Generalized Elias-Fano code for the compressed indexing of arbitrary integer sequences

Supplementary Material: Detailed Experimental Results

Michelangelo Pucci
ETH Zürich
Switzerland
mpucci@ethz.ch

Paolo Ferragina
Sant'Anna School of Advanced Studies
and University of Pisa
Italy
paolo.ferragina@santannapisa.it

We present the tables below, which contain detailed numerical data from our experiments. These tables are slightly wider than the standard margins, but we preferred to increase the font size to enhance their readability.

Table 1: Compression Ratio (%) Bold: Best, Underlined: Second, *Italics*: Third.

	General-purpose compressors					Special-purpose compressors											GEF variants						
Dataset	Brotli	Lz4	Snappy	Xz	Zstd	ALP	Camel	Chimp	Chimp128	DAC	ELF	Falcon	Gorilla	LeCo	NeaTS	TSXor	RLE-GEF	U-ĜEF	U-GEF*	B-ĜEF	B-GEF*	B*-ĜEF	B*-GEF*
IT	0.14	0.41	0.37	0.13	0.23	0.17	0.20	0.72	0.30	0.24	0.38	*0.11*	0.79	0.16	0.12	0.31	0.12	0.12	0.12	0.11	0.11	**0.10**	**0.10**
US	0.09	0.27	0.21	0.09	0.13	0.11	0.20	0.55	0.19	0.25	0.27	*0.07*	0.57	0.10	0.08	0.19	0.08	0.09	0.08	0.07	0.08	**0.06**	**0.06**
ECG	0.12	0.34	0.26	0.12	0.17	0.16	0.22	0.69	0.27	0.25	0.32	*0.12*	0.73	0.16	0.14	0.24	0.13	0.13	0.13	0.12	0.12	**0.11**	**0.11**
WD	0.28	0.53	0.54	0.24	0.34	0.25	0.40	0.84	0.44	0.26	0.54	0.23	0.91	0.25	0.25	0.47	**0.23**	0.23	0.23	0.23	<u>0.23</u>	*0.23*	*0.23*
AP	<u>0.13</u>	0.26	0.25	**0.12**	0.18	0.25	0.44	0.36	0.30	0.41	0.37	0.16	0.38	0.24	0.20	0.35	*0.14*	0.19	0.16	0.18	0.16	0.17	0.17
UK	0.09	0.27	0.21	0.09	0.13	0.12	0.15	0.47	0.23	0.26	0.36	*0.08*	0.54	0.11	0.09	0.16	0.09	0.11	0.10	0.09	0.09	**0.07**	**0.07**
GE	0.11	0.30	0.24	0.11	0.15	0.14	0.23	0.67	0.21	0.29	0.36	*0.11*	0.71	0.14	0.12	0.21	0.12	0.14	0.13	0.12	0.12	**0.10**	**0.10**
LON	0.19	0.50	0.49	*0.17*	0.33	0.26	0.37	0.62	0.55	0.47	0.64	**0.15**	0.63	0.25	0.18	0.66	0.19	0.19	0.18	0.19	<u>0.17</u>	0.19	0.19
LAT	0.24	0.52	0.51	*0.21*	0.41	0.25	0.26	0.61	0.44	0.47	0.63	**0.19**	0.64	0.30	0.22	0.51	0.24	0.24	0.23	0.24	<u>0.21</u>	0.23	0.23
DP	0.17	0.49	0.48	0.16	0.29	0.21	0.26	0.77	0.50	0.27	0.42	*0.15*	0.83	0.21	0.16	0.61	0.16	0.17	0.16	0.15	0.15	**0.13**	**0.13**
CT	0.16	0.43	0.38	0.16	0.25	0.15	0.23	0.74	0.36	0.19	0.66	0.14	0.87	0.15	0.14	0.31	*0.14*	0.14	0.14	0.14	0.14	**0.13**	**0.13**
DU	**0.08**	0.23	0.19	*0.08*	0.11	0.13	0.18	0.40	0.22	0.12	0.21	<u>0.08</u>	0.44	0.13	0.09	0.18	0.09	0.10	0.09	0.10	0.09	0.10	0.10
BT	<u>0.46</u>	0.67	0.69	**0.46**	0.58	*0.47*	0.53	0.84	0.48	0.57	0.82	0.52	0.93	0.55	0.65	0.56	0.53	0.53	0.53	0.53	0.53	0.51	0.51
BW	<u>0.41</u>	0.59	0.59	**0.36**	0.50	0.48	0.54	0.88	0.71	0.46	0.80	0.44	1.00	0.49	0.45	0.82	0.46	0.47	0.46	0.47	*0.44*	0.47	0.47
BM	<u>0.21</u>	0.44	0.39	**0.20**	0.29	0.33	0.41	0.65	0.41	0.37	0.65	0.22	0.73	0.31	0.24	0.48	0.24	0.24	0.24	0.25	*0.22*	0.24	0.24
BP	0.40	0.69	0.71	0.37	0.66	0.37	0.63	0.77	0.72	0.46	0.68	*0.36*	0.83	0.40	0.40	0.88	0.36	0.37	0.37	0.36	0.36	**0.34**	**0.34**

With $\hat{C}$ and C^*, we denote the GEF variant C that uses either its approximated or optimal split point, respectively.

Table 2: Compression Throughput (MB/s) Bold: Best, Underlined: Second, *Italics*: Third.

	General-purpose compressors					Special-purpose compressors											GEF variants						
Dataset	Brotli	Lz4	Snappy	Xz	Zstd	ALP	Camel	Chimp	Chimp128	DAC	ELF	Falcon	Gorilla	LeCo	NeaTS	TSXor	RLE-GEF	U-ĜEF	U-GEF*	B-ĜEF	B-GEF*	B*-ĜEF	B*-GEF*
IT	0.61	442.35	501.31	2.38	184.03	**1150.28**	152.99	352.88	94.01	270.83	185.40	132.70	*656.62*	<u>676.70</u>	0.93	25.79	595.60	312.58	240.28	305.28	230.45	305.55	288.16
US	0.47	592.79	605.66	1.91	263.73	**1166.41**	160.56	380.98	103.65	255.93	204.54	169.91	*659.78*	<u>690.99</u>	0.85	88.67	606.75	317.88	281.37	322.83	273.05	115.42	305.65
ECG	0.49	541.03	505.57	1.93	213.85	**1224.58**	166.22	351.66	98.23	263.07	192.74	126.51	509.76	<u>668.91</u>	0.97	62.13	*543.82*	277.47	252.39	263.93	230.98	305.53	269.00
WD	0.69	444.31	471.34	3.07	176.50	**1144.40**	114.37	348.78	100.61	242.56	164.79	100.97	<u>696.08</u>	*602.99*	1.21	16.28	519.86	294.93	191.65	294.48	186.73	313.38	283.93
AP	0.34	441.54	430.91	3.09	191.74	**1178.13**	126.03	421.52	100.00	175.94	185.42	116.38	<u>758.33</u>	*655.06*	0.86	50.73	593.83	322.83	202.86	333.74	201.83	348.38	300.21
UK	0.47	596.55	665.63	1.99	260.94	**1207.73**	182.03	372.88	198.32	242.99	184.54	161.15	<u>719.00</u>	*688.38*	0.79	89.32	627.52	318.40	270.75	321.79	264.61	147.30	284.17
GE	0.51	554.42	628.86	1.90	236.69	**1190.77**	145.32	358.71	153.74	259.14	184.56	148.98	<u>692.19</u>	*667.01*	0.81	71.50	612.28	296.10	241.46	298.83	232.76	303.48	281.54
LON	0.74	410.05	440.85	3.19	158.00	**1224.46**	130.99	511.80	271.65	212.62	183.34	124.01	*682.23*	664.94	0.84	19.73	<u>755.51</u>	399.76	191.12	401.54	192.42	426.22	374.51
LAT	0.72	406.75	446.68	3.22	155.01	**1200.09**	142.22	506.26	251.82	213.78	193.54	113.21	*676.35*	654.05	0.89	19.46	<u>748.54</u>	398.03	175.73	398.20	177.25	422.64	375.70
DP	0.67	397.49	395.39	2.62	172.53	**1174.17**	141.55	355.31	115.74	244.25	187.55	126.10	510.45	*558.47*	0.70	18.08	<u>595.14</u>	280.02	184.64	262.17	209.48	284.10	282.48
CT	0.61	385.29	464.48	2.18	173.27	**1157.93**	131.21	340.30	106.46	254.39	166.60	127.89	<u>680.35</u>	*622.12*	0.77	22.67	508.65	271.39	251.08	267.36	249.24	303.04	265.01
DU	0.41	667.27	*697.29*	2.21	297.39	**1187.17**	179.78	412.29	325.89	221.18	225.26	134.19	<u>777.18</u>	656.01	0.57	101.97	562.94	322.94	245.30	361.11	260.32	350.99	340.35
BT	0.58	392.03	358.25	2.84	144.51	**1098.87**	105.51	358.01	233.56	155.59	137.09	69.84	<u>665.07</u>	481.23	1.26	17.59	*567.58*	323.01	113.68	306.08	126.17	304.18	294.56
BW	0.55	364.43	382.55	2.89	149.80	367.32	100.44	337.59	232.14	157.03	123.73	72.65	**711.52**	*576.47*	1.11	18.06	<u>673.24</u>	310.63	128.14	348.15	114.65	325.03	273.04
BM	0.52	392.18	405.10	2.16	152.74	760.28	118.49	360.90	224.46	95.80	143.31	95.49	<u>667.98</u>	594.68	0.62	32.38	*656.96*	317.44	184.03	335.79	180.75	364.45	323.45
BP	0.54	336.05	311.96	2.23	117.38	450.78	99.88	343.65	237.94	146.31	154.83	82.51	**689.55**	*486.39*	1.17	16.32	<u>608.94</u>	275.50	160.01	265.43	152.98	290.93	265.02

With $\hat{C}$ and C^*, we denote the GEF variant C that uses either its approximated or optimal split point, respectively.

Table 3: Decompression Throughput (GB/s) Bold: Best, <u>Underlined</u>: Second, *Italics*: Third.

Dataset	General-purpose compressors					Special-purpose compressors											GEF variants			
	Brotli	Lz4	Snappy	Xz	Zstd	ALP	Camel	Chimp	Chimp128	DAC	ELF	Falcon	Gorilla	LeCo	NeaTS	TSXor	RLE-GEF	U-GEF	B-GEF	B*-GEF
IT	0.30	0.79	0.49	0.08	0.33	**3.94**	0.24	0.49	0.50	1.40	0.28	0.43	0.58	*2.81*	<u>3.20</u>	0.58	0.99	1.09	0.83	1.17
US	0.41	1.03	0.78	0.12	0.47	<u>4.09</u>	0.28	0.54	0.69	1.18	0.33	0.55	0.60	*2.93*	**4.43**	0.88	1.12	1.20	1.12	1.36
ECG	0.27	0.96	0.59	0.09	0.43	**5.53**	0.29	0.49	0.52	1.45	0.25	0.39	0.48	*2.71*	<u>3.92</u>	0.69	0.90	0.96	0.80	1.17
WD	0.22	1.06	0.42	0.05	0.34	**5.88**	0.16	0.47	0.47	1.14	0.25	0.30	0.61	*2.67*	<u>4.08</u>	0.56	0.73	0.80	0.72	1.13
AP	0.30	0.83	0.69	0.08	0.45	**5.71**	0.16	0.68	0.53	0.64	0.38	0.31	0.74	*2.59*	<u>3.78</u>	0.72	0.88	1.20	0.96	1.11
UK	0.41	1.04	0.81	0.12	0.48	**4.90**	0.33	0.53	0.56	1.03	0.34	0.53	0.64	*2.93*	<u>4.45</u>	0.86	1.13	1.19	1.01	1.31
GE	0.36	0.99	0.76	0.10	0.45	**6.17**	0.25	0.51	0.63	1.27	0.29	0.46	0.61	*2.85*	<u>4.57</u>	0.80	1.06	1.09	0.87	1.19
LON	0.29	0.92	0.44	0.06	0.31	**4.36**	0.17	0.80	0.50	1.40	0.31	0.33	0.60	*2.63*	<u>4.00</u>	0.55	1.38	1.22	0.96	1.59
LAT	0.23	0.96	0.44	0.05	0.30	**4.35**	0.20	0.77	0.43	1.36	0.31	0.29	0.59	*2.52*	<u>4.31</u>	0.52	1.36	1.12	0.95	1.54
DP	0.28	0.86	0.43	0.06	0.29	**6.40**	0.20	0.50	0.40	1.46	0.27	0.43	0.58	<u>2.64</u>	*2.25*	0.47	0.95	0.99	0.82	1.14
CT	0.27	0.73	0.50	0.07	0.31	**5.02**	0.24	0.45	0.47	1.30	0.28	0.45	0.58	<u>2.75</u>	*2.15*	0.58	0.74	0.89	0.74	1.13
DU	0.44	1.16	0.87	0.14	0.53	**7.94**	0.34	0.66	0.59	0.79	0.36	0.42	0.75	*2.84*	<u>3.54</u>	0.90	1.07	1.24	1.13	1.45
BT	0.16	0.88	0.42	0.03	0.31	**8.50**	0.09	0.50	0.45	1.17	0.25	0.18	0.60	<u>2.17</u>	*1.56*	0.50	0.82	0.84	0.81	1.07
BW	0.15	0.83	0.44	0.04	0.29	<u>3.43</u>	0.10	0.48	0.37	1.28	0.22	0.19	0.63	*2.20*	**4.75**	0.46	1.36	1.18	0.89	1.41
BM	0.22	0.84	0.49	0.06	0.36	**9.58**	0.16	0.50	0.51	0.84	0.27	0.25	0.61	*2.53*	<u>2.73</u>	0.58	1.59	1.06	0.84	1.50
BP	0.17	1.12	0.41	0.02	0.25	**6.15**	0.12	0.51	0.40	1.35	0.26	0.22	0.60	*2.38*	<u>4.40</u>	0.46	0.88	0.90	0.84	1.19

Table 4: Random Access Throughput (MB/s) Bold: Best, <u>Underlined</u>: Second, *Italics*: Third.

Dataset	General-purpose compressors					Special-purpose compressors											GEF variants			
	Brotli	Lz4	Snappy	Xz	Zstd	ALP	Camel	Chimp	Chimp128	DAC	ELF	Falcon	Gorilla	LeCo	NeaTS	TSXor	RLE-GEF	U-GEF	B-GEF	B*-GEF
IT	0.33	0.98	0.55	0.08	0.35	7.22	0.48	0.87	0.89	**140.37**	0.53	0.50	0.97	<u>121.42</u>	44.58	1.02	*64.78*	18.94	9.32	22.71
US	0.44	1.32	0.93	0.12	0.52	8.02	0.56	0.95	1.40	<u>76.99</u>	0.61	0.63	1.07	**134.66**	53.60	1.58	*68.01*	21.84	11.22	28.35
ECG	0.32	1.21	0.77	0.09	0.47	5.83	0.56	0.83	1.01	**139.14**	0.49	0.46	0.81	<u>121.78</u>	48.52	1.33	*60.28*	10.52	10.33	24.98
WD	0.23	1.37	0.48	0.05	0.37	7.36	0.31	0.82	0.90	<u>133.38</u>	0.48	0.33	1.00	**142.19**	11.47	1.02	*71.40*	26.14	14.64	30.24
AP	0.35	1.21	0.78	0.09	0.51	5.72	0.31	1.10	1.10	*66.26*	0.69	0.35	1.29	**135.02**	23.69	1.25	<u>75.79</u>	26.05	15.93	26.60
UK	0.45	1.35	0.99	0.12	0.53	6.03	0.67	0.95	1.13	129.98	0.63	0.60	1.13	**193.98**	69.66	1.57	*113.17*	31.08	15.92	36.17
GE	0.39	1.28	0.91	0.11	0.50	6.25	0.48	0.92	1.28	<u>179.22</u>	0.53	0.53	1.06	**202.92**	66.50	1.47	*119.35*	35.01	16.77	40.35
LON	0.31	1.16	0.49	0.06	0.35	10.50	0.33	1.42	1.01	<u>204.02</u>	0.58	0.36	1.07	**212.05**	76.67	0.99	*136.65*	35.69	19.68	46.74
LAT	0.25	1.20	0.49	0.05	0.33	10.39	0.39	1.37	0.87	<u>203.26</u>	0.58	0.31	1.06	**207.58**	72.32	0.94	*130.67*	34.98	19.55	46.35
DP	0.31	1.05	0.48	0.06	0.32	14.26	0.39	0.92	0.80	<u>319.40</u>	0.57	0.47	1.02	**419.93**	99.87	0.89	*212.01*	51.21	27.95	69.74
CT	0.28	0.89	0.57	0.07	0.35	17.28	0.47	0.83	1.06	<u>460.05</u>	0.55	0.50	1.07	**561.48**	116.44	1.12	*240.90*	68.91	34.23	82.39
DU	0.51	1.57	1.10	0.14	0.62	27.53	0.67	1.23	1.34	<u>393.38</u>	0.71	0.47	1.33	**953.09**	132.78	1.61	*320.34*	70.73	42.21	96.88
BT	0.18	1.13	0.47	0.03	0.34	21.83	0.19	0.94	0.94	<u>654.43</u>	0.46	0.19	1.07	**999.03**	122.00	0.97	*292.19*	67.32	38.91	94.63
BW	0.16	1.04	0.50	0.04	0.32	9.71	0.19	0.86	0.80	<u>764.17</u>	0.44	0.21	1.13	**984.79**	132.25	0.88	*297.95*	75.99	40.97	91.14
BM	0.26	1.11	0.59	0.06	0.43	26.32	0.32	1.02	1.12	<u>502.65</u>	0.53	0.28	1.17	**1083.34**	136.37	1.13	*344.71*	89.25	52.11	119.83
BP	0.21	1.62	0.48	0.03	0.32	28.46	0.25	1.06	0.94	<u>866.33</u>	0.55	0.26	1.24	**1159.38**	128.05	1.00	*317.75*	105.95	57.97	118.63