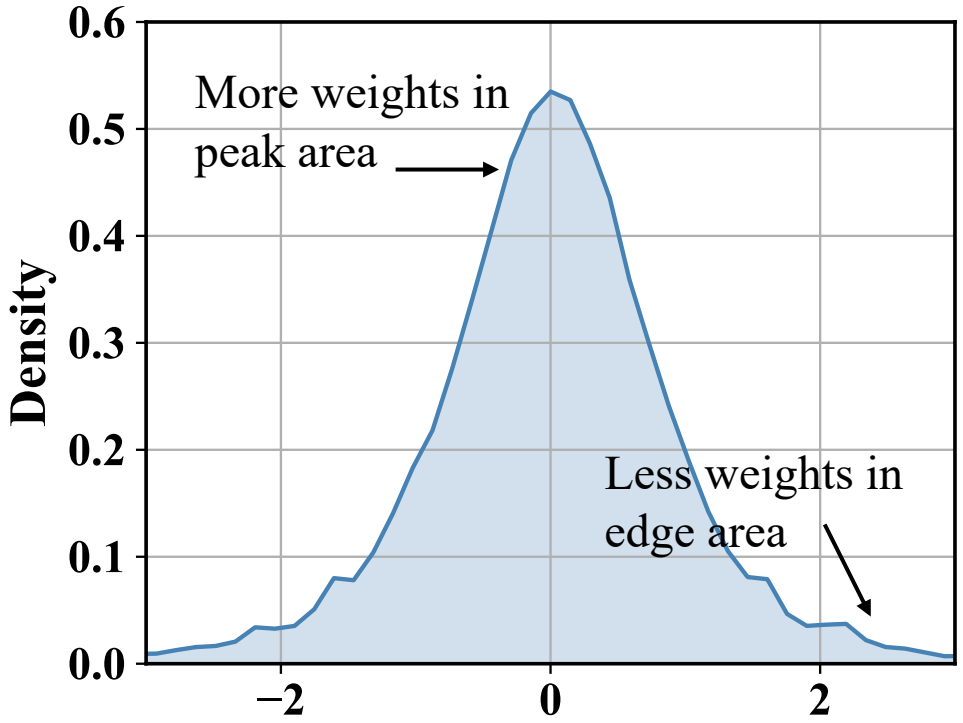
Quantization: Clipping and Projection

Clipping

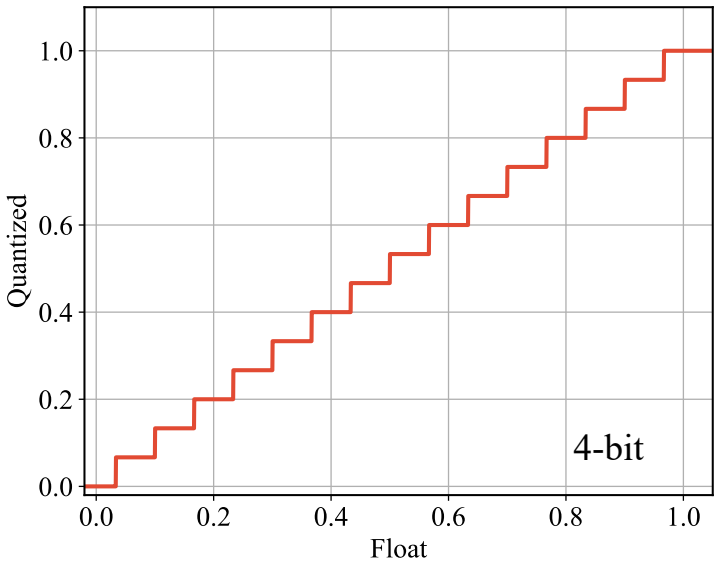
$$\dot{\mathcal{W}} = \text{Clip}(\mathcal{W}, -\alpha, +\alpha)$$

Projection

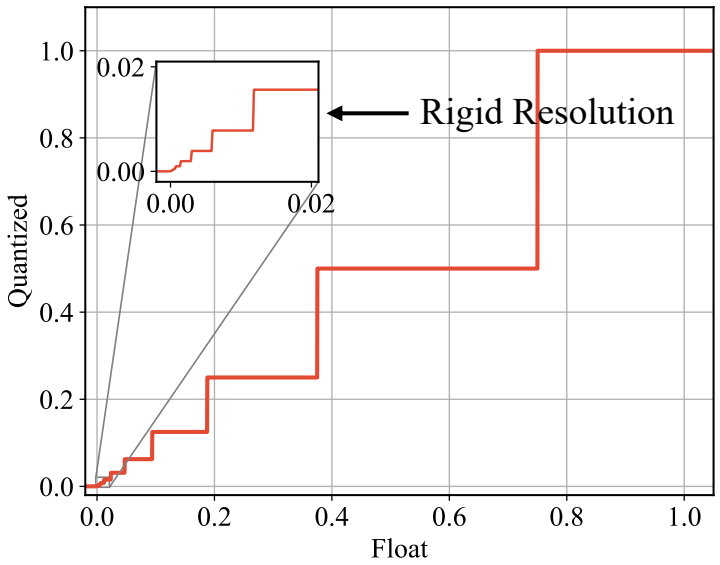
$$\hat{\mathcal{W}} = \Pi_{\mathcal{Q}} \dot{\mathcal{W}}$$



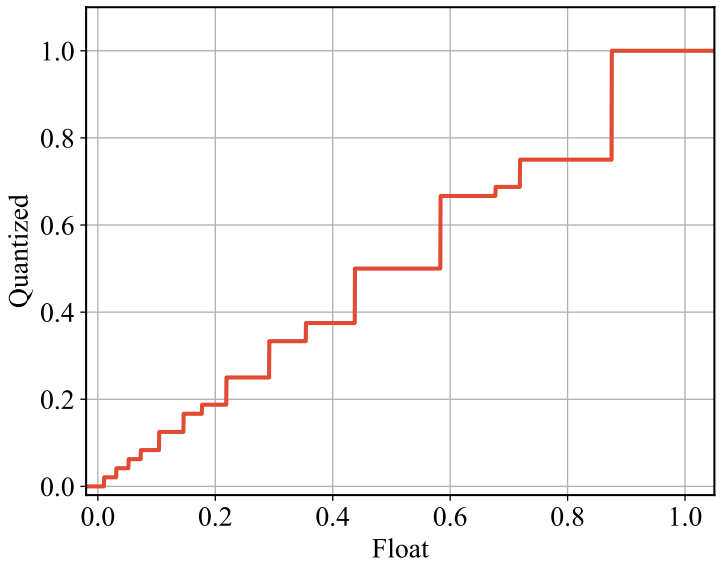
Uniform and Powers-of-Two Projection



Uniform Projection



Powers-of-Two Projection



Additive PoT Projection

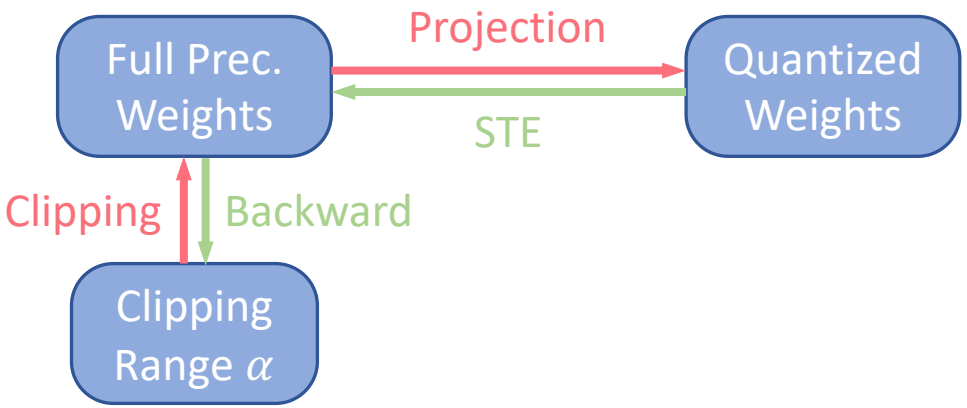
Generalized Definition

$$q = \underbrace{2^{x} + 2^{y} + 2^{z}}$$

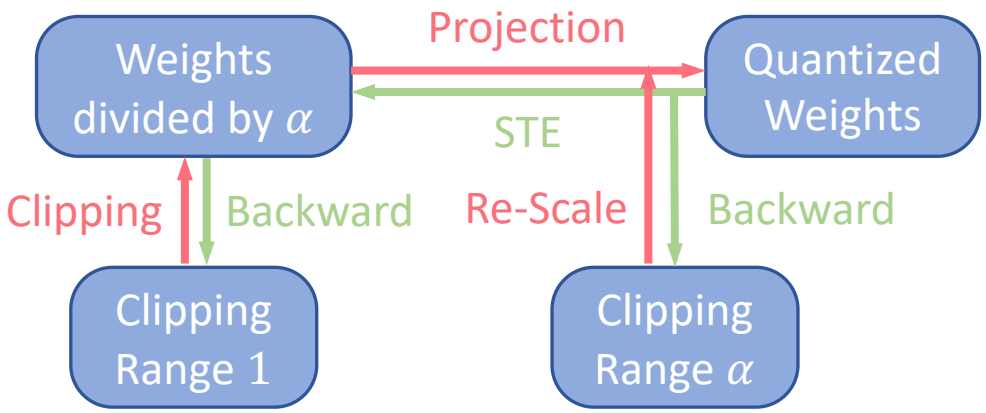
n additive terms, each of which has k-bit values

- Each quantization levels is viewed as n additive Power-of-Two terms
- Tuning k can change the distribution of quantization levels:
  i. When k is 1, Q resembles to uniform quantization
  ii. When k is 2, Q resembles to additive PoT quantization with 2 terms
  iii. When k is b, Q resembles to vanilla PoT quantization

Reparameterized Clipping Function

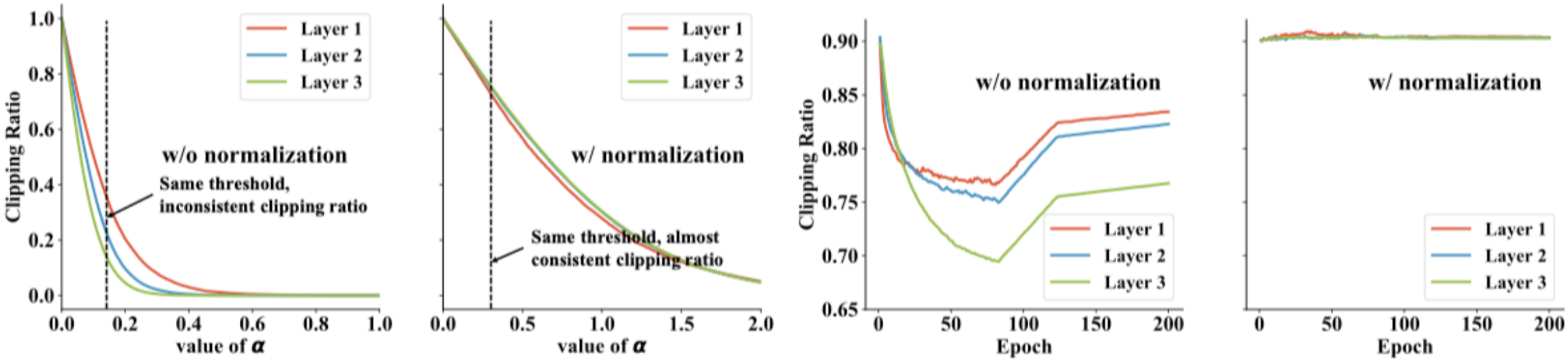


Learning α in PACT



Learning α in RCF

Weights Normalization



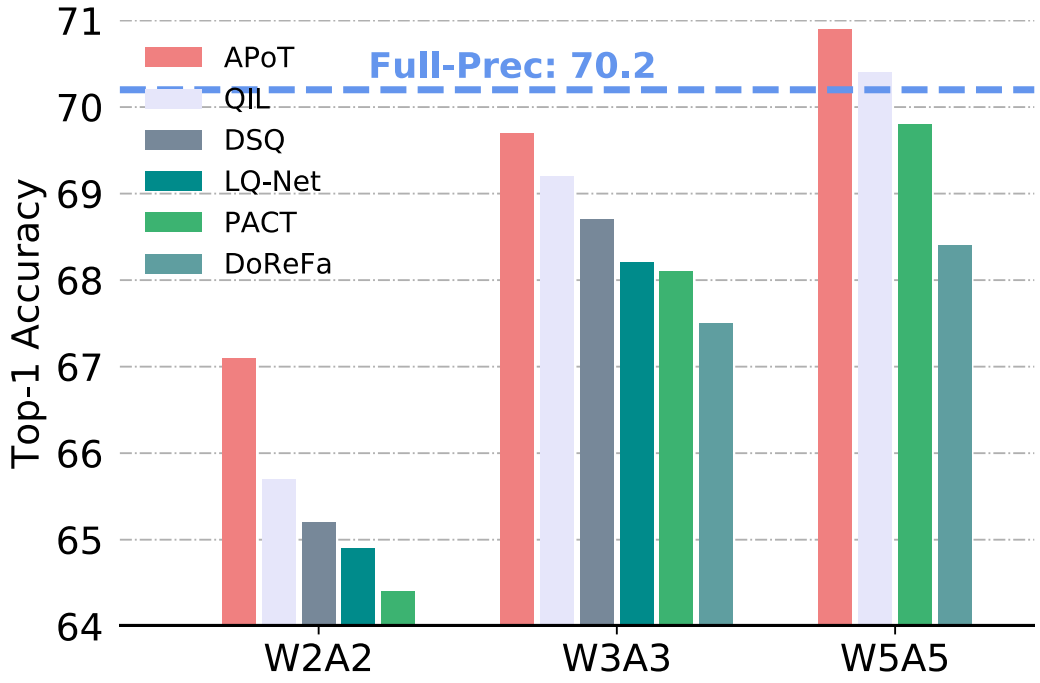
(a) Evolution of clipping ratio with fixed weights

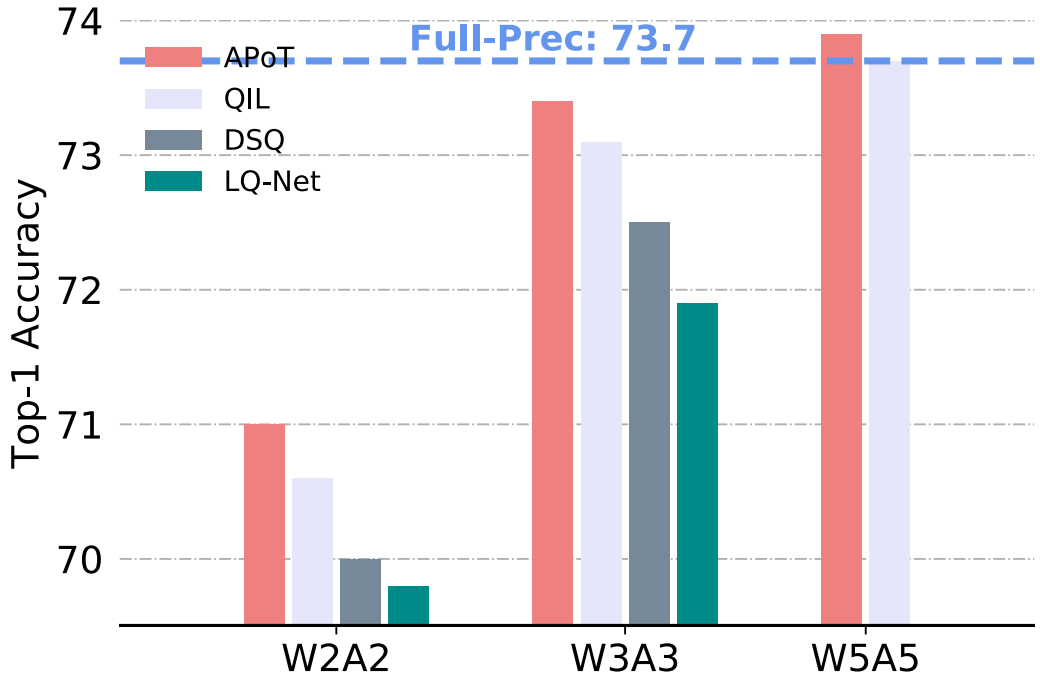
(b) Evolution of clipping ratio with fixed threshold

Forward and Backward Algorithm

- FORWARD
  i. Normalize weights to zero mean and unit variance
  ii. Apply RCF to clip the weights and the activations
  iii. Apply APoT Projection to the weights and the activations
  iv. Quantized Convolution
- BACKWARD
  i. Compute the gradients of weights before normalization
  ii. Compute the gradients of clipping threshold
  iii. Update the weights and the clipping threshold

Results: Compared with Existing Methods





Results: Ablation Study

Table 3: Comparison of quantizer, weight normalization and RCF of ResNet-18 on ImageNet.

Method	Precision	WN	RCF	Acc.-1	RCF	Acc.-1	Model Size	FixOPS
Full Prec.	32 / 32	-	-	70.2	-	70.2	46.8 MB	1.82G
APoT	5 / 5	✓	✓	70.9	✗	70.0	7.22 MB	616M
PoT	5 / 5	✓	✓	70.3	✗	68.9	7.22 MB	582M
Uniform	5 / 5	✓	✓	70.7	✗	69.4	7.22 MB	781M
Lloyd	5 / 5	✓	✓	70.9	✗	70.2	7.22 MB	1.81G
APoT	3 / 3	✓	✓	69.9	✗	68.5	4.56 MB	298M
Uniform	3 / 3	✓	✓	69.4	✗	67.8	4.56 MB	357M
Lloyd	3 / 3	✓	✓	70.0	✗	69.0	4.56 MB	1.81G
APoT	3 / 3	✗	✓	2.0	✗	68.5	4.56 MB	198M

Thank You