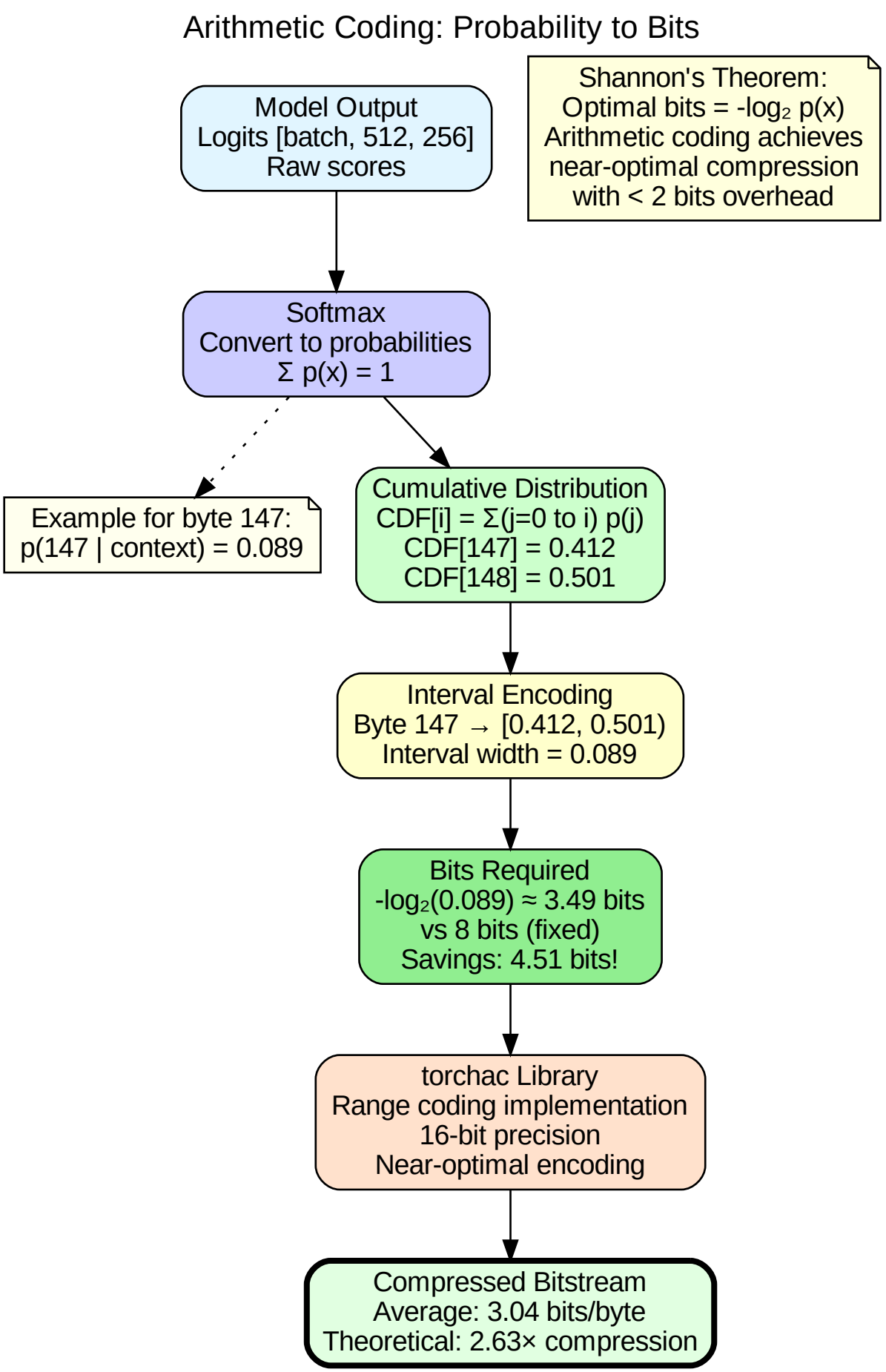
Arithmetic Coding: Probability to Bits
Model Output
Logits [batch, 512, 256]
Raw scores
Shannon's Theorem:
Optimal bits = -log₂ p(x)
Arithmetic coding achieves
near-optimal compression
with < 2 bits overhead
Softmax
Convert to probabilities
Σ p(x) = 1
Example for byte 147:
p(147 | context) = 0.089
Cumulative Distribution
CDF[i] = Σ(j=0 to i) p(j)
CDF[147] = 0.412
CDF[148] = 0.501
Interval Encoding
Byte 147 → [0.412, 0.501)
Interval width = 0.089
Bits Required
-log₂(0.089) ≈ 3.49 bits
vs 8 bits (fixed)
Savings: 4.51 bits!
torchac Library
Range coding implementation
16-bit precision
Near-optimal encoding
Compressed Bitstream
Average: 3.04 bits/byte
Theoretical: 2.63× compression