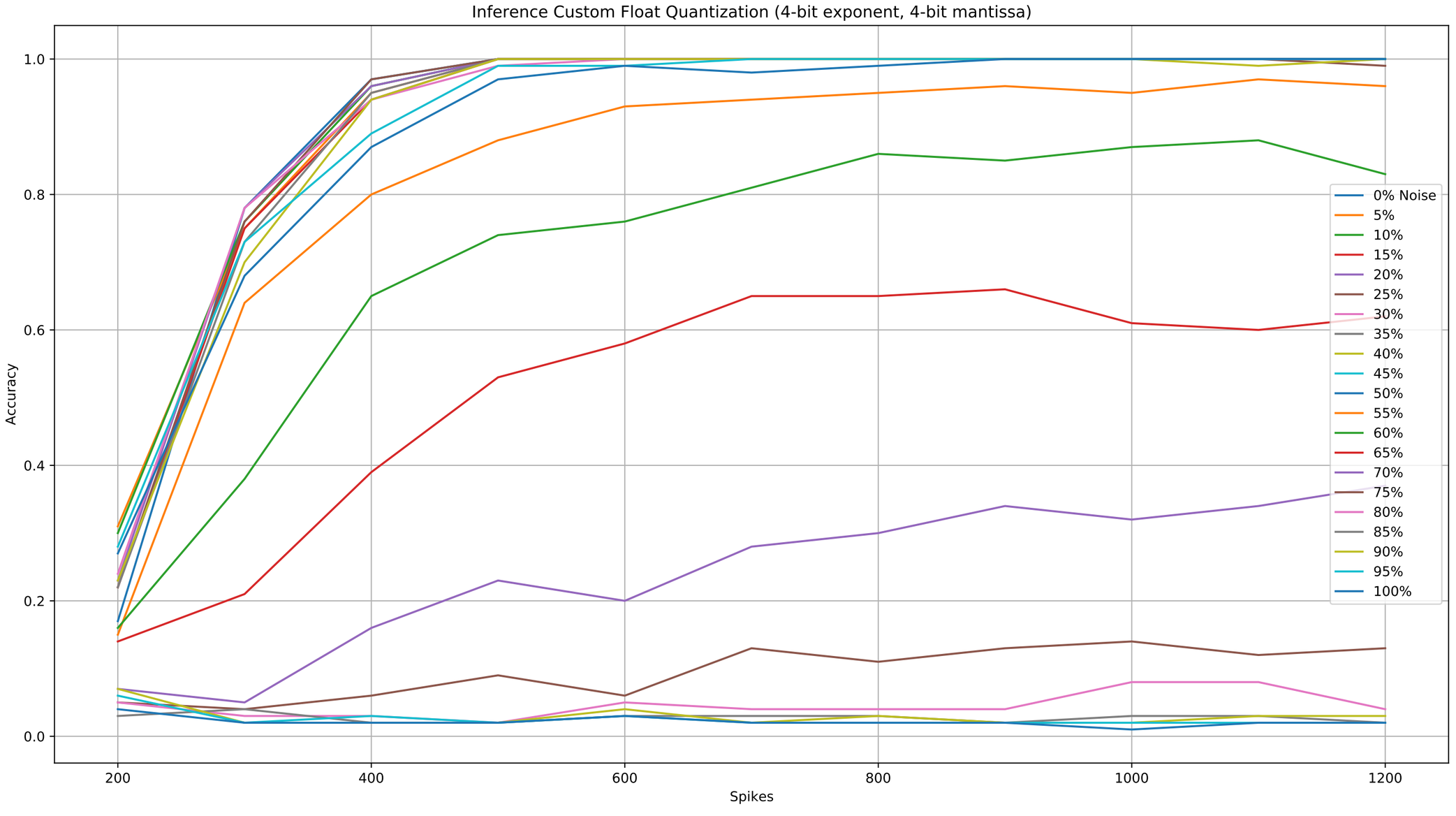
Inference Custom Float Quantization (4-bit exponent, 4-bit mantissa)
Accuracy
Spikes
0.0
0.2
0.4
0.6
0.8
1.0
200
400
600
800
1000
1200
0% Noise
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
55%
60%
65%
70%
75%
80%
85%
90%
95%
100%