Models		Iter.	Speed	WMT'14		WMT'16	
				EN-DE	DE-EN	EN-RO	RO-EN
AT	Transformer *base* (teacher)	N	1.0×	**27.48**	**31.39**	**33.70**	**34.05**
	Transformer *base* (12-1)	N	2.4×	26.21	30.80	33.17	33.21
	+ KD	N	2.5×	27.34	30.95	33.52	34.01
Iterative NAT	iNAT (Lee et al., 2018)	10	1.5×	21.61	25.48	29.32	30.19
	Blockwise (Stern et al., 2018)	≈ N/5	3.0×	27.40	-	-	-
	InsT (Stern et al., 2019)	≈ log N	4.8×	27.41	-	-	
	CMLM (Ghazvininejad et al., 2019)*	10	1.7×	27.03	30.53	33.08	33.31
	LevT (Gu et al., 2019)	Adv.	4.0×	27.27	-	-	33.26
	KERMIT (Chan et al., 2019)	≈ log N	-	27.80	30.70	-	-
	LaNMT (Shu et al., 2020)	4	5.7×	26.30	-	-	29.10
	SMART (Ghazvininejad et al., 2020b)*	10	1.7×	27.65	31.27	-	-
	DisCO (Kasai et al., 2020a)*	Adv.	3.5×	27.34	31.31	33.22	33.25
	Imputer (Saharia et al., 2020)*	8	3.9×	**28.20**	**31.80**	**34.40**	**34.10**
Fully NAT	Vanilla-NAT (Gu et al., 2018a)	1	15.6×	17.69	21.47	27.29	29.06
	LT (Kaiser et al., 2018)	1	3.4×	19.80	-	-	-
	CTC (Libovický and Helcl, 2018)	1	-	16.56	18.64	19.54	24.67
	NAT-REG (Wang et al., 2019)	1	-	20.65	24.77	-	-
	Bag-of-ngrams (Shao et al., 2020)	1	10.0×	20.90	24.60	28.30	29.30
	Hint-NAT (Li et al., 2018)	1	-	21.11	25.24	-	-
	DCRF (Sun et al., 2019)	1	10.4×	23.44	27.22	-	-
	Flowseq (Ma et al., 2019)	1	1.1×	23.72	28.39	29.73	30.72
	ReorderNAT (Ran et al., 2019)	1	16.1×	22.79	27.28	29.30	29.50
	AXE (Ghazvininejad et al., 2020a)*	1	15.3×	23.53	27.90	30.75	31.54
	EM+ODD (Sun and Yang, 2020)	1	16.4×	24.54	27.93	-	-
	GLAT (Qian et al., 2020)	1	15.3×	25.21	29.84	31.19	32.04
	Imputer (Saharia et al., 2020)*	1	18.6×	25.80	28.40	32.30	31.70
	Ours (Fully NAT)	1	17.6×	11.40	16.47	24.52	24.79
	+ KD	1	17.6×	19.50	24.95	29.91	30.25
	+ KD + CTC	1	16.8×	26.51	30.46	33.41	34.07
	+ KD + CTC + VAE	1	16.5×	**27.49**	31.10	**33.79**	33.87
	+ KD + CTC + GLAT	1	16.8×	27.20	**31.39**	33.71	**34.16**