Table 2: Performance of BLEU score on WMT14 En-De/De-En and IWSLT14 De-En tasks. The number in the parentheses denotes the performance gap between NART models and their ART teachers. ”/” denotes that the results are not reported. LSTM-based results are from [2, 27]; CNN-based results are from [5, 28]; Transformer [1] results are based on our own reproduction.[6]

	WMT14		IWSLT14		
Models	En-De	De-En	De-En	Latency	Speedup
Autoregressive models					
LSTM-based [2]	24.60	/	28.53	/	/
CNN-based [5]	26.43	/	32.84	/	/
Transformer [1] (beam size = 4)	27.41	31.29	33.26	$387ms^{\ddagger}$	1.00×
Non-autoregressive models					
FT [6]	17.69 (5.76)	21.47 (5.55)	/	$39ms^{\dagger}$	$15.6\times^{\dagger}$
FT [6] (rescoring 10)	18.66 (4.79)	22.41 (4.61)	/	$79ms^{\dagger}$	$7.68\times^{\dagger}$
FT [6] (rescoring 100)	19.17 (4.28)	23.20 (3.82)	/	$257ms^{\dagger}$	$2.36\times^{\dagger}$
IR [9] (adaptive refinement)	21.54 (3.03)	25.43 (3.04)	/	/	$2.39\times^{\dagger}$
LT [15]	19.80 (7.50)	/	/	$105ms^{\dagger}$	/
LT [15] (rescoring 10)	21.00 (6.30)	/	/	/	/
LT [15] (rescoring 100)	22.50 (4.80)	/	/	/	/
CTC [13]	17.68 (5.77)	19.80 (7.22)	/	/	$3.42\times^{\dagger}$
ENAT-P [29]	20.26 (7.15)	23.23 (8.06)	25.09 (7.46)	$25ms^{\dagger}$	$24.3\times^{\dagger}$
ENAT-P [29] (rescoring 9)	23.22 (4.19)	26.67 (4.62)	28.60 (3.95)	$50ms^{\dagger}$	$12.1\times^{\dagger}$
ENAT-E [29]	20.65 (6.76)	23.02 (8.27)	24.13 (8.42)	$24ms^{\dagger}$	$25.3\times^{\dagger}$
ENAT-E [29] (rescoring 9)	24.28 (3.13)	26.10 (5.19)	27.30 (5.25)	$49ms^{\dagger}$	$12.4\times^{\dagger}$
NAT-REG [8]	20.65 (6.65)	24.77 (6.52)	23.89 (9.63)	$22ms^{\dagger}$	$27.6\times^{\dagger}$
NAT-REG [8] (rescoring 9)	24.61 (2.69)	28.90 (2.39)	28.04 (5.48)	$40ms^{\dagger}$	$15.1\times^{\dagger}$
VQ-VAE [16] (compress 8×)	26.70 (1.40)	/	/	$81ms^{\dagger}$	$4.08\times^{\dagger}$
VQ-VAE [16] (compress 16×)	25.40 (2.70)	/	/	$58ms^{\dagger}$	$5.71\times^{\dagger}$
Non-autoregressive models (Ours)					
NART	20.27 (7.14)	22.02 (9.27)	23.04 (10.22)	$26ms^{\ddagger}$	$14.9\times^{\ddagger}$
NART (rescoring 9)	24.22 (3.19)	26.21 (5.08)	26.79 (6.47)	$50ms^{\ddagger}$	$7.74\times^{\ddagger}$
NART (rescoring 19)	24.99 (2.42)	26.60 (4.69)	27.36 (5.90)	$74ms^{\ddagger}$	$5.22\times^{\ddagger}$
NART-CRF	23.32 (4.09)	25.75 (5.54)	26.39 (6.87)	$35ms^{\ddagger}$	$11.1\times^{\ddagger}$
NART-CRF (rescoring 9)	26.04 (1.37)	28.88 (2.41)	29.21 (4.05)	$60ms^{\ddagger}$	$6.45\times^{\ddagger}$
NART-CRF (rescoring 19)	26.68 (0.73)	29.26 (2.03)	29.55 (3.71)	$87ms^{\ddagger}$	$4.45\times^{\ddagger}$
NART-DCRF	**23.44 (3.97)**	**27.22 (4.07)**	**27.44 (5.82)**	$37ms^{\ddagger}$	$10.4\times^{\ddagger}$
NART-DCRF (rescoring 9)	**26.07 (1.34)**	**29.68 (1.61)**	**29.99 (3.27)**	$63ms^{\ddagger}$	$6.14\times^{\ddagger}$
NART-DCRF (rescoring 19)	**26.80 (0.61)**	**30.04 (1.25)**	**30.36 (2.90)**	$88ms^{\ddagger}$	$4.39\times^{\ddagger}$