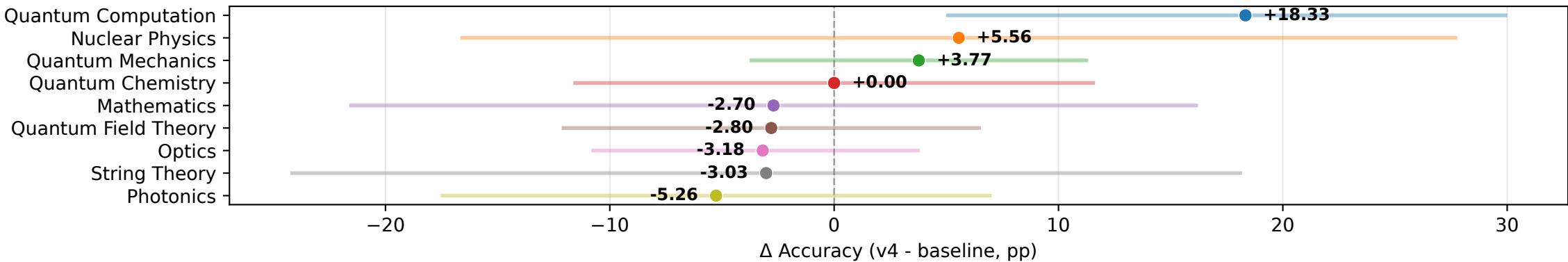
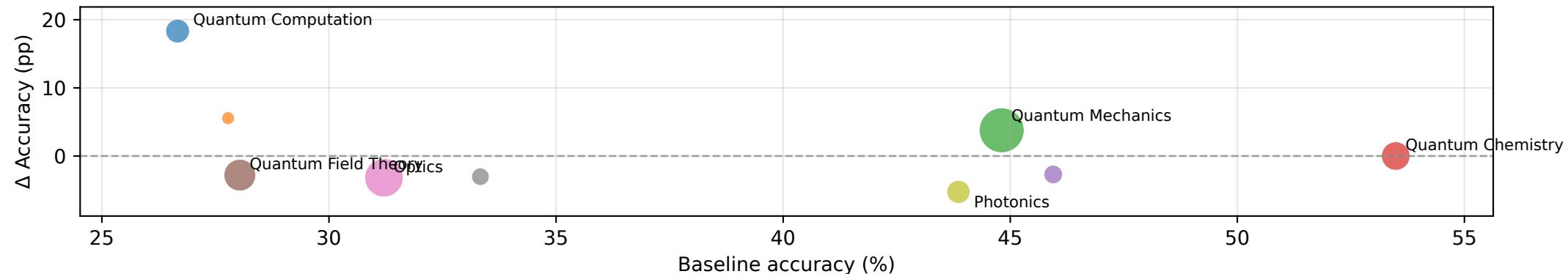


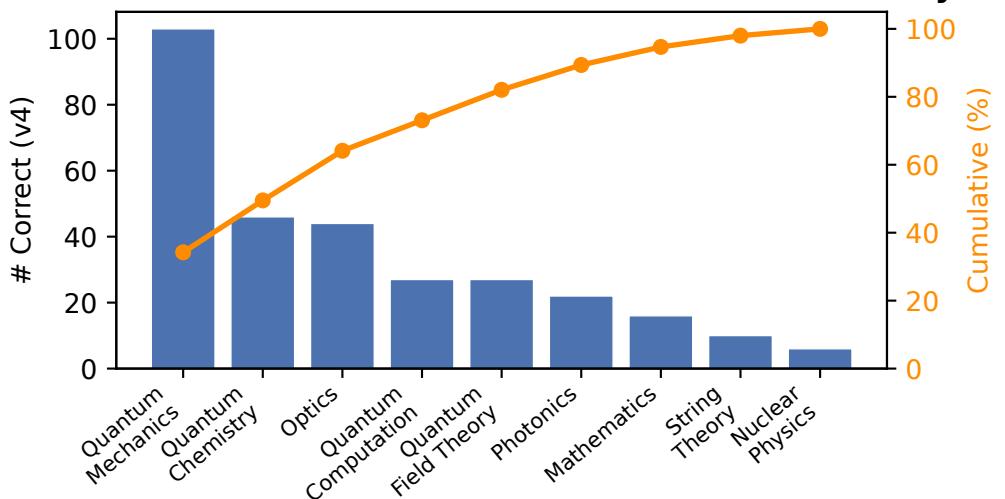
**(a) Subdomain gains/losses with bootstrap 95% CI (paired)**



**(b) Gain vs baseline strength (bubble size  $\propto$  #questions)**



**(c) Pareto: subdomain contribution to v4 accuracy**



**(d)  $\Delta pp$  by question type  $\times$  subdomain ( $n \geq 2$ )**

