Parallel Processing Architectures and Power Efficiency in Smart Camera Chips

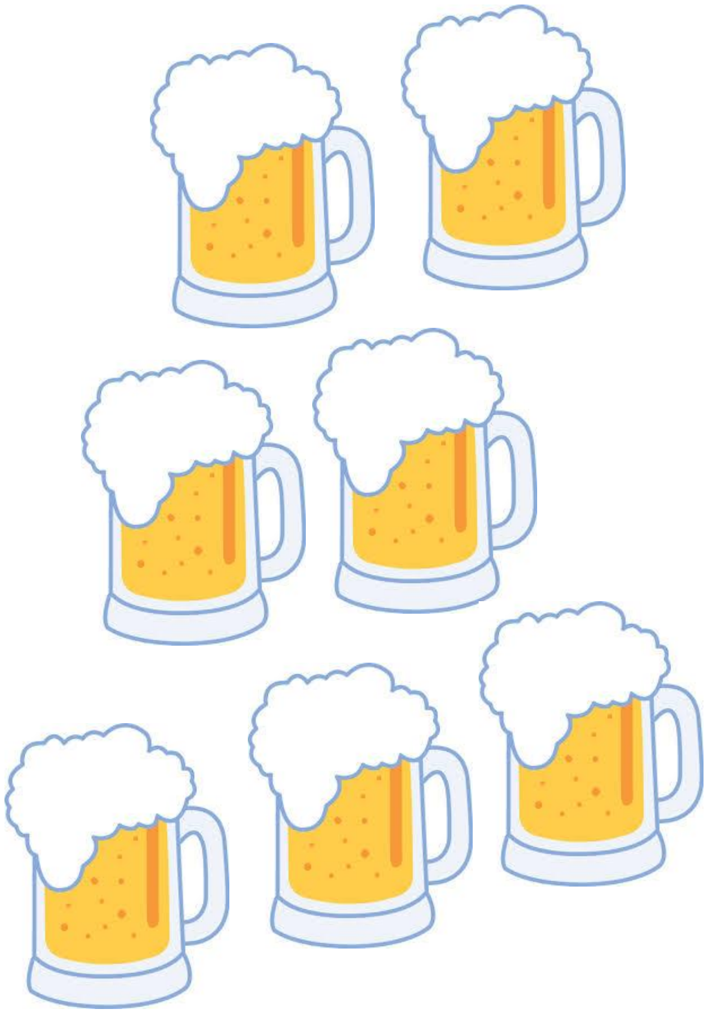
Ricardo Carmona-Galán, Jorge Fernández-Berni,
M. Trevisi and Ángel Rodríguez-Vázquez

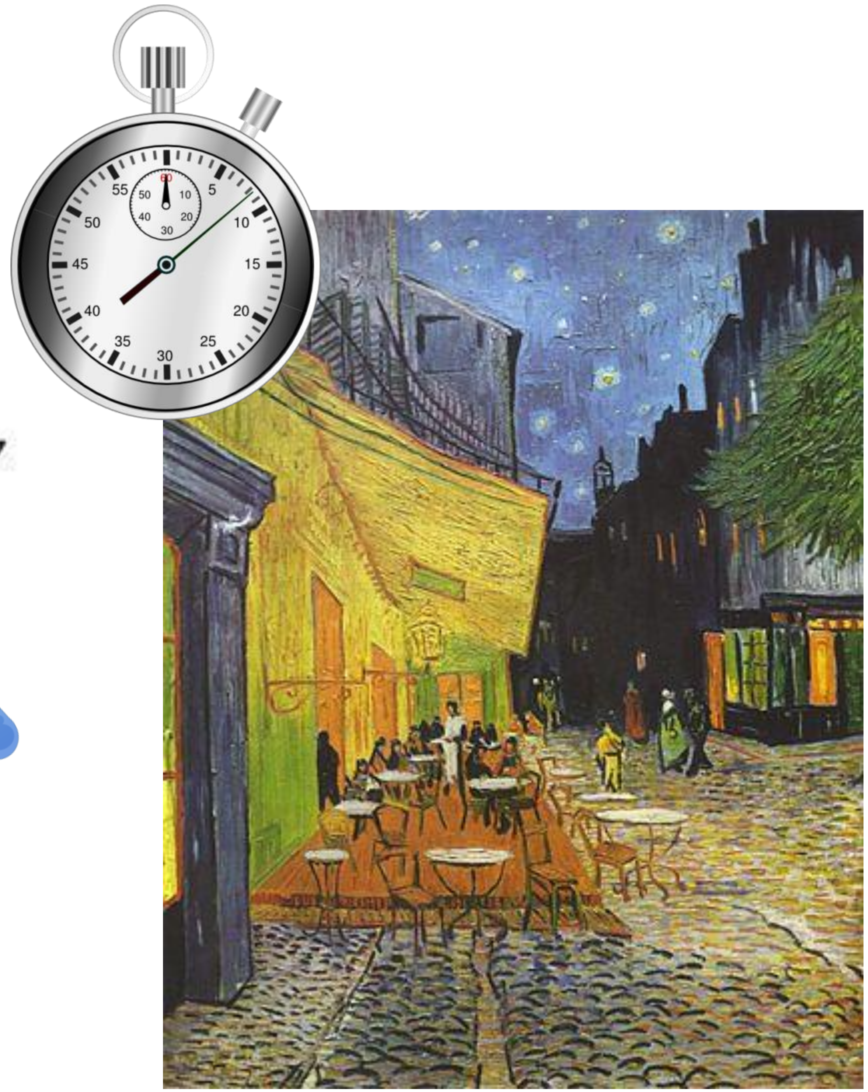
rcarmona@imse-cnm.csic.es
www.imse-cnm.csic.es/~rcarmona

Instituto de Microelectrónica de Sevilla (IMSE-CNM)
CSIC-Universidad de Sevilla (Spain)

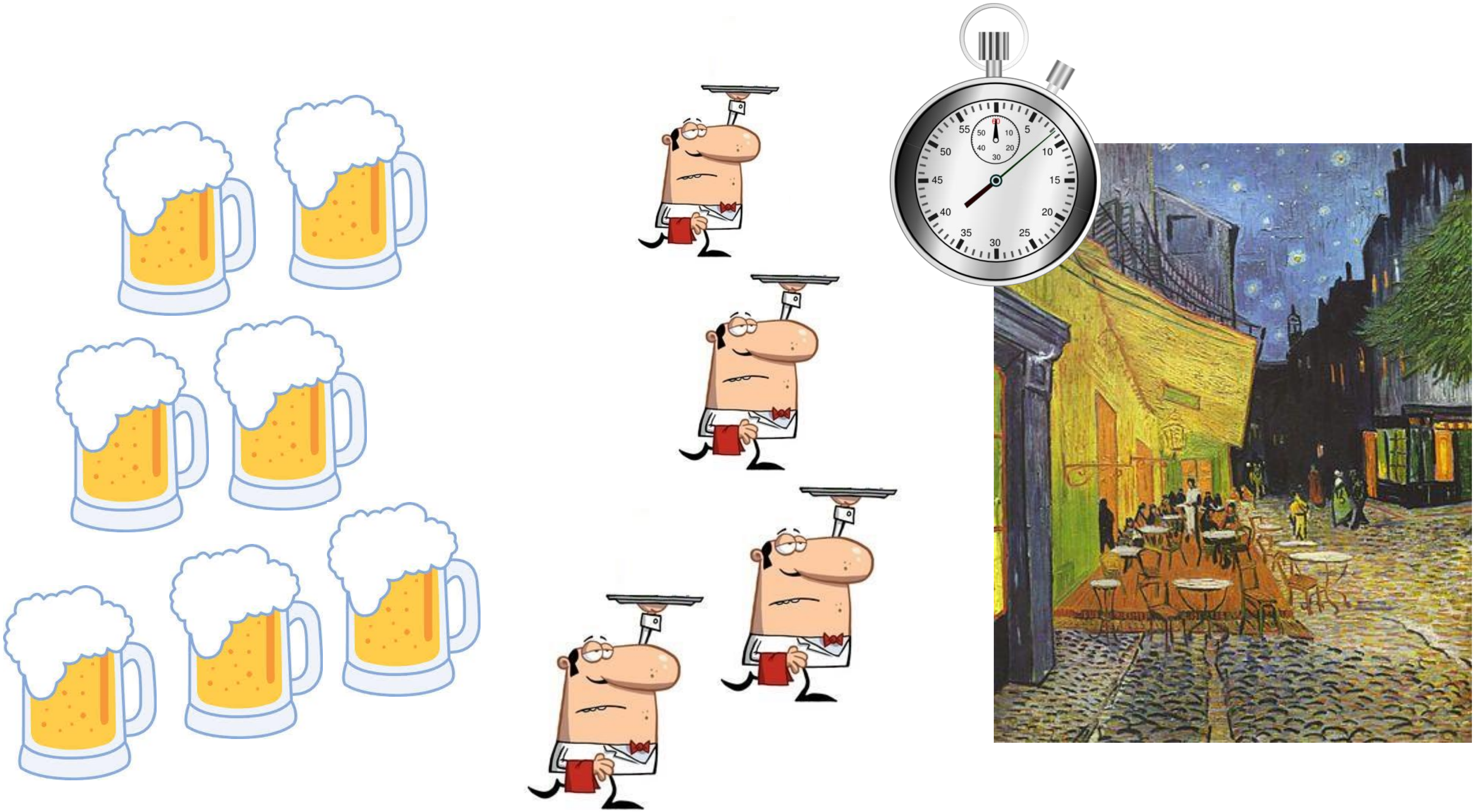
WASC 2014, Pisa (Italy)

Task parallelization





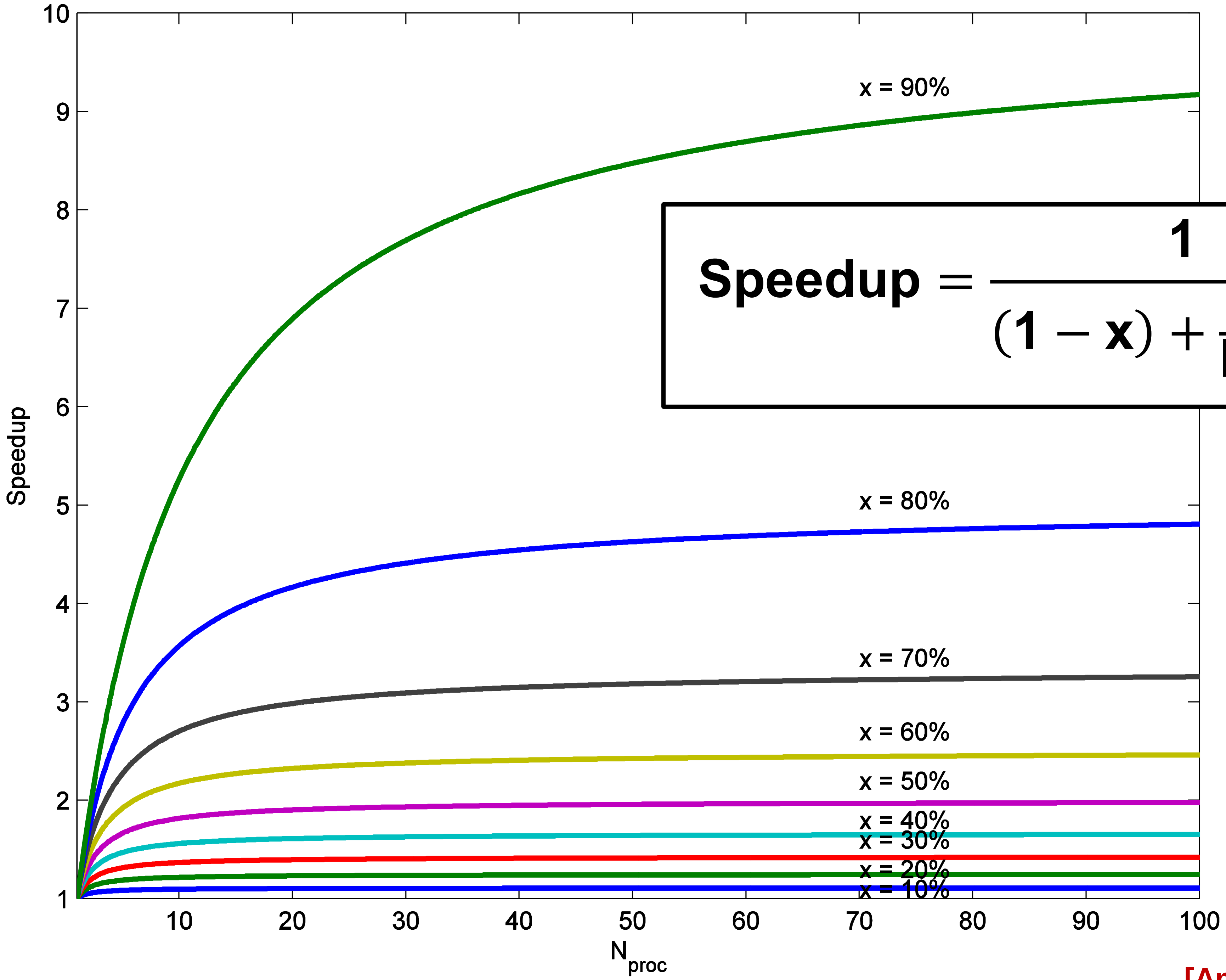
Task parallelization



Task parallelization

- Distributing tasks between several processors working in parallel speeds up processing
- Constrained by the degree of parallelization that can be achieved

Amdahl's law



$$\text{Speedup} = \frac{1}{(1 - x) + \frac{x}{N_{proc}}}$$

[Amdahl 1967]

Amdahl's law

- Favors the use of a single-core system
- But…problems have grown and parallel processing is the **only** alternative to operate onto a large amount of data in a certain amount of time

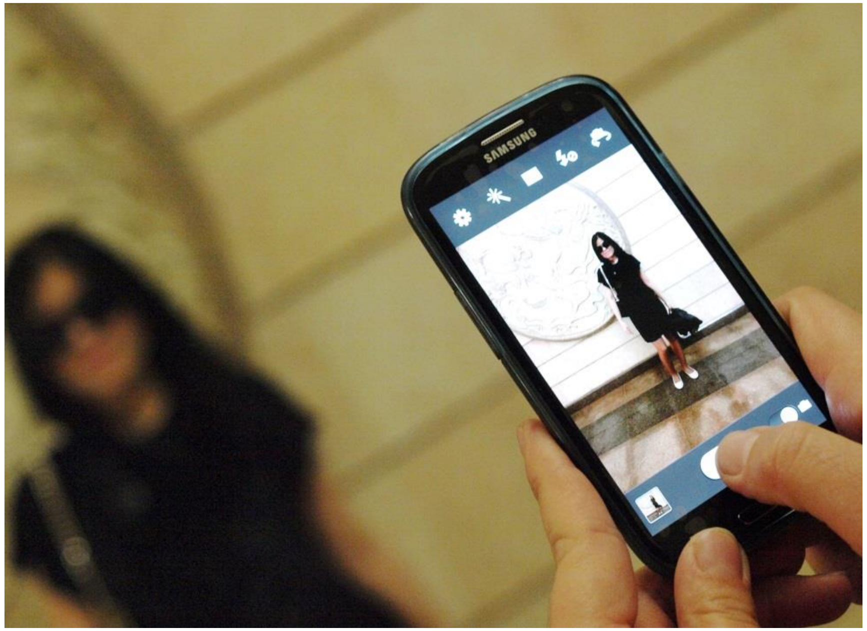
Performance vs. power efficiency

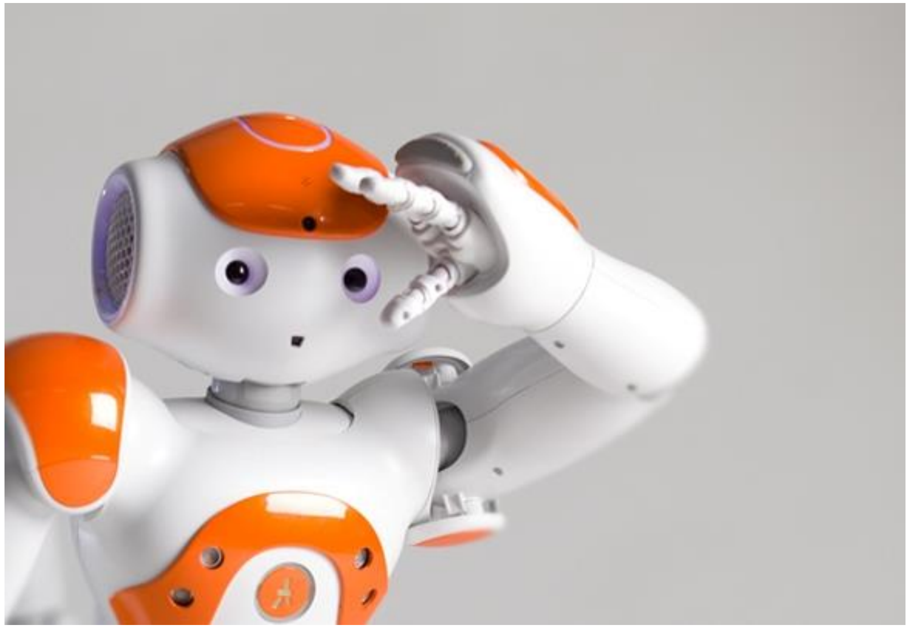
GOPS vs. GOPS/W

…or MOPS/mW, or nJ/OP

Performance vs. power efficiency

vs.



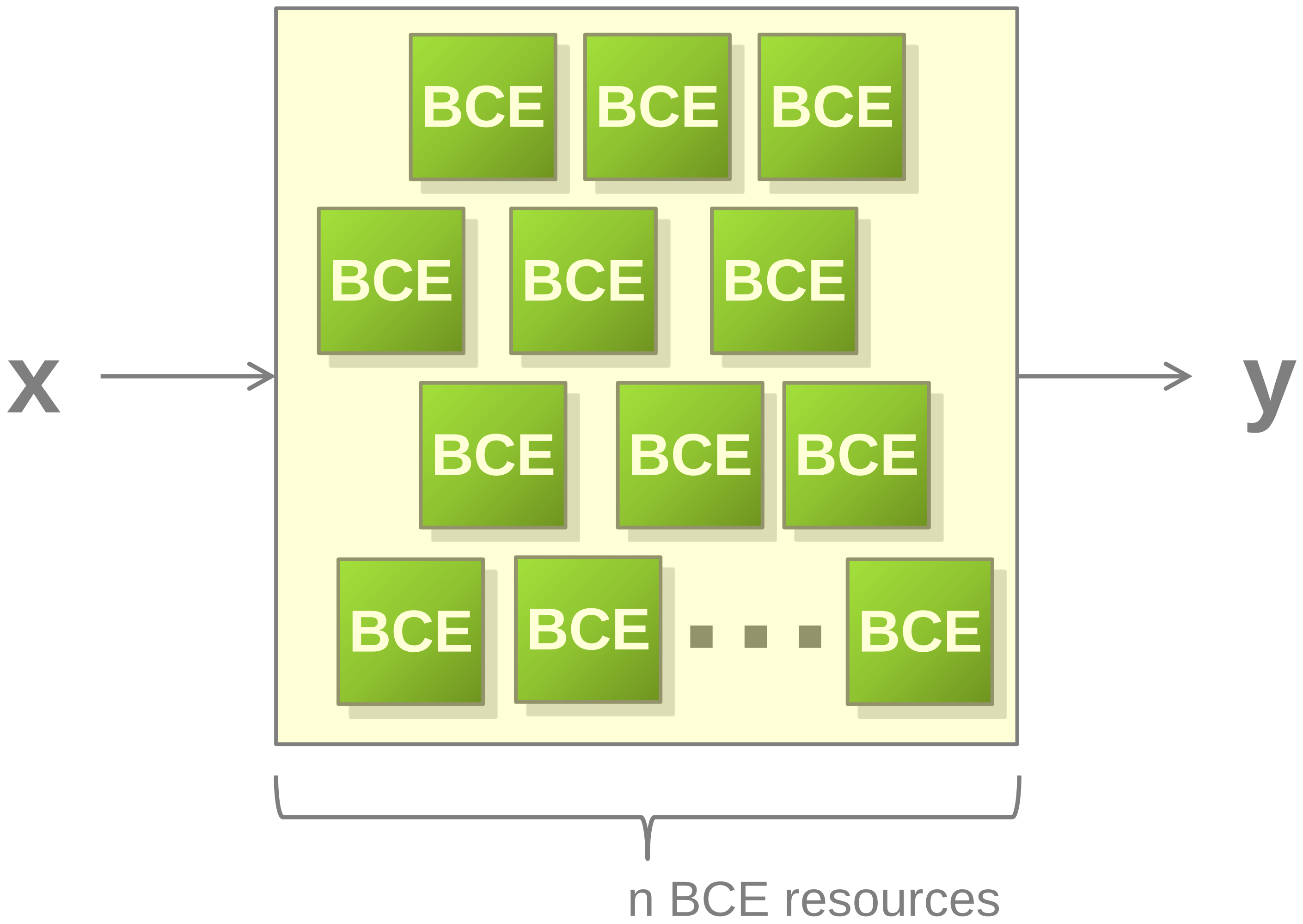


Basic core equivalent

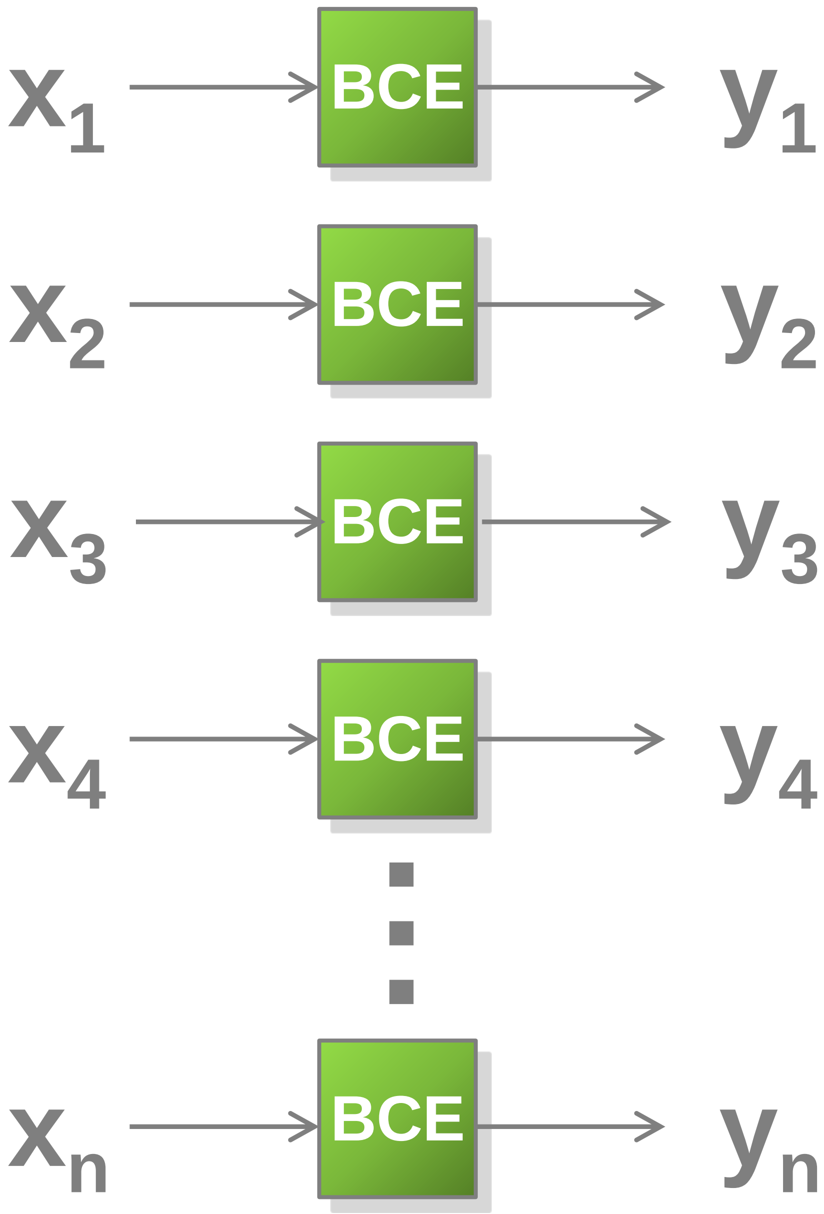
[Hill & Marty 2008]

- Time to perform an elementary operation → t_0
- Elementary performance → $G_0 = 1 / t_0$
- Energy required to realize an elementary op. → e_0
- Power consumption of one BCE → $P_0 = e_0 / t_0$

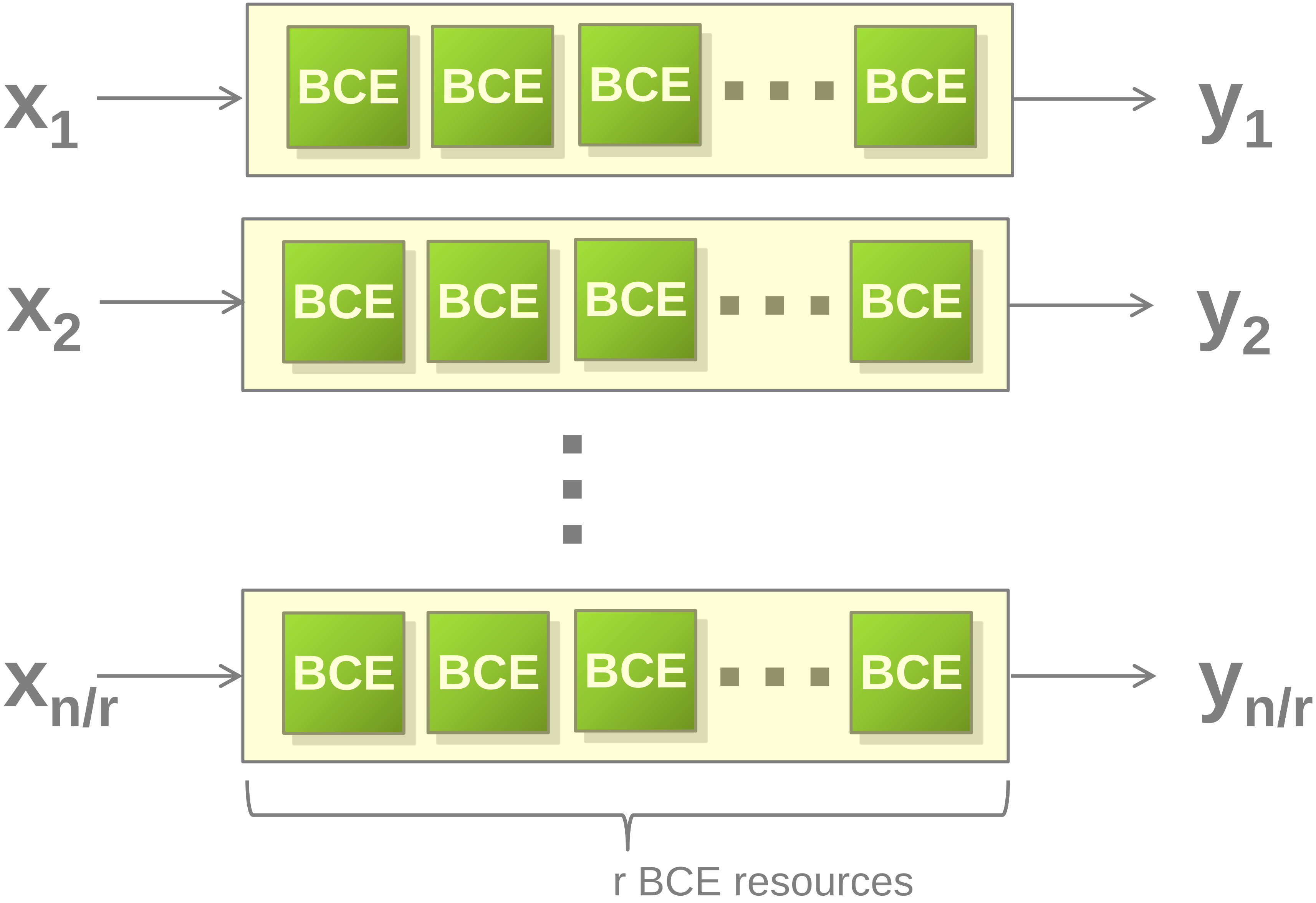
Single n-BCE core



n 1-BCE cores in parallel



n/r r-BCE cores in parallel



Pollack's rule

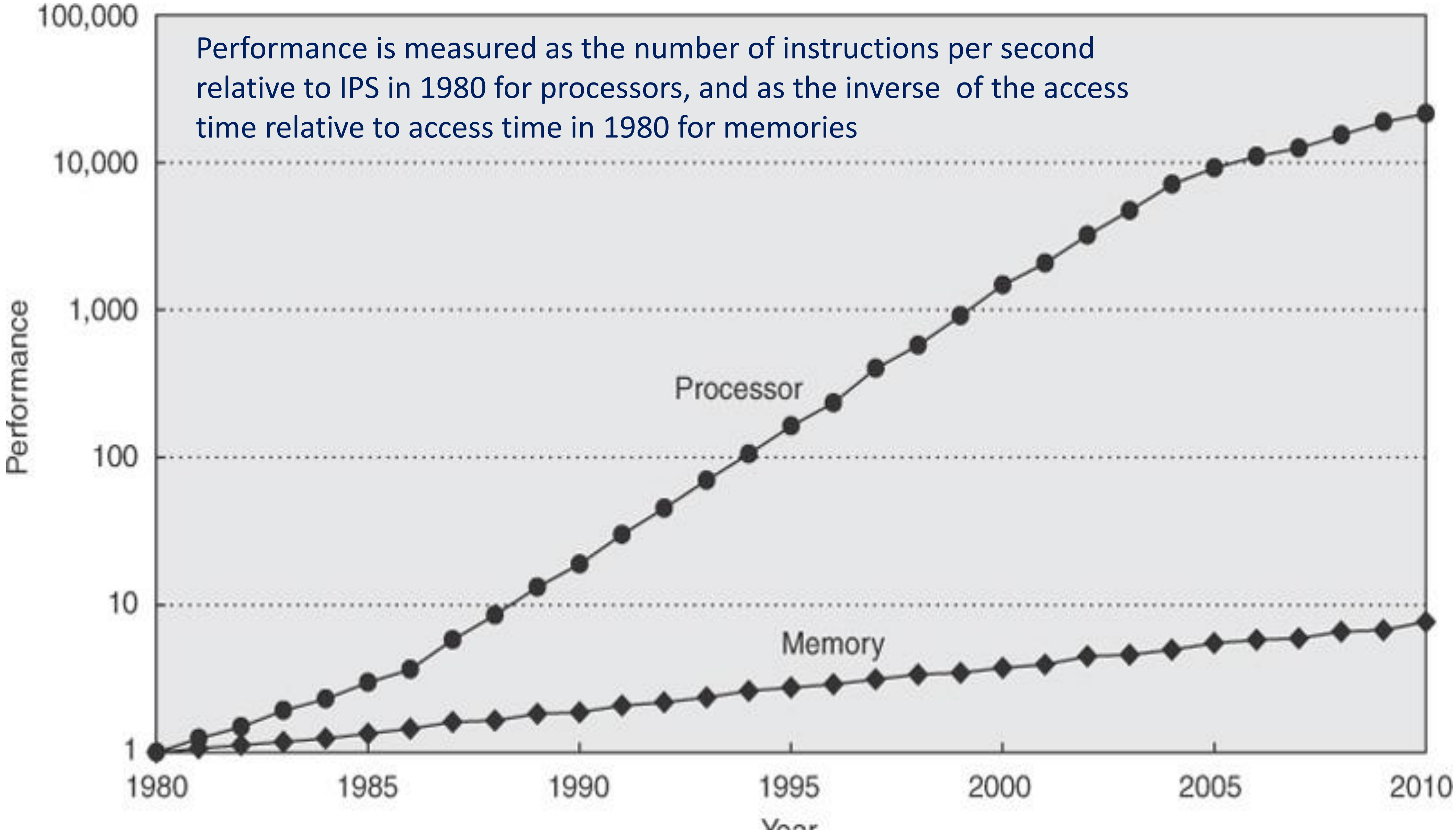
Performance scales with the square root of complexity

[Borkar 2007]

$$G(n,r)=\left(\frac{n}{r}\right)\sqrt{r}G_0=\frac{n}{\sqrt{r}}G_0$$

- Single n-BCE core: r = n → $G(n,n)=\sqrt{n}G_0$
- n 1-BCE cores in parallel: r = 1 → $G(n,1)=nG_0$

Processor/memory performance gap



[Hennessy & Patterson 2006]

Processing speed

$$t(n,r) = \frac{\sqrt{r}}{n} t_0$$

- Single n-BCE core: r = n → $t(n,n) = t_0 / \sqrt{n}$
- n 1-BCE cores in parallel: r = 1 → $t(n,1) = t_0 / n$

Energy required to operate

$$e(n,r) = n \cdot e_0$$

… which is independent of the degree of parallelization

Power consumption

$$P(n,r) = \frac{e(n,r)}{t(n,r)} = \frac{n^2}{\sqrt{r}} P_0$$

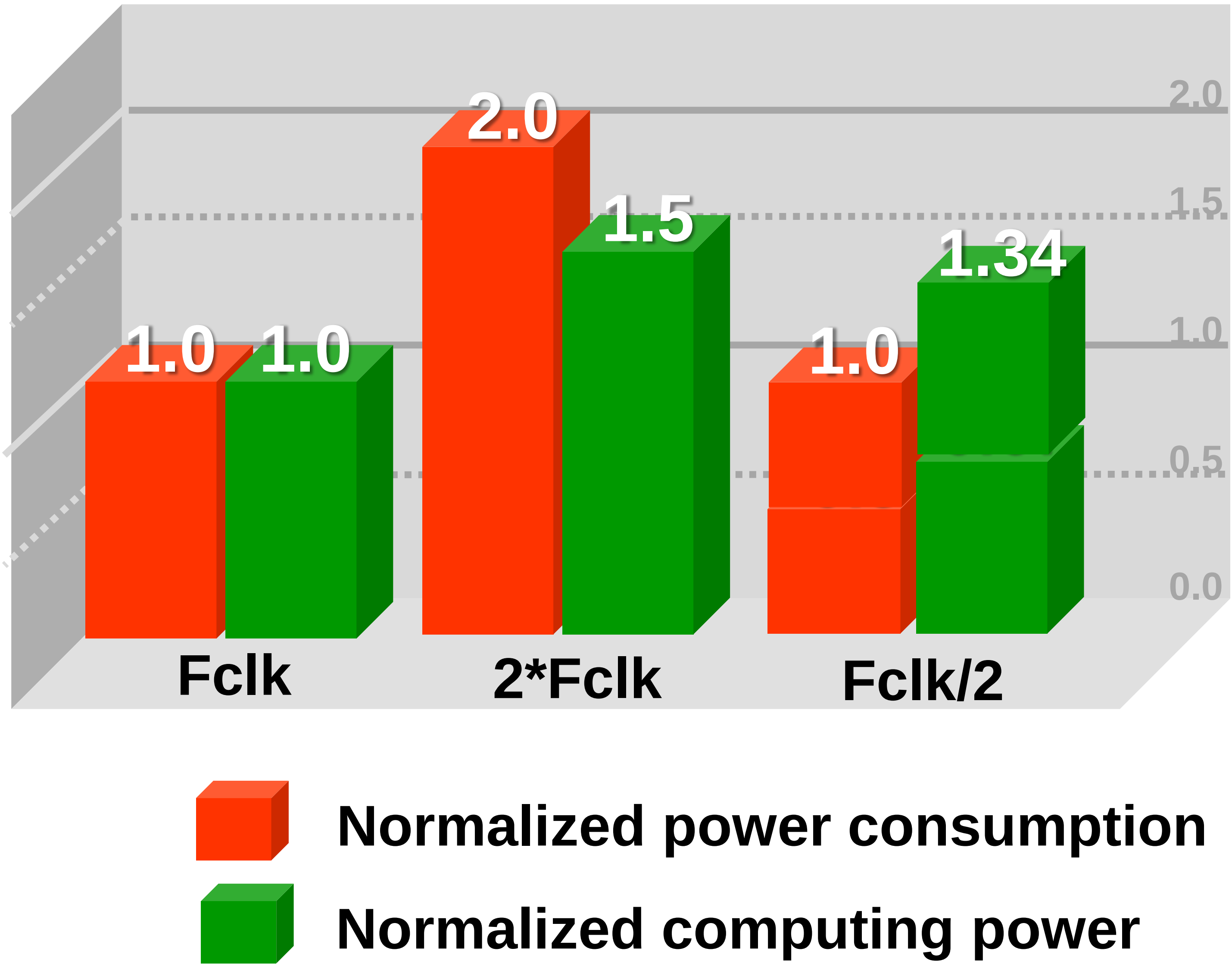
Power efficiency

$$\frac{G(n,r)}{P(n,r)} = \frac{G_0}{nP_0}$$

Power efficiency

$$\frac{G(n,r)}{P(n,r)} \propto \left(\frac{n}{r}\right)^k$$

Multicore architectures



A survey of multicore processors

First author	Year	Tech. (nm)	Nproc	Clk (MHz)	Area (mm2)	Power (mW)	GOPS
Gerosa	2008	45	1	1600.0	25.96	4000.0	3.85
Intel	2010	45	2	1600.0	51.92	8000.0	8.03
Hinrichs	2000	500	4	66.0	187.68	650.0	1.30
Shiota	2005	90	4	533.0	122.57	5000.0	51.20
Chien	2008	180	4	50.0	8.91	21.6	0.80
Minsu Kim	2009	130	4	200.0	4.30	51.8	54.00
Freescale	2011	40	4	1200.0	6.90	3800.0	12.00
Se-Hyun Yang	2012	32	4	1500.0	118.00	4000.0	14.00
Rohrer	2005	90	5	2500.0	62.00	50000.0	9.50
Kaul	2009	45	5	2800.0	0.75	278.0	17.17
Nvidia	2010	40	8	1000.0	49.00	500.0	4.60
Yuyama	2010	45	8	648.0	153.76	3070.0	114.51
Weihu Wu	2011	65	8	1050.0	299.80	40000.0	128.00
Weihu Wu	2013	35	8	1350.0	182.50	40000.0	172.80
Youngmin	2013	28	8	1800.0	123.71	6000.0	30.00
T.-H. Chen	2009	130	10	200.0	10.11	329.0	236.35
Ramacher	2001	350	16	100.0	506.00	8000.0	53.00
Chia-Hsia Yang	2009	90	16	16.0	8.88	275.0	50.00
Zhiyi Yu	2012	65	16	800.0	9.10	320.0	22.22
Donghyun Kim	2009	180	18	400.0	37.50	540.0	81.60
Yiping Dong	2011	90	20	1000.0	25.00	1131.7	3.10
Clermidy	2010	65	23	790.0	30.00	500.0	37.00
Peng Ou	2013	65	24	850.0	18.80	523.0	20.40
Xun He	2011	65	32	750.0	25.00	3830.0	375.00
Zhiyi Yu	2008	180	36	475.0	32.10	1152.0	21.62
Kwanho Kim	2008	130	64	200.0	36.00	392.0	96.00
Fick	2012	130	64	10.0	13.30	5.7	0.05
Hui Xu	2012	40	64	333.0	210.00	1700.0	852.00
Phi-Hung Pham	2013	130	64	174.0	23.00	200.0	11.20
Kwanho Kim	2009	130	65	200.0	36.00	583.0	125.00
Ozaki	2011	65	65	210.0	8.82	11.2	2.50
Khailany	2007	130	82	800.0	155.00	10496.0	256.00

First author	Year	Tech. (nm)	Nproc	Clk (MHz)	Area (mm2)	Power (mW)	GOPS
Kyo	2003	180	128	100.0	121.00	4000.0	51.20
Shorin Kyo	2008	130	128	100.0	100.00	2000.0	100.00
Chih-Chi Cheng	2009	180	128	50.0	70.50	374.0	76.80
Seungjin Lee	2010	130	128	200.0	4.22	92.0	76.80
Jae-Sung Yoon	2013	180	128	200.0	28.75	413.0	153.60
Joo-Young Kim	2010	180	130	400.0	49.00	695.0	201.40
Jimwook Oh	2013	130	157	200.0	32.00	534.0	342.00
Truong	2009	65	167	1070.0	0.71	47.5	1.08
Miao	2008	180	256	40.0	2.25	8.7	0.21
Arakawa	2008	65	260	250.0	152.83	783.0	90.00
Abbo	2008	90	320	84.0	74.00	600.0	107.00
Chuan-Yung Tsai	2012	65	360	250.0	20.25	351.0	360.00
Lopich	2011	350	418	75.0	9.00	26.4	1.00
Junyoung Park	2013	130	432	200.0	28.00	270.0	271.40
Dudek	2005	600	441	2.5	10.00	40.0	1.10
Graupner	2003	600	512	—	10.00	21.3	0.03
Tanabe	2012	40	549	266.0	44.54	748.6	463.90
Wen-Chia Yang	2011	350	1024	10.0	13.86	21.0	8.19
Jinwook Oh	2011	130	1025	200.0	13.50	75.0	49.14
Carmona	2003	500	2048	10.0	78.33	300.0	470.00
Noda	2007	90	2048	200.0	3.10	250.0	40.00
Kurafuji	2011	65	3328	560.0	24.00	545.0	191.00
Komuro	2003	500	4096	10.0	49.00	280.0	14.64
Qingyu Lin	2009	180	4096	40.0	5.25	82.5	2.10
Jendernalik	2013	350	4096	10.0	9.80	0.3	0.04
Zhang	2011	180	4128	100.0	13.50	450.0	44.01
Rossi	2010	90	4319	250.0	110.00	1450.0	120.00
Seungjin Lee	2011	130	4920	200.0	4.50	84.0	24.00
Seungjin Lee	2010	130	6412	400.0	50.00	704.0	228.00
Ikenaga	2000	250	16384	56.0	273.70	2300.0	640.00
Linan	2004	350	16384	100.0	145.18	4000.0	330.00
Komuro	2009	350	76800	50.0	78.55	41.6	3340.00
Dongsuk Jeon	2013	28	79400	27.0	2.22	2.7	149.30

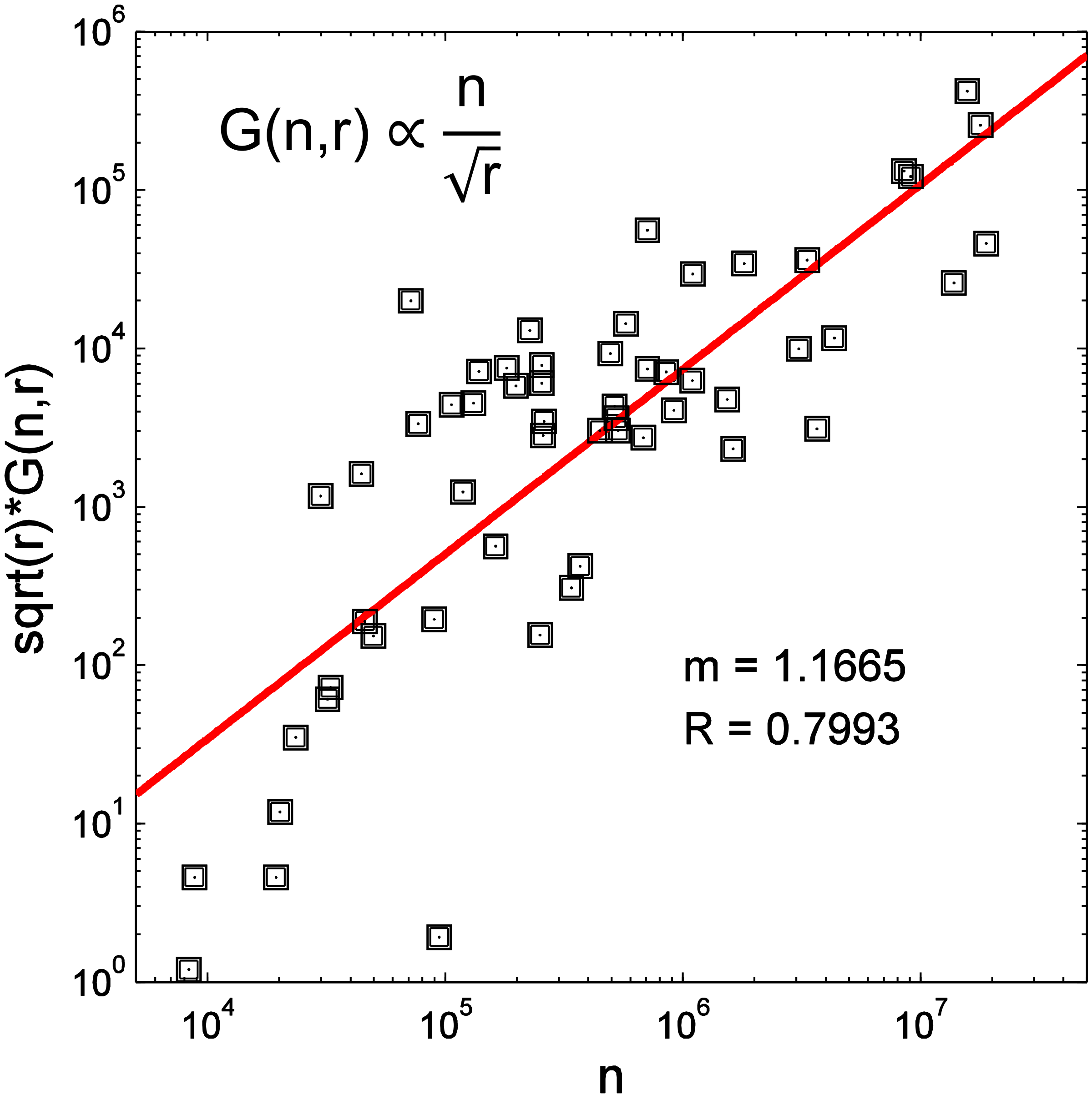
www.imse-cnm.csic.es/mondego/public/processor_comp.xlsx

Normalization: area of BCE

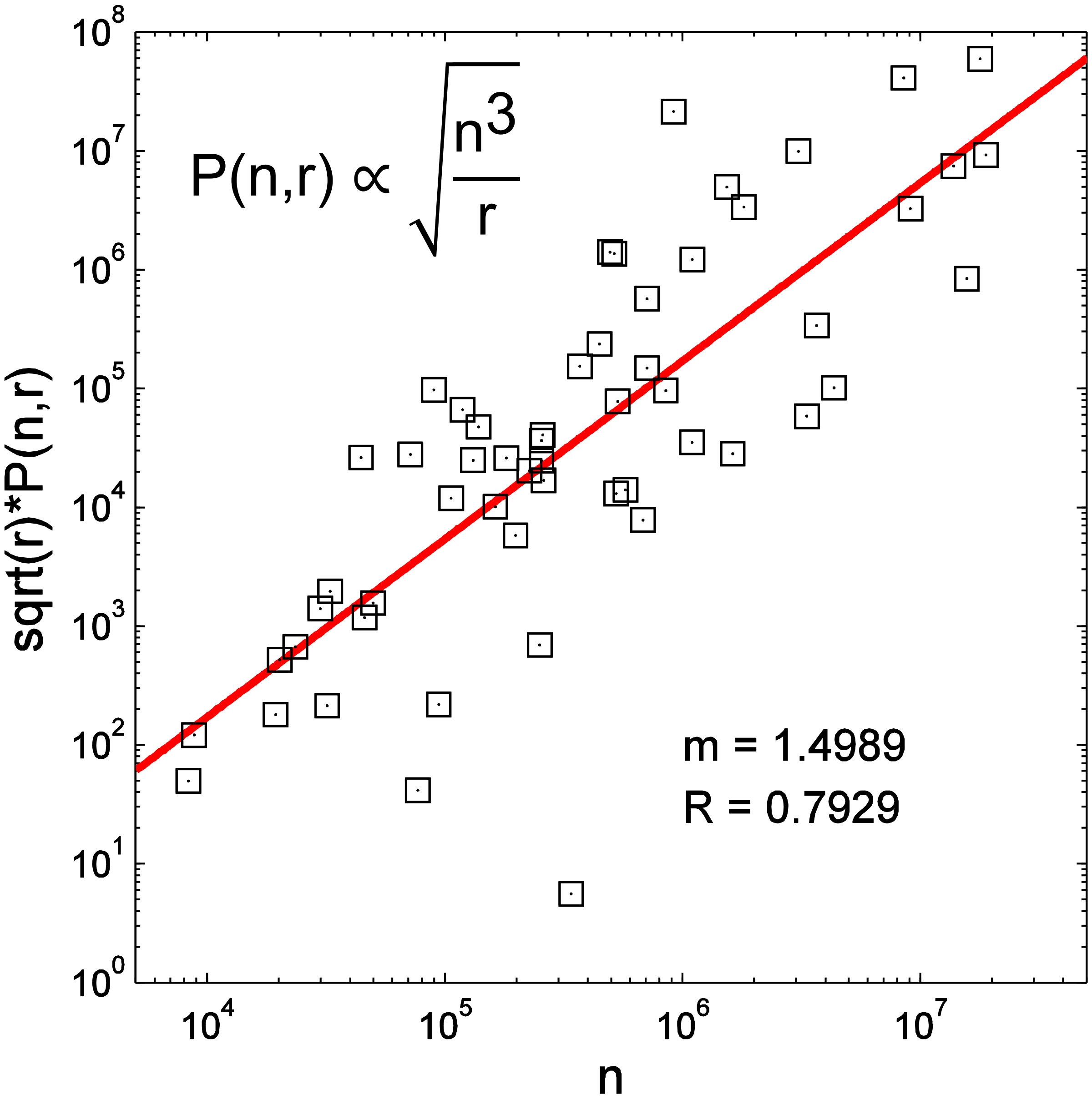
$$A_0 \equiv \min\left(\frac{A}{\lambda^2 N_{proc}}\right)$$

- Total number of resources $\rightarrow$ $n = \dfrac{A}{\lambda^2 A_0}$
- Total resources per core $\rightarrow$ $r = \dfrac{n}{N_{proc}}$

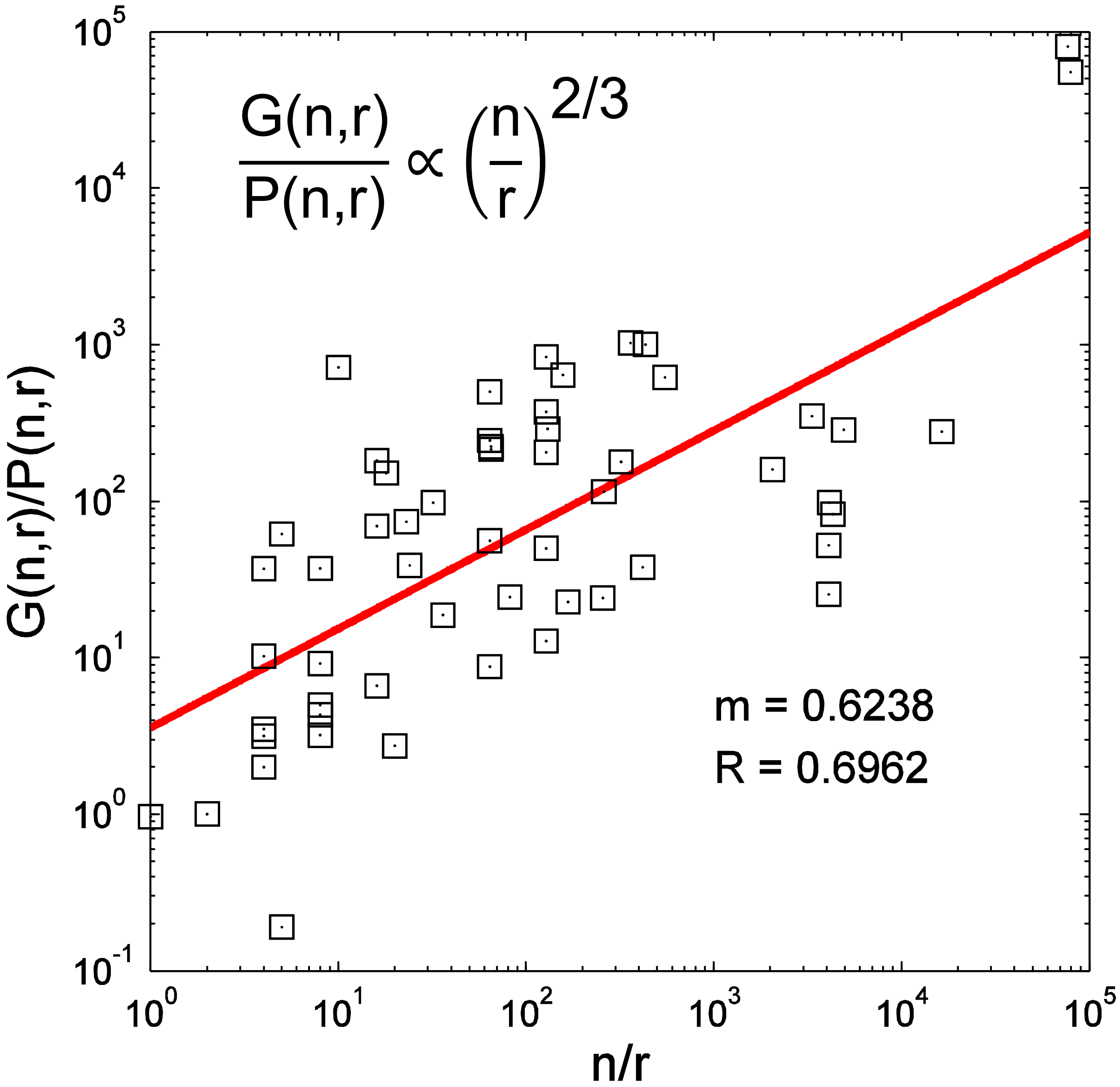
Pollack's rule



Power consumption vs. n



Power efficiency vs. n/r



Conclusions

- Parallelizing the operation of hardware resources has an incidence in power efficiency
- Increase in performance is easily predicted
- Estimation of power efficiency is more involved
- The roots of the gain are in the distribution of computing and memory resources
- The formal cause for the relation found is still pending

Acknowledgements

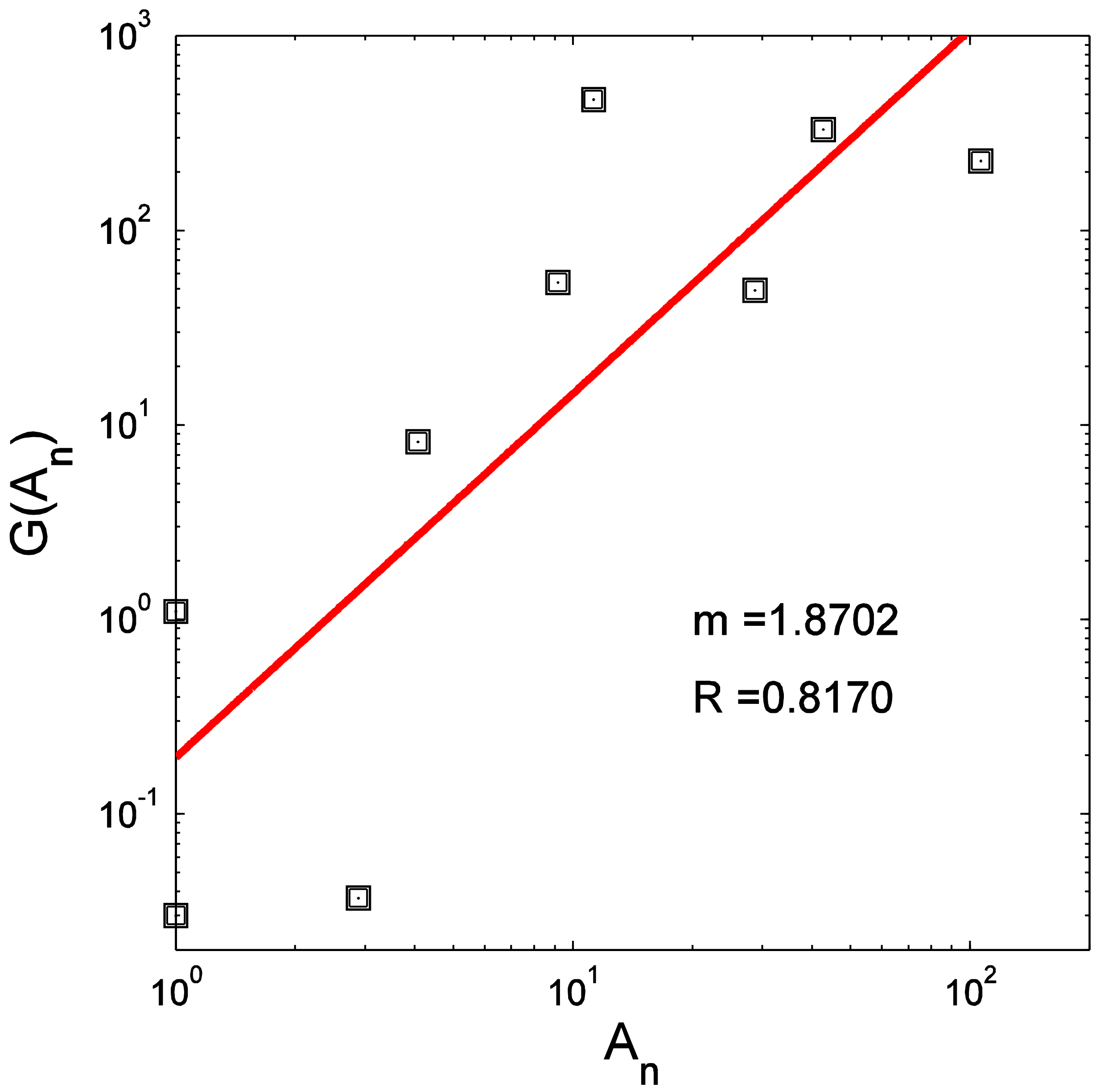
This work has been funded by:

- The Spanish Government through projects TEC2012-38921-C02 MINECO (ERDF/FEDER), IPT-2011-1625-430000 MINECO, IPC-20111009 CDTI (ERDF/FEDER)
- Junta de Andalucía through project TIC 2338-2012 CEICE
- the Office of Naval Research (USA) through grant no. N000141410355.

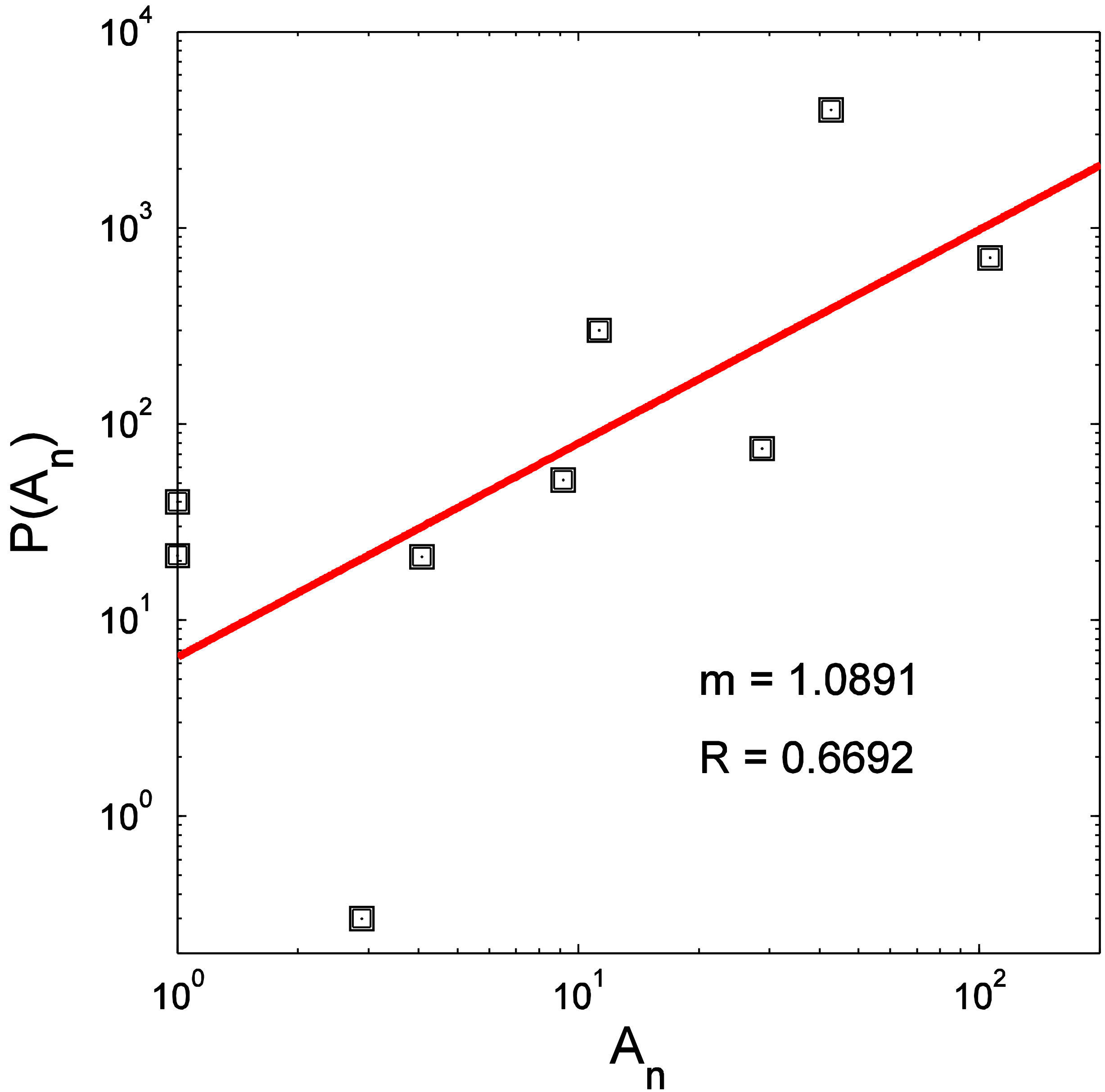
Analog array processor examples

First author	Year	Tech. (nm)	Nproc	Clk (MHz)	Area (mm2)	Power (mW)	GOPS
Minsu Kim	2009	130	4	200.00	4.30	51.8	54.00
Jinwook Oh	2011	130	1025	200.00	13.50	75.0	49.14
Seungjin Lee	2010	130	6412	400.00	50.00	704.0	228.00
Wen-Chia Yang	2011	350	1024	10.00	13.86	21.0	8.19
Jendernalik	2013	350	4096	10.00	9.80	0.3	0.0369
Linan	2004	350	16384	100.00	145.18	4000.0	330.00
Carmona	2003	500	2048	10.00	78.33	300.0	470.00
Dudek	2005	600	441	2.50	10.00	40.0	1.10
Graupner	2003	600	512	—	10.00	21.3	0.03

Power vs. complexity



Performance vs. complexity



References

- G. M. Amdahl, “Validity of the single processor approach to achieving large scale computing capabilities,” in *Proc. of the Spring Joint Computer Conference (AFIPS),* 1967, pp. 483–485.
- S. Borkar, “Thousand core chips: a technology perspective,” in *Proceedings of the 44th annual Design Automation Conference (DAC),* 2007, pp. 746–749.
- J. L. Gustafson, “Reevaluating Amdahl’s law,” *Communications of the ACM*, vol. 31, no. 5, pp. 532–533, May 1988.
- J. L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach* (4th ed.). Morgan Kaufman, 2006.
- M. Hill and M. Marty, “Amdahl’s law in the multicore era,” *Computer*, vol. 41, no. 7, pp. 33 –38, July 2008.
- S. Moreno-Londono and J. Pineda de Gyvez, “Extending Amdahl’s law for energy-efficiency,” in *Int. Conf. on Energy Aware Computing (ICEAC),* Dec. 2010, pp. 1–4.
- M. V. Wilkes, “The memory gap and the future of high performance memories,” *SIGARCH Computer Architecture News*, vol. 29, no. 1, pp. 2–7, Mar. 2001.