



**Figure R1:** Ablation of stochasticity strength and BERT planner in RNA generation. Sweeping  $\eta$  from 0 to 2 with a step size of 0.25 over 10 runs per setting, showing the standard error. The use of a Planner significantly boosts performance, with  $\eta = 1.25$  yielding robust improvements for the P2+BERT Planner.

**Table R1:** Ablation of sampling strategies on code generation (HumanEval pass@1) and story infilling (ROUGE scores). We sweep  $\eta$  from 0 to 2 with step size of 0.1 and report the optimal result of  $\eta = 0.7$  on humaneval and  $\eta = 0.1$  on story infilling.

Method	pass@1 $\uparrow$	ROUGE-1 $\uparrow$	ROUGE-2 $\uparrow$	ROUGE-L $\uparrow$
RDM sampling ( $\eta = 1$ )	0.132	20.31	2.83	18.16
P2 ( $\eta = 0.7/0.1$ )	<b>0.180</b>	<b>25.27</b>	<b>7.36</b>	<b>23.25</b>
Greedy ancestral ( $\eta = 0.0$ )	0.161	24.68	7.12	22.85
Vanilla ancestral ( $\eta = 0.0$ )	0.121	17.18	2.72	15.57
DFM ( $\eta = 0.7/0.1$ )	0.116	16.62	2.42	15.23

**Table R2:** Overhead comparison of different planner sizes with a 150M denoiser. Benchmarked on NVIDIA A100 80GB GPUs.

Planner	Elapsed Time (s) $\downarrow$	Token/sec $\uparrow$
No Planner	<b>33.71</b>	<b>673.16</b>
ESM2-8M	39.25	509.55
ESM2-35M	45.30	441.46
ESM2-150M	50.21	398.36
ESM2-650M	119.45	167.44
ESM2-3B	508.10	39.36