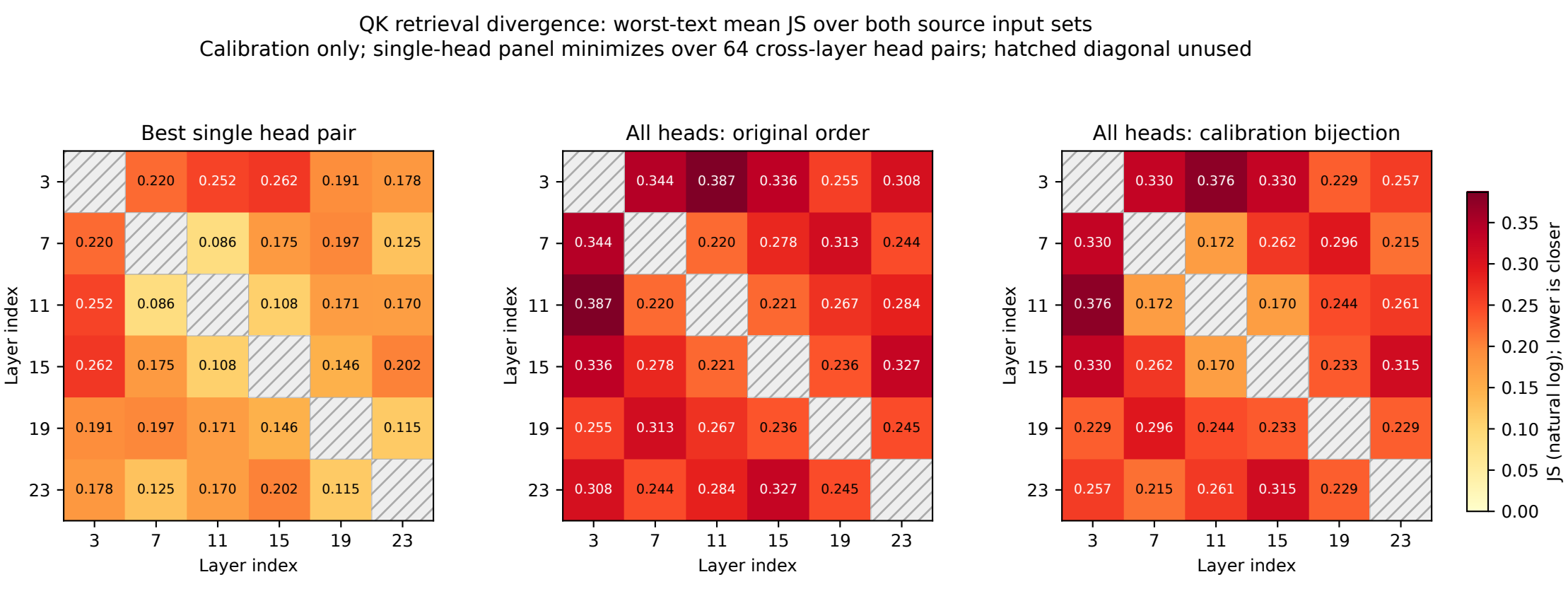

QK retrieval divergence: worst-text mean JS over both source input sets
Calibration only; single-head panel minimizes over 64 cross-layer head pairs; hatched diagonal unused
Best single head pair
All heads: original order
All heads: calibration bijection
Layer index
Layer index
3 7 11 15 19 23
0.220 0.252 0.262 0.191 0.178
0.220 0.086 0.175 0.197 0.125
0.252 0.086 0.108 0.171 0.170
0.262 0.175 0.108 0.146 0.202
0.191 0.197 0.171 0.146 0.115
0.178 0.125 0.170 0.202 0.115
0.344 0.387 0.336 0.255 0.308
0.344 0.220 0.278 0.313 0.244
0.387 0.220 0.221 0.267 0.284
0.336 0.278 0.221 0.236 0.327
0.255 0.313 0.267 0.236 0.245
0.308 0.244 0.284 0.327 0.245
0.330 0.376 0.330 0.229 0.257
0.330 0.172 0.262 0.296 0.215
0.376 0.172 0.170 0.244 0.261
0.330 0.262 0.170 0.233 0.315
0.229 0.296 0.244 0.233 0.229
0.257 0.215 0.261 0.315 0.229
0.35 0.30 0.25 0.20 0.15 0.10 0.05 0.00
JS (natural log); lower is closer