How much of the global information and communication explosion is driven by *more*, and how much by *better* technology?

Martin Hilbert
United Nations ECLAC, Dag Hammarskjold 3477, Santiago, Chile;
Pontificia Universidad Católica de Chile, Facultad de Comunicaciones.
Email: martinhilbert@gmail.com

This is a preprint of an article accepted for publication in

Journal of the American Society for Information Science and Technology
(Volume 65, Issue 4, pages 856–861, April 2014)
of The Association for Information Science &Technology
http://www.asis.org/jasist.html

ABSTRACT:

Technological change in the digital age is a combination of both, more and better technology. The article quantifies how much of the technologically-mediated information and communication explosion during the period of digitization (1986 – 2007) was driven by the deployment of additional technological devices, and how much by technological progress in hardware and software. We find that technological progress has contributed between two to six times more than additional technological infrastructure. While infrastructure seems to reach a certain level of saturation at roughly 20 storage devices per capita and 2-3 telecommunication subscriptions per capita, informational capacities are still exploding. Besides progress in better hardware, software for information compression turns out to be an important and often neglected driver of the global growth of technologically mediated information and communication capacities.

Keywords: measurement methodology, information and communication technology, technological change, information theory, infrastructure, hardware, software.

The effort to quantify the amount of information in society goes at least back to the groundbreaking works of Machlup (1962) and Porat (1977), who worked with economic variables, and Ito (1981) and Pool (1983), who counted words. The digital age led to a new generation of studies that quantifies the world's technological capacity to handle information directly in bits and bytes (e.g. Lyman, et al., 2003; Gantz, et al., 2008; Bohn and Short, 2009; Short, Bohn and Baru, 2011; Neuman, Park, and Panek, 2012; for the history and a comparison of the approaches see Hilbert, 2012). We work with the most comprehensive of these exercises (Hilbert and López, 2012a, 2012b), which is based on more than 1,100 different sources (see Appendix A) and provides consistent time series for the period of digitization (1986-2007) for more than 60 analog and digital categories of storage (in bits), communication (in bits per second), and computation technologies (in MIPS). While other studies have analyzed this global information explosion in terms of its technological constituents (Hilbert and López, 2011), its international distribution (Hilbert, 2013a), and its content (Hilbert, 2013b), this article asks how much of the increasing technological capacity has been provided by (a) the installation of additional technological devices and subscriptions, (b) better hardware, and (c) more efficient (compression) software.

Measuring Technological Information and Communication Capacity

Equation (1) is applied to three main *groups g* of technologies (telecommunication, storage, and computation), which consist of the 52 most common *technologies t* in both analog and digital formats (Appendix Tables S-2 - S-4), with 261 t_{kyu} *subtypes of technologies with different performances* for a given year (66 for computation, 172 for storage, and 23 for telecom, for details see Appendix Table S-1, also Hilbert, 2013a, as well as Hilbert and López, 2012a, 2012b).

$$\sum_{\substack{over\ all\ t_{kyu} \\ of\ group\ g}} \left(\left[\text{number of devices or subscriptions } t_{kyu}\right] * \left[\text{performance per device or subscription } t_{kyu}\right]\right) = \text{technological capacity}_{\text{of group g}} \quad (1)$$

- *Storage* is measured in optimally compressed bits and estimates the installed capacity, which evaluates the maximum available storage ("as if all storage were full"). We include the 12 most widely used families of analog (such as books and VHS cassettes) and the 13 most prominent families of digital *storage* (such as hard disks, DVDs and memory cards).
- *Communication* is measured in optimally compressed bits per second and estimates the effective traffic capacity ("only those bits that are effectively transmitted per year"). The inventory covers 6 analog and 5 digital unidirectional *broadcast* technologies (such as radio and TV), as well as 3

bidirectional analog *telecommunication* technologies and their 4 most common digital heirs (fixed and mobile phone and internet).

- *Computation* is measured in instructions per second and represents the maximally available installed hardware capacity ("as if all computers ran all the time"). We include 6 families of general-purpose computers (e.g. PC, servers, videogame consoles) and 3 groups of application-specific embedded computer (digital signal processors, microcontrollers and graphic processing units).

For the case of computation, our measure of performance is equal to the hardware capacity of the computers. For the cases of storage and communication we fine-tune the performance indicator by not only considering the contributions of hardware, but also of software compression algorithms. Depending on the rate of information compression, a certain kind of hardware can hold and communicate different amounts of information bits (Shannon, 1948). Using an analogy, the underlying logic is comparable with filling a certain number of buckets or tubes (infrastructure) of different sizes (hardware) with content of different levels of granularity (software compression). The more fine-grained the filling, the more content units fit into each hardware device. Our data show that the same amount of telecom hardware can roughly send three times more information in 2007 than in 1986, thanks to advances in compression softwares like ZIP, GIF, JPEG or MPEG (see Hilbert, 2011). We have decided to normalize our estimates on the highest conceivable compression rate achievable in 2007, which we call the optimal compression rate (or 2007-entropic compression, since optimal compression approaches the entropy of the source, according to Shannon, 1948) (for more see Appendix B, also Hilbert and López, 2012b).

We can now decompose the growth factor ($^{t+1}/_{t}$) of the total technological capacity to store and communicate information into its three components: (a) the number of devices or subscriptions (infrastructure), (b) the (physical) size of their hardware, and (c) the granularity of their filling (software compression); whereas (b) and (c) represent the "performance of devices" variable from equation (1):

$$\text{(2)} \qquad \frac{Technological\ capacity_{t+1}}{Technological\ capacity_{t}} =$$

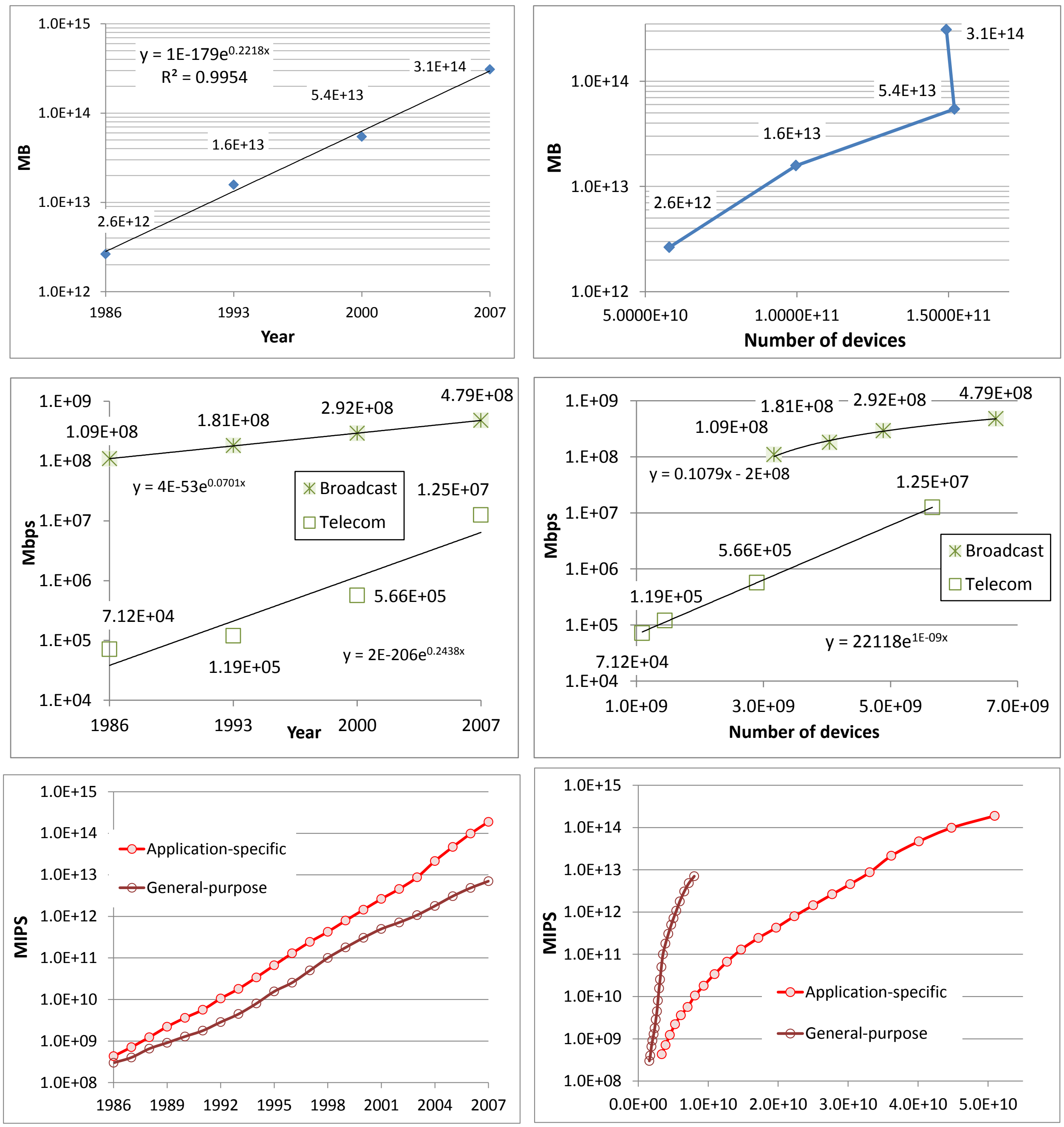
$$= \left(\frac{\#devices_{t+1}}{\#devices_{t}}\right) \times \left(\frac{\emptyset\ hardware\ performance_{t+1}}{\emptyset\ hardware\ performance_{t}}\right) \times \left(\frac{\emptyset\ software\ performance_{t+1}}{\emptyset\ software\ performance_{t}}\right)$$

Appendix B contains a hypothetical illustration of equation (2), as well as a reformulation that facilitates its use in practice (equation S-1).

Quantity versus Performance: More or Better?

Comparing the left-hand side of Figure 1 (informational capacity as a function of years), with the right-hand side (informational capacity as a function of the number of devices) shows that the world's informational capacity has grown exponentially over recent decades, while both, more and better devices made their separate contributions. The case of storage illustrates very clearly that even a stagnating quantity of technology can still result in exponential growth of the installed capacity (compare Figures 1a and 1b). Since 2000, the number of storage devices has reached a level of saturation at the threshold of 22-23 storage devices per capita (see Appendix Figure S-2b). Notwithstanding, the world's total storage capacity in optimally compressed MB has grown with a compound annual growth rate (CAGR) of 28 % per year between 2000 and 2007. This is five times faster than the growth of global GDP during the same years. While broadcast and telecommunication subscriptions (Figure 1c and 1d) have grown at CAGRs of 4 % and 8 % respectively (increasingly reaching a level of saturation in developed countries at around 2-3 telecom subscriptions per capita; see ITU, 2012, p. 179), the world's telecommunication capacity in optimally compressed kbps grew with an average of 28 % per year (broadcast capacity CAGR of 7 %). Figure 1f shows that the number of application-specific embedded computers (such as in household appliances, cars, or monitors) has increased much faster (CAGR of 86%) than the number of humanly guided general-purpose computers (such as PCs or smart phones) (CAGR of 61%). The Figures show that a combination of different trajectories of technological change (consisting of different combinations of more and better technology) leads to distinguishable paths.

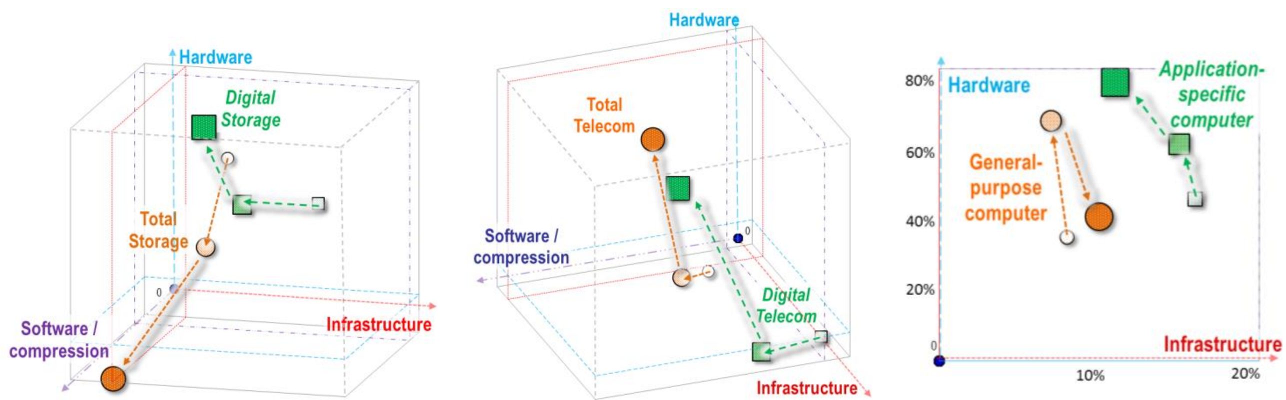
Figure 1: Semi-logarithmic plots of capacity as a function of years (left-hand side) and as a function of installed devices (right-hand side). (a) and (b) installed capacity of storage in optimally compressed Megabytes (MB); (c) and (d) effective traffic capacity of broadcasting and telecommunication in optimally compressed Megabits per second (Mbps); (e) and (f) installed capacity of computation in millions-instructions per second (MIPS).



Inside the black box that drives the global information explosion

The two digit rates of change of our informational capacities dwarf all kinds of orders of magnitude social scientists are used to work with. We can now use equation (2) to decompose these impressive total growth rates into its three main drivers: more technology, better hardware, and better software compression. The results are presented in Table 1 and visualized in Figure 2.

Figure 2: Graphic schematization of Table 1; 1986-1993 (small bright points); 1993-2000 (medium points); and 2000-2007 (large dark points): (a) storage; (b) telecom; (c) computation.



In the case of computation we only deal with two drivers (equation (1), Figure 2c) (we do not consider the improvements of software algorithms due to lack of adequate data, see Appendix B). It shows that better hardware has been the main driver of technological change (i.e. “Moore’s law”). Especially the hardware capacity of application-specific computers has seen extraordinary increases (from CAGR of 47% for 1986-1993; 63% for 1993-2000, and 81% for 2000-2007), which is mainly driven by the outstanding progress of graphic processing units in monitors.

The case of telecommunication considers all three dimensions, including improvements in software compression (equation (2)). Table 1 shows that the early days of digitization of telecom networks (20% digitized in 1986, 69% in 1993, 98% in 2000; see Hilbert and López, 2011) were driven by the contributions of compression of formerly analog phone systems. The period 1993-2000 was characterized by the rapid expansion of infrastructure (mainly driven by the global flood of mobile phones), and 2000-2007 was driven by increases in hardware performance (especially broadband, with a growing weight of fiber-optics, see also the visualization in Figure 2b). During this period, digital telecom infrastructure grew at a CAGR of 15.8 %, but digital telecom hardware performance (more bandwidth per subscription) grew with 18.5 % annually. Together with the contribution of software compression (CAGR 13.9 %) digital telecommunication capacity grew 56.2 % per year (see Table 1; following equation (2): 1.158 * 1.185 * 1.139 = 1.562).

Software compression has also played an important role during the time of digitization of the world's storage capacity (global storage capacity was 25 % digitized in 2000, and 94% in 2007; see Hilbert and López, 2011), contributing with a 38 % growth rate to technological memory during 2000-2007. Average hardware performance per digital storage device continuously increased, while the hardware contribution of the average storage device constantly declined (in relative terms) due to sharply decreasing average analog performance (given a sharp decline in bit-heavy analog video cassettes, like VHS, i.e. during 2000-2007). While these high-level numbers disguise the fact that the total capacity consists of a myriad of different technological solutions with vastly heterogeneous average performances (see Appendix, Figure S-9), a general tendency is unequivocal.

Conclusions

A decomposition of technological change in the digital age shows that the world's technological capacity to store, communicate and compute information has mainly been driven by better, and not merely more technology. The only exception to this trend is broadcasting, which was hardly digitized during the period of the study (only 25 % of broadcasted information was digitized in 2007). Our approach includes progress in software compression and it shoes that software compression rates play an important and often neglected role when quantifying informational capacities, especially during periods of digitization.

On the one hand, this finding is of methodological interest and stresses the importance of normalizing information capacities on justifiable compression rates (see Hilbert and López, 2012b). On the other hand, it is also of practical relevance, because the vast majority of digital policies are based on mere proxies of our informational capacities, such as on statistics of the number of technological devices and subscriptions (Hilbert, 2011; ITU, 2012, Ch.5; Hilbert, 2013). We have seen that the amount of technological devices and subscriptions are increasingly playing a secondary role in the digital age. This implies that the basic roll-out of the technological infrastructure of the global "Network Society" (Castells, 2009) has reached a certain level of saturation, while the capacity on basis of this fundamental infrastructure is continuously advancing. Despite stagnating stocks of devices and subscriptions, we continue to expand our technological information, communication, and computation capacities through incessant technological progress. This implies that any quantitative assessment of the digital age requires an accounting that goes beyond the traditional inventory of technological devices and subscriptions, and captures informational capacities in terms of both more and better technologies (see e.g. Hilbert, 2013).

Table 1: Compound annual growth rates (CAGR) for 1986-1993; 1993-2000; 2000-2007 (percentage points) of the drivers of storage (in optimally compressed MB); broadcasting and telecom (in optimally compressed Mbps); and computation (in MIPS) according to equation (2). Note: * Or year of introduction.

STORAGE	Drivers	CAGR 86-93	CAGR 93-00	CAGR 00-07	CAGR 86*-07	Main driver of technological change
TOTAL storage	Infrastructure	8.1	6.2	-0.3	4.6	Tech. progress software (compression)
	Hardware	**24.1**	**9.2**	-6.9	8.0	
	Software (content compression)	-3.7	3.0	**38.0**	**11.0**	
	TOTAL	**29.1**	**19.4**	**28.2**	**25.5**	
Digital storage	Infrastructure	**26.1**	15.4	7.3	16.1	Tech. progress hardware
	Hardware	19.3	**20.6**	**33**	**24.2**	
	Software	5.5	14.2	8.4	9.3	
	TOTAL	59.0	59.0	54.7	57.6	
BROADCAST						
TOTAL broadcast	Infrastructure	3.5	2.8	**4.5**	**3.6**	Diffusion
	Hardware	**4.0**	**3.2**	-0.1	2.4	
	Software (content compression)	-0.2	0.9	2.8	1.2	
	TOTAL	7.5	7.0	7.3	7.3	
Digital broadcast	Infrastructure	-	-	**36.7**	**36.7**	Diffusion
	Hardware	-	-	-6.3	-6.3	
	Software	-	-	-0.1	-0.1	
	TOTAL	-	-	27.9	27.9	
TELECOM						
TOTAL telecom	Infrastructure	4.1	10.4	10.0	8.1	Tech. progress software (compression)
	Hardware	-1.9	1.7	**23.2**	7.1	
	Software (content compression)	**5.4**	**11.3**	14.9	**10.4**	
	TOTAL	7.7	24.9	55.7	27.9	
Digital telecom	Infrastructure	**28.5**	**25.9**	15.8	**23.3**	Diffusion
	Hardware	-0.1	-2.4	**18.5**	4.9	
	Software	0.2	6.9	13.9	6.8	
	TOTAL	28.6	31.4	56.2	38.2	
COMPUTATION						
TOTAL General-purpose	Infrastructure	8	7	10	8	Tech. progress hardware
	Hardware	**36**	**70**	**42**	**49**	
	TOTAL	47	82	56	61	
TOTAL Application-specific	Infrastructure	16	15	11	14	Tech. progress hardware
	Hardware	**47**	**63**	**81**	**63**	
	TOTAL	70	87	101	86	

References

Bohn, R., & Short, J. (2009). *How Much Information? 2009 Report on American Consumers*. Global Information Industry Center, UC San Diego.

Castells, M. (2009). *The Rise of the Network Society: The Information Age: Economy, Society, and Culture Volume I* (2nd ed.). Wiley-Blackwell.

Gantz, J., Chute, C., Manfrediz, A., Minton, S., Reinsel, D., Schlichting, W., & Toncheva, A. (2008). *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011*. IDC (International Data Corporation) sponsored by EMC.

Hilbert, M. (2011). Mapping the dimensions and characteristics of the world's technological communication capacity during the period of digitization. *Working Paper*. Presented at the 9th World Telecommunication/ICT Indicators Meeting, Mauritius: International Telecom.Union (ITU).

Hilbert, M. (2012). How to Measure "How Much Information"? Theoretical, methodological, and statistical challenges for the social sciences. *International Journal of Communication*, *6* (Introduction to Special Section on "How to measure 'How-Much-Information'?"), 1042–1055.

Hilbert, M. (2013a). Technological information inequality as an incessantly moving target: Redistribution of the global, international and national information and communication capacities between 1986 and 2010. currently under review.

Hilbert, M. (2013b). What Is the Content of the World's Technologically Mediated Information and Communication Capacity: How Much Text, Image, Audio, and Video? The Information Society, 30(2), 127–143.

Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, *332*(6025), 60 –65.

Hilbert, M., & López, P. (2012a). How to Measure the World's Technological Capacity to Communicate, Store and Compute Information? Part I: results and scope. *International Journal of Communication*, *6*, 956–979.

Hilbert, M., & López, P. (2012b). How to Measure the World's Technological Capacity to Communicate, Store and Compute Information? Part II: measurement unit and conclusions. *International Journal of Communication*, *6*, 936–955.

Ito, Y. (1981). The Johoka Shakai approach to the study of communication in Japan. In C. Wilhoit & H. de Bock (Eds.), *Mass Communication Review Yearbook* (Vol. 2, pp. 671–698). CA: Sage.

ITU (International Telecommunication Union). (2012). *Measuring the Information Society 2012*.

Lyman, P., Varian, H., Swearingen, K., Charles, P., Good, N., Jordan, L., & Pal, J. (2003). *How much information? 2003*. UC at Berkeley.

Machlup, F. (1962). *The production and distribution of knowledge in the United States*. Princeton University Press.

Neuman, R., Park, Y., & Panek, E. (2012). Tracking the Flow of Information into the Home: An Empirical Assessment of the Digital Revolution in the U.S. from 1960 - 2005. *International Journal of Communication*, Special Section on "How to measure 'How-Much-Information'?"(6), 1022–1041.

Pool, I. de S. (1983). Tracking the Flow of Information. *Science*, *221*(4611), 609–613.

Porat, M. U. (1977, May). *The Information Economy: Definition and Measurement*. U.S. Government Printing Office, Washington, (Stock No. 003-000-00512-7).

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.

Short, J., Bohn, R., & Baru, C. (2011). *How Much Information? 2010 Report on Enterprise Server Information*. Global Information Industry Center, UC San Diego.