Monotonic Infinite Lookback Attention for Simultaneous Machine Translation (Appendices)

Naveen Arivazhagan* **Colin Cherry*** **Wolfgang Macherey** **Chung-Cheng Chiu**

Semih Yavuz **Ruoming Pang** **Wei Li** **Colin Raffel**

Google

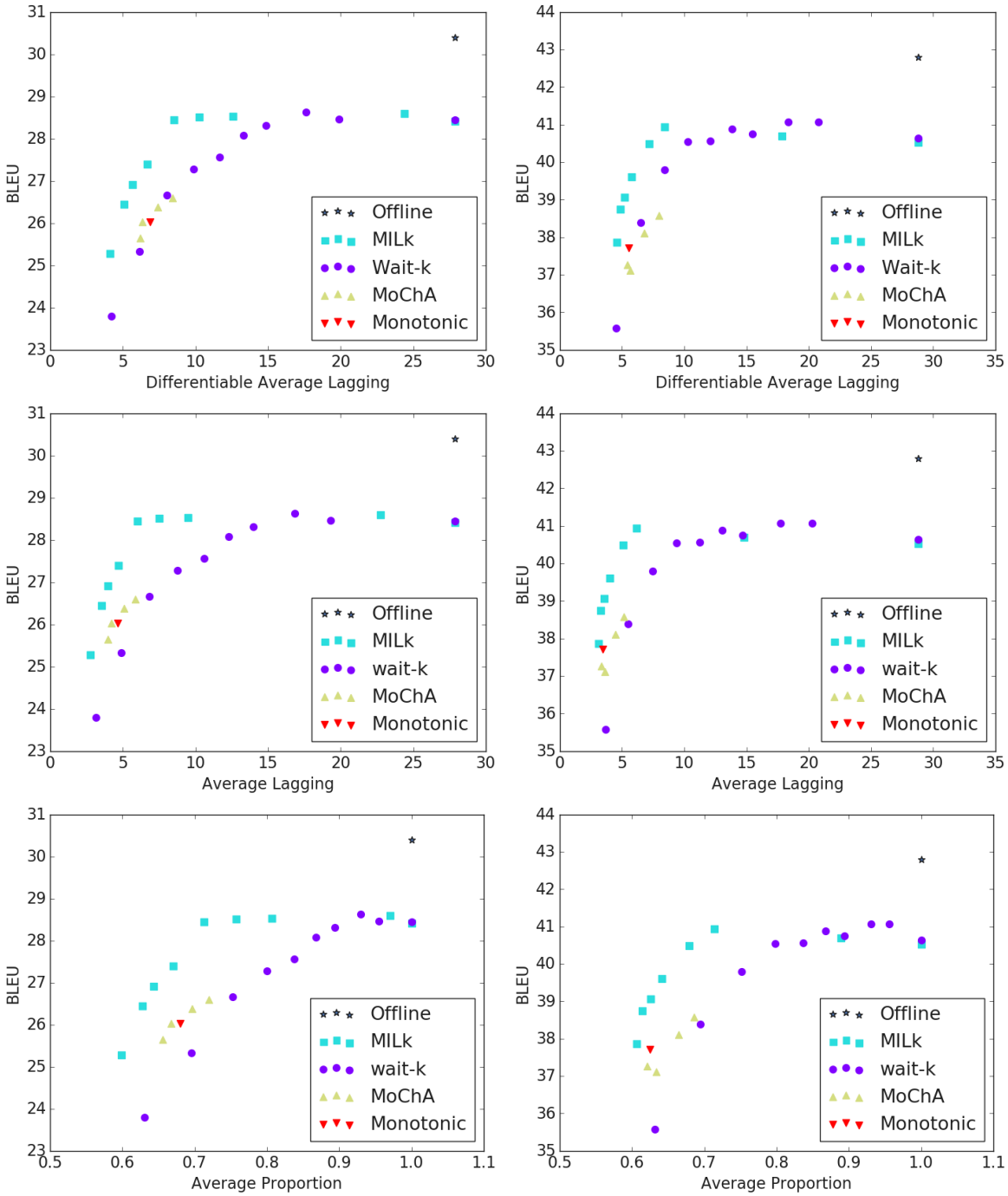
navari,colincherry,wmach,chungchengc@google.com

syavuz,rpang,mweili,craffel@google.com

A Expanded Results

We provide full-sized versions of our Quality-Latency curves from Section 5.3 in Figure 1. We also provide a complete table of results in Tables 1 and 2. As in the main text, DAL is Differentiable Average Lagging, AL is Average Lagging and AP is Average Proportion. wait-k is parameterized by k, MoChA by its chunk size cs and MILk by its latency weight λ. Results for EnFr MILk with $\lambda = 0.75$ are omitted, as it failed to converge.

*Equal contributions.



(a) German-to-English WMT15 (DeEn) test set.

(b) English-to-French WMT14 (EnFr) test set.

Figure 1: BLEU versus latency (top: differentiable average lagging, middle: average lagging, bottom: average proportion) for our two language pairs (left: DeEn, right: EnFr).

	Method	BLEU	DAL	AL	AP
wait-k	$k = 002$	19.5	2.5	1.5	0.56
	$k = 004$	23.8	4.2	3.1	0.63
	$k = 006$	25.3	6.1	4.9	0.70
	$k = 008$	26.7	8.1	6.8	0.75
	$k = 010$	27.3	9.9	8.8	0.80
	$k = 012$	27.6	11.7	10.6	0.84
	$k = 014$	28.1	13.3	12.3	0.87
	$k = 016$	28.3	14.9	14.0	0.89
	$k = 020$	28.6	17.6	16.9	0.93
	$k = 024$	28.5	19.9	19.3	0.95
	$k = 300$	28.4	27.9	27.9	1.00
MoChA	$cs = 01$	26.0	6.9	4.7	0.68
	$cs = 02$	25.6	6.2	4.0	0.66
	$cs = 04$	26.0	6.4	4.2	0.67
	$cs = 08$	26.4	7.4	5.1	0.70
	$cs = 16$	26.6	8.4	5.9	0.72
MILk	$\lambda = 0.75$	25.3	4.1	2.8	0.60
	$\lambda = 0.50$	26.4	5.1	3.5	0.63
	$\lambda = 0.40$	26.9	5.7	4.0	0.64
	$\lambda = 0.30$	27.4	6.7	4.7	0.67
	$\lambda = 0.20$	28.4	8.5	6.0	0.71
	$\lambda = 0.10$	28.5	10.3	7.5	0.76
	$\lambda = 0.05$	28.5	12.6	9.5	0.81
	$\lambda = 0.01$	28.6	24.4	22.7	0.97
	$\lambda = 0.00$	28.4	27.9	27.9	1.00

Table 1: Complete DeEn test set results, backing the curves in Figure 1a.

	Method	BLEU	DAL	AL	AP
wait-k	$k = 002$	28.9	2.9	2.1	0.57
	$k = 004$	35.6	4.5	3.7	0.63
	$k = 006$	38.4	6.5	5.5	0.70
	$k = 008$	39.8	8.4	7.5	0.75
	$k = 010$	40.5	10.3	9.4	0.80
	$k = 012$	40.6	12.1	11.3	0.84
	$k = 014$	40.9	13.9	13.0	0.87
	$k = 016$	40.7	15.5	14.7	0.89
	$k = 020$	41.1	18.3	17.7	0.93
	$k = 024$	41.1	20.8	20.3	0.96
	$k = 300$	40.6	28.8	28.8	1.00
MoChA	$cs = 01$	37.7	5.5	3.4	0.63
	$cs = 02$	37.3	5.4	3.3	0.62
	$cs = 04$	37.1	5.6	3.6	0.63
	$cs = 08$	38.1	6.8	4.5	0.66
	$cs = 16$	38.6	7.9	5.1	0.69
MILk	$\lambda = 0.50$	37.9	4.6	3.1	0.61
	$\lambda = 0.40$	38.7	4.9	3.3	0.61
	$\lambda = 0.30$	39.1	5.2	3.6	0.63
	$\lambda = 0.20$	39.6	5.8	4.0	0.64
	$\lambda = 0.10$	40.5	7.2	5.1	0.68
	$\lambda = 0.05$	40.9	8.4	6.2	0.71
	$\lambda = 0.01$	40.7	17.9	14.8	0.89
	$\lambda = 0.00$	40.5	28.8	28.8	1.00

Table 2: Complete EnFr test set results, backing the curves in Figure 1b.