

Iteration Time (sec) Benchmark

Measured on a 20 GB MIG slice from an 80 GB NVidia A100.

1 slice

3 slices

6 slices

batch 16

	1 probe	3 probes	6 probes	12 probes
ptyrad	6.32 ± 0.2	6.34 ± 0.05	7.35 ± 0.09	9.88 ± 0.07
ptyshv	6.87 ± 0.4	11.8 ± 0.7	19.2 ± 0.9	34.2 ± 2
py4dstem	7.02 ± 0.3	14.6 ± 0.9	25.8 ± 1	49.1 ± 2

	1 probe	3 probes	6 probes	12 probes
ptyrad	8.62 ± 0.1	9.02 ± 0.08	11.5 ± 0.09	16.1 ± 0.09
ptyshv	12 ± 0.6	25 ± 1	45.5 ± 2	85 ± 3
py4dstem	16.6 ± 0.8	39.2 ± 2	73.3 ± 3	142 ± 5

	1 probe	3 probes	6 probes	12 probes
ptyrad	13.3 ± 0.07	14.6 ± 0.1	19 ± 0.09	26.6 ± 0.09
ptyshv	20.5 ± 1	49.2 ± 2	90.1 ± 5	182 ± 7
py4dstem	31 ± 2	74.2 ± 3	144 ± 7	283 ± 8

batch 32

	1 probe	3 probes	6 probes	12 probes
ptyrad	3.37 ± 0.1	4.03 ± 0.04	5.35 ± 0.1	7.9 ± 0.1
ptyshv	3.75 ± 0.2	6.35 ± 0.4	10.2 ± 0.5	18.2 ± 0.8
py4dstem	3.6 ± 0.2	7.37 ± 0.4	13.3 ± 0.5	25.4 ± 1

	1 probe	3 probes	6 probes	12 probes
ptyrad	5.38 ± 0.09	6.97 ± 0.06	9.3 ± 0.08	13.8 ± 0.1
ptyshv	6.16 ± 0.4	13.1 ± 0.6	23 ± 1	43.7 ± 2
py4dstem	8.37 ± 0.5	20 ± 1	37.6 ± 2	73.5 ± 3

	1 probe	3 probes	6 probes	12 probes
ptyrad	9.47 ± 0.1	12.4 ± 0.1	16.1 ± 0.1	23.6 ± 0.09
ptyshv	10.8 ± 0.6	25.4 ± 1	46 ± 2	90.2 ± 4
py4dstem	15.5 ± 0.8	38.7 ± 1	73.3 ± 3	147 ± 5

batch 64

	1 probe	3 probes	6 probes	12 probes
ptyrad	2.28 ± 0.06	3.19 ± 0.05	4.44 ± 0.04	7 ± 0.06
ptyshv	2.23 ± 0.2	3.76 ± 0.2	6.02 ± 0.4	10.1 ± 0.8
py4dstem	1.84 ± 0.07	3.92 ± 0.2	7.22 ± 0.3	13.8 ± 0.6

	1 probe	3 probes	6 probes	12 probes
ptyrad	4.3 ± 0.06	5.96 ± 0.06	8.19 ± 0.05	12.8 ± 0.1
ptyshv	3.6 ± 0.2	7.48 ± 0.3	13.2 ± 0.8	24.6 ± 1
py4dstem	4.24 ± 0.1	10.5 ± 0.4	20.3 ± 0.8	39.7 ± 2

	1 probe	3 probes	6 probes	12 probes
ptyrad	8.22 ± 0.05	11 ± 0.05	14.7 ± 0.08	22.3 ± 0.1
ptyshv	6.11 ± 0.2	13.9 ± 0.8	25.7 ± 1	49.6 ± 2
py4dstem	7.91 ± 0.3	20.3 ± 0.8	39.7 ± 2	78.9 ± 4

batch 128

	1 probe	3 probes	6 probes	12 probes
ptyrad	1.83 ± 0.03	2.7 ± 0.04	3.97 ± 0.03	6.52 ± 0.03
ptyshv	1.6 ± 0.05	2.65 ± 0.09	4 ± 0.08	6.72 ± 0.2
py4dstem	1.15 ± 0.05	2.36 ± 0.08	4.28 ± 0.1	8.18 ± 0.2

	1 probe	3 probes	6 probes	12 probes
ptyrad	3.78 ± 0.03	5.33 ± 0.05	7.63 ± 0.05	12.1 ± 0.04
ptyshv	2.62 ± 0.02	5.51 ± 0.2	9.51 ± 0.3	17.5 ± 0.8
py4dstem	2.54 ± 0.1	6.3 ± 0.2	12.1 ± 0.4	23.4 ± 0.7

	1 probe	3 probes	6 probes	12 probes
ptyrad	7.57 ± 0.04	10.2 ± 0.05	14 ± 0.05	21.4 ± 0.06
ptyshv	4.39 ± 0.05	9.95 ± 0.2	18.7 ± 0.3	34.9 ± 1
py4dstem	4.59 ± 0.2	12 ± 0.4	23.5 ± 0.8	46.2 ± 2

batch 256

	1 probe	3 probes	6 probes	12 probes
ptyrad	1.59 ± 0.006	2.46 ± 0.006	3.72 ± 0.01	6.26 ± 0.02
ptyshv	1.33 ± 0.06	2 ± 0.07	2.98 ± 0.1	4.85 ± 0.2
py4dstem	0.852 ± 0.03	1.7 ± 0.05	2.97 ± 0.09	5.51 ± 0.2

	1 probe	3 probes	6 probes	12 probes
ptyrad	3.47 ± 0.006	5.04 ± 0.02	7.29 ± 0.02	11.8 ± 0.02
ptyshv	2.15 ± 0.07	4.17 ± 0.1	7.2 ± 0.3	13.4 ± 0.2
py4dstem	1.77 ± 0.03	4.33 ± 0.1	8.11 ± 0.2	15.6 ± 0.4

	1 probe	3 probes	6 probes	12 probes
ptyrad	7.14 ± 0.01	9.74 ± 0.02	13.5 ± 0.02	21 ± 0.03
ptyshv	3.41 ± 0.1	7.54 ± 0.2	13.9 ± 0.4	26.5 ± 0.3
py4dstem	3.11 ± 0.08	8.23 ± 0.3	15.7 ± 0.4	30.7 ± 1

batch 512

	1 probe	3 probes	6 probes	12 probes
ptyrad	1.47 ± 0.006	2.34 ± 0.01	3.61 ± 0.005	6.15 ± 0.01
ptyshv	1.22 ± 0.03	1.81 ± 0.04	2.61 ± 0.09	4.26 ± 0.2
py4dstem	0.813 ± 0.05	1.46 ± 0.06	2.43 ± 0.05	4.43 ± 0.1

	1 probe	3 probes	6 probes	12 probes
ptyrad	3.33 ± 0.01	4.87 ± 0.01	7.13 ± 0.009	11.7 ± 0.01
ptyshv	2.01 ± 0.06	3.86 ± 0.09	6.52 ± 0.2	11.8 ± 0.2
py4dstem	1.58 ± 0.06	3.58 ± 0.1	6.49 ± 0.1	12.5 ± 0.3

	1 probe	3 probes	6 probes	12 probes
ptyrad	6.97 ± 0.01	9.52 ± 0.01	13.3 ± 0.01	20.9 ± 0.01
ptyshv	3.15 ± 0.09	6.86 ± 0.08	12.3 ± 0.2	23.5 ± 0.4
py4dstem	2.66 ± 0.08	6.58 ± 0.1	12.5 ± 0.3	24.4 ± 0.6

batch 1024

	1 probe	3 probes	6 probes	12 probes
ptyrad	1.47 ± 0.1	2.28 ± 0.006	3.55 ± 0.008	6.11 ± 0.007
ptyshv	1.24 ± 0.1	1.66 ± 0.05	2.39 ± 0.1	3.8 ± 0.2
py4dstem	0.773 ± 0.07	1.35 ± 0.03	2.28 ± 0.03	4.07 ± 0.07

	1 probe	3 probes	6 probes	12 probes
ptyrad	3.25 ± 0.008	4.8 ± 0.007	7.08 ± 0.009	11.7 ± 0.004
ptyshv	1.87 ± 0.04	3.55 ± 0.08	6 ± 0.1	11 ± 0.3
py4dstem	1.44 ± 0.03	3.27 ± 0.04	6.01 ± 0.07	OOM

	1 probe	3 probes	6 probes	12 probes
ptyrad	6.85 ± 0.008	9.45 ± 0.007	13.2 ± 0.008	OOM
ptyshv	2.88 ± 0.07	6.3 ± 0.09	11.5 ± 0.2	OOM
py4dstem	2.44 ± 0.02	6.11 ± 0.07	11.6 ± 0.1	OOM