From Bit to It: How a Complex Metabolic Network Transforms Information into Living Matter

Andreas Wagner

SFI WORKING PAPER: 2007-09-035

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu

SANTA FE INSTITUTE

From bit to it: How a complex metabolic network transforms information into living matter.

Andreas Wagner

Department of Biochemistry, University of Zurich
Building Y27, Winterthurerstrasse 190
CH-8057 Zurich, Switzerland
Email: aw@bioc.unizh.ch
Phone (Secr.): + 41-44-635-6142
FAX: +41-44-635-6144

Abstract

Organisms live and die by the amount of information they acquire about their environment. The systems analysis of complex metabolic networks allows us to ask how such information translates into fitness. A metabolic network transforms nutrients into biomass. The better it uses information on available nutrient availability, the faster it will allow a cell to divide. I here use metabolic flux balance analysis to show that the accuracy *I* (in bits) with which a yeast cell can sense a limiting nutrient's availability relates logarithmically to fitness as indicated by biomass yield and cell division rate. For microbes like yeast, natural selection can resolve fitness differences of genetic variants smaller than 10^{-6}, meaning that cells would need to estimate nutrient concentrations to very high accuracy (>22 bits) to ensure optimal growth. I argue that such accuracies are not achievable in practice. Natural selection may thus face fundamental limitations in maximizing the information processing capacity of cells. The analysis of metabolic networks opens a door to understanding cellular biology from a quantitative, information-theoretic perspective.

Background

Organisms need to acquire information about their environment in order to survive and reproduce. They need to respond to information about changes in temperature, soil conditions, water availability, nutrient supply, predation pressure, and many other factors. The ability to acquire and use such information arguably affects organismal fitness [1-6]. However, we know nothing about the quantitative relationship between such information and fitness.

In microbes, an important fitness component is a cell's growth or cell division rate. The selection pressure to grow rapidly during times of nutrient availability has left clear traces in the evolutionary record, such as strong microbial codon usage biases that allow high translation efficiency of abundant proteins [7, 8]. It has recently become possible to make quantitative predictions about a cell's maximal division rate, based on nearly complete information about the metabolic networks that sustain cellular life [9-16]. These metabolic networks comprise of the order of 10^3 chemical reactions for free-living organisms. Their structure has been elucidated in several organisms by manual curation, aided by functional genomic data [9, 11, 12]. Flux balance analysis allows one to predict those flows of matter –metabolic fluxes – through each reaction of a network that are consistent with the laws of mass conservation. More precisely, flux balance analysis predicts ratios of metabolic fluxes, i.e., values of metabolic fluxes relative to a reference flux that needs to be determined independently, for example through experimental measurements. Together with the known biomass composition of an organism, flux balance analysis can then also identify the metabolic fluxes that maximize biomass production. For organisms such as *Saccharomyces cerevisiae* and common growth substrate compositions, such as minimal media with glucose as a sole carbon source, the predictions of flux balance about maximal biomass yield are in good agreement with experimental data, where available [14, 17, 18]. For other organisms and more unusual environments [19-21] this does not always hold. Strikingly, however, even in this case laboratory evolution experiments can produce strains of organisms that show the maximally predicted biomass yield within a short amount of time [13, 19, 20, 22]. I here use flux balance analysis of the yeast metabolic network to explore the relationship between environmental information and how rapidly an organism produces biomass per unit time.

The kind and concentration of available growth substrates influence a cell's maximal biomass yield [23-26]. Cells have developed elaborate nutrient sensing mechanisms to respond to changes in nutrient abundances. For example, in yeast, dedicated glucose sensor proteins (*Snf*3, *Rgt*2) as well as glycolytic intermediates form the beginning of a signaling cascade. This cascade produces an integrated cellular response that includes the expression of glucose transporter genes (*HXT1-HXT7*), the expression of glycolytic genes, as well as the repression of many other genes [24, 25]. Even though long-studied, glucose sensing is still only qualitatively and incompletely understood. This holds to an even greater extent for other sensing mechanisms, such as those for phosphate and nitrogen [24, 25]. The accuracy of the sensing mechanism is clearly important for optimal growth, but there is an important asymmetry: Overestimation of a nutrient concentration, e.g. through overexpression of catabolic

genes, will not lead to sub-optimal growth due to metabolic undercapacity, because the available nutrients can still be maximally used. In contrast, underestimation and the resulting undercapacity will lead to reduced growth. Information is thus of greatest value from a metabolic perspective if it prevents underestimation of nutrient concentrations, and thus undercapacity of a nutrient utilization system. I will thus focus primarily on the consequences of sensing errors that lead to underestimates of nutrient concentrations. Note that overestimation of nutrient concentration may lead to sub-optimal growth for other reasons, reasons that cannot currently be modeled in a metabolic context, and that are discussed below.

Information acquired through nutrient sensing can be represented as follows. Consider k nutrients and their actual concentrations N_i ($1 \leq i \leq k$) in a cell's environment. If a cell underestimates the actual nutrient concentration in the environment for any one nutrient, then its "measurement" N_i^m of the actual nutrient concentration is such that $N_i^m < N_i$. Now subdivide the interval $(0, N_i)$ into n_i equal subintervals. If the cell can place (measure) the concentration of i within the interval $(N_i(n_i-1)/n_i, N_i)$ then it has $I_i = \log_2 n_i$ bits of information about this nutrient. If the measurement error, when viewed as a random variable, has a symmetric or a uniform distribution *within this interval*, then the expected sensing error is $E_i = N_i / 2n_i$. Nutrient information and sensing error thus relate to each other as

$$I_i = \log_2\left(\frac{N_i}{2E_i}\right)$$

Although E_i and I_i are equivalent, using I_i has two advantages. First, its units (bits) are canonical measures of information [27]; second, information content is additive, that is, if a cell has I_j bits of information on nutrient j ($1 \leq j \leq k$), then it has $I = \sum_j I_j$ bits of information about its entire nutrient environment, if the nutrients occur independently from one another, or if the cell measures them independently from each other.

Flux balance analysis allows us to immediately assess the fitness value of nutrient information for a cell, because a cell's maximal biomass yield is a function of the measured nutrient concentrations $(N_1^m, \ldots, N_k^m)$ and the extent to which these concentrations differ from their actual value $(N_1, \ldots, N_k)$.

Results

Diminishing returns on improved information acquisition

The relationship between information and fitness is best explored for a defined environment, such as a minimal growth medium. The environment I use contains NH_3, inorganic phosphate (P_i), sulfate, and glucose as the sole carbon source. Oxygen is available as a terminal electron acceptor. For simplicity, I first focus on a scenario where

information about all substrates except glucose is perfectly accurate. I assume that the biomass yield Y per unit time is linearly proportional to a cell's division rate G, a measure of fitness. In other words, $Y=cG$, c being some constant. I express the effect of incomplete information on biomass yield Y as $s=1-Y/Y_{\max}=1-G/G_{\max}$, where Y_{max} and $G_{\max}$ are the maximally achievable biomass yields and cell division rates, respectively, i.e., the yields and rates for perfectly accurate glucose information. The quantity s can also be thought of as a selection coefficient, as a measure by how far a cell's fitness $w=1-s=G/G_{max}$ is reduced by incomplete information. Figure 1a shows how a cell's fitness depends on the amount of information the cell can acquire about substrate concentration. Specifically, the figure shows that the logarithm of fitness depends linearly on information in bits. The relationship of s and information is especially simple if a binary logarithm is used to scale s, i.e., $-\log_2(s)=I_{GLC}+1$. This simple relationship emerges numerically from flux balance analysis, but it also has a straightforward intuitive explanation. If zero bits of information are available for a growth-limiting nutrient, then under the assumptions used here, the cell's "guess" about nutrient concentrations will be randomly distributed in the interval $(0, N_i)$, with an expected value of $N_i/2$. At this expected value, the division rate of a cell will be half the maximal growth rate, such that $s=1/2$. The above relationship between s and I then holds, because $-\log_2(1/2)=1=I+1$. If one bit of information is available ($I=1$), then the cell's measurement will be randomly distributed in the interval $(N_i/2, N_i)$, with an expected value of $3N_i/4$, leading to $s=1/4$, and $-\log_2(1/4)=2=I+1$. The same line of reasoning applies to ever increasing values of I. The key assumption in this intuitive explanation is that if one nutrient is growth-limiting, then cell division rate depends linearly on the cell's ability to utilize this nutrient. This is not obvious *a priori*, because the nutrient's metabolic products may be fed into many different pathways that produce essential biomass components. The distribution of these products among different pathways, and the cell's final resulting division rate, might in principle depend on the concentration of the nutrient *and* on that of other nutrients. However, flux balance analysis shows that the dependency between nutrient concentration and biomass yield is quite simple and linear.

Nutrient sensing has much greater impact if the amount of information acquired by a cell is low (Figure 1a). For instance, an increase in available glucose information from 1 bit (low-high) to 2 bits (four distinct concentration values) causes a 36% increase in growth rate, whereas an increase from 14 to 15 bit causes an increase of 0.0012 percent. For the purpose of comparison, the figure also shows the relationship between fitness reduction and sensing error in percent. Fitness reduction s decreases linearly with decreasing sensing error (note the double-logarithmic scale). Quantitatively very similar linear-log and linear relationships hold for the other four nutrients (data for oxygen and ammonium are shown in the inset for Figure 1a). In sum, the logarithm of division rate scales linearly with nutrient information in bits, and increased information acquisition carries diminishing fitness returns.

Even very small sensing errors cause adaptively significant growth-rate differences

How large must a growth rate difference s (due to imperfect nutrient sensing) be in order to matter to natural selection? The influence of genetic drift dominates over that of natural selection, if a reduction in growth rate is $s<1/4N_e$ for diploid cells, where N_e is the

effective population size [28, 29]. N_e, in turn, can be estimated from the synonymous nucleotide diversity π in a population, and the per-generation mutation rate μ as $N_e=\pi/4\mu$. Thus, if a growth rate difference is smaller than $s=\mu/\pi$ in a population, then the associated growth rate difference is too small to be seen by natural selection. The rate of mutations per nucleotide and generation in *Saccharomyces cerevisiae* has been estimated at $\mu=2.2\times10^{-10}$ [30]. In the closest wild relative of yeast synonymous π has been estimated as $\pi=0.003$ [31]. The above parameters yield $s=7.33\times10^{-8}$ as a "critical" growth rate difference that can still be seen by natural selection. No data on *synonymous* nucleotide diversity are available for *S. cerevisiae* itself, but a recent estimate [32] on *overall* nucleotide diversity of $\pi=0.0046$ (which is typically smaller than synonymous diversity) suggests an upper bound of $s=4.78\times10^{-8}$, rendering the critical s I use here conservative. Growth rate differences below this value are more strongly influenced by drift than by selection.

The critical selection coefficient s is indicated in Figure 1a by a horizontal line. Below this line, any gains in information do have negligible effect. Specifically, information gains exceeding $I=-\log_2 s-1\approx22$ bits are selectively neutral for yeast.

Reduced fitness value of information for imperfect sensing of several nutrients.

Thus far, I assumed limited information for only one nutrient, but what if information is limited for more than one nutrient? Consider a genotype that systematically underestimates the availability of one substrate, such as glucose, because of a poor sensing mechanism. The resulting undercapacity to metabolize this substrate renders the substrate growth-limiting. In this situation, accurate sensing of the availability of another substrate, such as ammonium may not increase fitness. The reason is that growth is limited not by a lack of information about ammonium, but by lack of information about glucose. As an example, Figure 1b shows how information about ammonium and glucose abundance (x- and y-axes) interact to produce observed growth rate differences (z-axis). If ammonium sensing is highly accurate, then increasing information about glucose concentrations causes a linear increase in fitness (the plane parallel to the yz-axes, at an accuracy of 24 bits for ammonium availability) exactly as in Figure 1a. However, if ammonium sensing is poor (the same plane, but at zero bits of ammonium information) then good glucose sensing yields no growth-rate gain, because it is the poor ammonium-sensing that effectively limits growth. If ammonium-sensing accuracy is intermediate, then an improvement in glucose sensing causes a growth-rate increase (smaller s) up to some number of bits. From that point on, additional glucose-information has no effect, because ammonium-sensing has become growth-limiting. Exactly the same considerations hold if the places of ammonium or glucose are interchanged (or for any other two nutrients), which causes the symmetry of the piecewise linear surface in Figure 1b.

Figure 1c shows a different representation of the relationship between the fitness value of information about one growth substrate, and variation in information about another growth substrate. The figure shows how increasing information about glucose concentration (horizontal axis) affects biomass yield and thus fitness (vertical axis), if at the same time information measured about one or more *other* nutrients varies randomly. For example, the black bars correspond to a situation where only one other substrate,

oxygen, is measured at accuracies randomly and uniformly distributed between 0 and 16 bits. For increasing glucose information, the growth rate loss becomes smaller, but never as small as when *only* glucose information limits growth (note the logarithmic and linear vertical axes in Figures 1a and 1c, respectively). Even at glucose information of 16 bits, the biomass yield loss by a cell relative to the maximal growth rate is of the order of s=0.02 (2 percent, Figure 1c, black bars), several orders of magnitude greater than the $s \approx 10^{-8}$ observed if oxygen sensing is perfectly accurate. This means that inaccurate sensing of one growth substrate severely limits the value of information about other growth substrates. This limitation becomes more severe as the number of growth substrates for which uncertainty exists increases (black to white bars in Figure 1c). Although the data is shown for specific growth substrates, the results are insensitive to the kind of nutrients for which imprecise information is available.

Figure 1d, finally, shows the effect on biomass yield of the total amount of information available, i.e., summed over five key growth substrates in a minimal glucose medium, where information on each growth substrate can vary between 0 and 16 bits. Biomass yield increases (s decreases) only slightly as the amount of available information increases, until much information (>60 bits) becomes available, at which point every additional bit has a large effect. This means that the benefits of growth substrate information are limited by the poorest sensing process in a metabolic system. Only if every substrate-sensing process has high accuracy, does increasing information about any one substrate provide large benefits.

Discussion

The notion that cells process information is not new [33]. However, it is usually expressed qualitatively, without reference to the amounts of information involved and what exactly is being processed. Using metabolic flux balance analysis, I here take a small step towards a quantitative approach to information processing in cells. Specifically, I show that nutrient sensing inaccuracy is translated into reduced cell growth. In the simplest possible case of information limitation in only one nutrient, the relationship between cell growth and information (in bits) is best expressed as $-\log_2(s)=I+1$, where s is a selection coefficient, the difference between the maximal growth rate with perfect sensing and the actually attained growth rate.

Population genetic considerations show that very small selection coefficients of $s<10^{-6}$ are still visible to natural selection in microbes like yeast. Very small inaccuracies (≈0.0001 percent measurement error; Figure 1a) can thus still lead to growth rate loss with evolutionary consequences. Yeast cells would need to sense nutrients at accuracies greater than 22 bits to ensure optimal growth. Organisms with larger population sizes would need even greater accuracies.

Although nutrient sensing mechanisms are only incompletely understood [23-25], several lines of evidence suggest that the needed accuracies are unlikely to be achievable in practice. First, with an average volume of 9.5×10^{-14} liter for a yeast cell [34], a typical nutrient concentration of 10 mmol l^{-1} translates into 5.72×10^{8}molecules. Temporal random fluctuations scale as the square root of the number of molecules [35], and will be of the order of 0.004 percent, more than an order of magnitude greater than the necessary

accuracy. At physiological nutrient concentrations, the needed sensing accuracy thus is greater than random fluctuations in molecule numbers.

Second, although it is generally unknown how accurately cells can sense molecule concentrations, some benchmarks come from the accuracy with which cells can sense concentration *differences*. Eukaryotic cells, including yeast, can detect concentration differences of 1-10% across the length of a cell [36]. Much smaller *E. coli* cells can detect concentration differences of 0.01% by integrating information over time during chemotactic swimming [37-41]. However, these accuracies are two orders of magnitude or more smaller than the measurement error associated with 22 bit sensing accuracy of absolute nutrient concentrations.

A third line of evidence comes from measurements of gene expression noise. For optimal utilization of a nutrient, several classes of molecules need to be expressed at a minimum level determined by the nutrient concentration. These include the nutrient sensors, the signaling molecules needed to communicate the sensing information to regulators of gene expression, and the nutrient utilization enzymes themselves. Sensing is suboptimal if any one of these classes of molecules is expressed at too low a level. Concentrations of all gene products fluctuate in a cell due to gene expression noise [38-42]. Although highly expressed proteins show low expression noise, even highly expressed yeast proteins may fluctuate in concentration by about 10% around their mean [41]. If concentrations of sensing and utilization molecules need to be fine-tuned for high sensing accuracy, then sufficiently high sensing accuracy is not realistic. In sum, fluctuations of nutrient concentrations, limits to detection of concentration differences, and gene expression noise will conspire to prevent high-accuracy sensing of nutrient concentrations needed for optimal growth.

A number of caveats to this approach are in order. First, I have here emphasized sensing errors that lead to underestimation of substrate availabilities, because only such errors lead to an undercapacity to metabolize nutrients. Overestimation would lead to overcapacity of nutrient utilization systems, which in itself would still lead to maximal growth. However, if overestimation causes a systematic overexpression of signaling or utilization molecules, overestimation could carry an increased cost of gene expression. Even though few genes might be involved, these costs are not necessarily low. For example, in case of the lactose operon of *E. coli*, overexpression of only three lactose utilization genes in the absence of lactose leads to a 5% reduction in growth rate (s=0.05) [26], which is a very large fitness loss compared to the small values visible to natural selection [43]. Hundreds of genes can change expression in response to nutrient availability in yeast [44]. Because many of these genes do not encode metabolic enzymes, it is not straightforward to predict the ensuing change in expression cost from a metabolic model. In addition, available *quantitative* information about expression changes for such genes has very limited accuracy for mRNA, and is generally unavailable for proteins. Thus, although it would be highly desirable to understand both the growth cost of underestiming and overestimating nutrient concentrations [44], a quantitative analysis of such costs must await more complete characterization of sensing pathways and gene expression changes therein.

A second caveat is that selection may act concurrently on multiple attributes of a metabolic system, not only on nutrient sensing. One example comes from glucose limitation experiments in chemostats, where a population's environment is held constant

for hundreds of generations. Consider a nutrient whose actual concentration is N, and the value of its concentration sensed by a cell (possibly with some inaccuracies) is $S(N)$. If the import system for this nutrient is far from being saturated, which is likely if the nutrient is at concentrations sufficiently low to be growth limiting, then the cell's uptake rate U of this nutrient is likely to be proportional to $S(N)$, with some proportionality constant c, i.e., $U=cS(N)$. In this paper, I focus on the value of improved nutrient sensing $S(N)$. However, in addition, the proportionality constant c itself may be subject to selection pressure, thereby increasing the efficiency of uptake for a given $S(N)$. For example, in yeast populations cultivated during several hundred generations in a chemostat, such increase in uptake efficiency occurs through gene duplication and increased expression of hexose transporter genes [45, 46]. Note, however, that the constant chemostat conditions of such experiments are likely to be rare in the wild, where an exponentially growing population rapidly exhausts any limiting nutrient source.

A final caveat is that it may not always be possible to sense the availability of two compounds independently from one another. One example is the sensing of glucose and protons (H^+), where in yeast the glucose sensor *Snf3p* is known to activate the proton pumping plasma membrane ATPase [47].

Nutrient sensing systems are only as strong as their weakest link: Inaccurate sensing of one nutrient may strongly reduce the fitness benefits of high quality sensing of other nutrients. However, it is easy to see how multiple independent mutations, each in a different nutrient sensing system, may favor incremental improvements in the sensing of all nutrients through natural selection. The reason is that an allele that increases sensing quality for one nutrient will increase fitness whenever that nutrient is limiting, and be driven to fixation during such times. This increases the value of better information for other nutrients, and favors alleles that improve information acquisition for these nutrients, thus increasing the respective mutations in frequency, and so on. At the end point of many such evolutionary cycles stands a cell that achieves the best possible nutrient sensing, given biophysical and population size constraints.

Recent work suggests that the lens of natural selection can see seemingly minute changes in transcriptome and proteome composition, such as single amino acid changes and small changes in the expression of one gene [48]. The observations made here likewise emphasize the importance of natural selection to shape nutrient sensing accuracy. In addition, they suggest the existence of biophysical constraints that may severely limit the outcome of selection on high-accuracy nutrient sensing to biophysically achievable, but suboptimal solutions. This perspective only becomes possible through a system-wide analysis of a metabolic network. An important task of future work would be to *quantify* the constraints natural selection faces in optimizing how cells acquire information.

Methods

Flux balance analysis [49] uses information about the stoichiometry of all enzymatic reactions known to occur in an organism, which is encapsulated in a stoichiometry matrix **S**. At a metabolic steady-state, the vector of allowable metabolic fluxes **v** describing the rates through each reaction in the network must fulfill the condition **Sv**=0 so as to not violate mass conservation. For each **v** that is a solution of this equation, c**v** (c being some real constant) is also a solution, such that one can think of **v** as specifying relative *ratios* of fluxes through a metabolic network. Further constraints, such as irreversibility of some reactions, and experimentally measured uptake fluxes of external substrates can reduce the number of allowable fluxes **v** in steady-state. Within the space of allowable fluxes one can then use linear programming to determine the fluxes that maximize or minimize any quantity that can be expressed as linear combinations of individual metabolic fluxes. An especially important such quantity is the biomass growth flux itself.

The following substrate uptake fluxes (mmol substrate /h /g dry weight) were used here (rounded to two significant digits): Glucose: 15.3; O_2: 2.4; Sulfate: 0.034; inorganic phosphate P_i: 0.09; ammonium NH_3: 2.45. Among these values, the glucose uptake flux v_G and oxygen uptake flux v_{O2} stem from experimental measurements in an aerated batch culture of *S. cerevisiae* [50]. I constrained these two fluxes and then determined the maximal growth flux v_{max} given these constraints, where the stoichiometry of the growth flux is given in [12]. I then constrained the growth flux to v_{max} (while still constraining v_G, v_{O2} to the above values), and minimized the sulfate uptake flux for which this growth flux could be observed. The resulting sulfate uptake flux (see value above) is the smallest sulfate uptake flux that can sustain the observed v_{max} with the given glucose uptake flux. I subsequently carried out analogous minimization procedures for the remaining two growth substrates, thus arriving at the values listed above. This combination of values has the advantage that reduction in any one nutrient uptake flux will lead to a reduction in growth rate. In other words, no nutrient is in excess, and accurate nutrient availability estimation is critical to sustain maximal growth. All analyses were carried out with a publicly available yeast metabolic model [12] and with the FBA package "sbrt" (Wright and Wagner, unpublished), using the commercial linear programming package CPLEX (ILOG, Mountain View, CA.). To estimate expected growth rates for an amount of information I_i available for a given nutrient i, I translated I_i into the appropriate number of measurement intervals n_i, according to the relation $I_i=\log_2 n_i$ discussed in the main text. The expected measured nutrient value then calculates as $N_i(2n_i-1)/2n_i$, where N_i is the actual nutrient concentration represented through one of the uptake fluxes listed above. Flux balance analysis was carried out with this expected value to determine the growth rate achievable for the corresponding amount of information.

As stated, nutrient concentrations are here represented through nutrient uptake rates. That is, I implicitly assume that if the concentration of a *limiting* nutrient changes by x%, then a cell with access to perfect nutrient sensing would also change its uptake rate of the nutrient by x%. However, it must be clarified that nutrient uptake rates are not necessarily proportional to nutrient concentrations, even if cells have perfect information. Specifically, if there is so much of a nutrient that the nutrient uptake transporters are

saturated, changes in extracellular nutrient concentrations will have no effect on nutrient uptake. However, this scenario is very different from my focus here, namely an environment where the concentration of individual nutrients limits growth, and where information about such nutrients matters to the cell.

Authors' Contributions

AW designed and executed the research, analyzed the data, and wrote the manuscript.

Acknowledgments

I would like to thank Jeremiah Wright for many valuable discussions, as well as Attila Becskei for useful comments, and the paper's reviewers for valuable suggestions. This work was in part supported by SNF grant 315200-116814. I thank the Santa Fe Institute for its continued support.

References

1. Bergstrom C, Lachmann M: Shannon information and biological fitness. In: *IEEE Workshop on Information Theory: 2004*: IEEE, Pasadena, CA; 2004: 50-54.
2. Stephens D, Krebs J: Foraging theory. Princeton, NJ: Princeton University Press; 1986.
3. Seger J, Brockmann HJ: What is bet-hedging? . *Oxford Surverys in Evolutionary Biology* 1987, 4:182-211.
4. Kussell E, Kishony R, Balaban NQ, Leibler S: Bacterial persistence: A model of survival in changing environments. *Genetics* 2005, 169(4):1807-1814.
5. Smith JM: The idea of information in biology. *Quarterly Review of Biology* 1999, 74(4):395-400.
6. Stephens DW: Variance and the value of information. *American Naturalist* 1989, 134(1):128-140.
7. Sharp PM, Li WH: The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Molecular Biology and Evolution* 1987, 4(#3):222-230.
8. Sharp PM, Li W-H: The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* 1986, 15(3):1281-1295.
9. Edwards JS, Palsson BO: Systems properties of the Haemophilus influenzae Rd metabolic genotype. *Journal of Biological Chemistry* 1999, 274(#25):17410-17416.
10. Schilling CH, Schuster S, Palsson BO, Heinrich R: Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress* 1999, 15(#3):296-303.
11. Edwards JS, Palsson BO: The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the*

National Academy of Sciences of the United States of America 2000, 97(10):5528-5533.
12. Forster J, Famili I, Fu P, Palsson B, Nielsen J: Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research* 2003, 13:244-253.
13. Fong SS, Palsson BO: Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. *Nature Genetics* 2004, 36(10):1056-1058.
14. Blank LM, Kuepfer L, Sauer U: Large-scale C-13-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biology* 2005, 6(6):R49.
15. Kuepfer L, Sauer U, Blank LM: Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Research* 2005, 15(10):1421-1430.
16. Blank LM, Lehmbeck F, Sauer U: Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *Fems Yeast Research* 2005, 5(6-7):545-558.
17. Segre D, Vitkup D, Church G: Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the USA* 2002, 99(15112-15117).
18. Papp B, Pal C, Hurst LD: Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 2004, 429(6992):661-664.
19. Ibarra RU, Edwards JS, Palsson BO: Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 2002, 420(6912):186-189.
20. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO: Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004, 429(6987):92-96.
21. Fischer E, Sauer U: Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. *Nature Genetics* 2005, 37(6):636-640.
22. Fong SS, Nanchen A, Palsson BO, Sauer U: Latent pathway activation and increased pathway capacity enable Escherichia coli adaptation to loss of key metabolic enzymes. *Journal of Biological Chemistry* 2006, 281(12):8024-8033.
23. Van Nuland A, Vandormael P, Donaton M, Alenquer M, Lourenco A, Quintino E, Versele M, Thevelein JM: Ammonium permease-based sensing mechanism for rapid ammonium activation of the protein kinase A pathway in yeast. *Molecular Microbiology* 2006, 59(5):1485-1505.
24. Mouillon JM, Persson BL: New aspects on phosphate sensing and signalling in Saccharomyces cerevisiae. *Fems Yeast Research* 2006, 6(2):171-176.
25. Rolland F, Winderickx J, Thevelein JM: Glucose-sensing and -signalling mechanisms in yeast. *Fems Yeast Research* 2002, 2(2):183-201.
26. Dekel E, Alon U: Optimality and evolutionary tuning of the expression level of a protein. *Nature* 2005, 436(7050):588-592.
27. Shannon C: A mathematical theory of communication. . *Bell Systems Technical Journal* 1948, 27:379-423.
28. Li W-H: Molecular Evolution. Massachusetts: Sinauer; 1997.

29. Hartl DL, Clark AG: Principles of population genetics, 2nd edn. Sunderland, MA: Sinauer associates; 1989.
30. Drake JW, Charlesworth B, Charlesworth D, Crow JF: Rates of spontaneous mutation. *Genetics* 1998, 148(#4):1667-1686.
31. Johnson LJ, Koufopanou V, Goddard MR, Hetherington R, Schafer SM, Burt A: Population genetics of the wild yeast Saccharomyces paradoxus. *Genetics* 2004, 166(1):43-52.
32. Aa E, Townsend J, Adams R, Nielsen K, Taylor J: Population structure and gene evolution in Saccharomyces cerevisiae. *FEMS Yeast Research* 2006, 6:702-715.
33. Bray D: Protein molecules as computational elements in living cells. *Nature* 1995, 376(6538):307-312.
34. Sherman F: Getting started with yeast. *Guide To Yeast Genetics and Molecular and Cell Biology, Pt B* 2002, 350:3-41.
35. Berg HC: Random walks in biology. Princeton, NJ: Princeton University Press; 1983.
36. Arkowitz RA: Responding to attraction: chemotaxis and chemotropism in Dictyostelium and yeast. *Trends in Cell Biology* 1999, 9(1):20-27.
37. Neidhardt FC, Ingraham J, Schaechter M (eds.): Physiology of the bacterial cell. Sunderland, MA: Sinauer; 1990.
38. McAdams HH, Arkin A: It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics* 1999, 15(2):65-69.
39. Blake W, Kaern M, Cantor C, Collins J: Noise in eukaryotic gene expression. *Nature* 2003, 422(6932):633-637.
40. Elowitz M, Levine A, Siggia E, Swain P: Stochastic gene expression in a single cell. *Science* 2002, 297(5584):1183-1186.
41. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N: Noise in protein expression scales with natural protein abundance. *Nature Genetics* 2006, 38(6):636-643.
42. Becskei A, Serrano L: Engineering stability in gene networks by autoregulation. *Nature* 2000, 405:590-593.
43. Wagner A: Energy constraints on the evolution of gene expression. *Molecular Biology and Evolution* 2005, 22(6):1365-1374.
44. Tai SL, Boer VM, Daran-Lapujade P, Walsh MC, de Winde JH, Daran JM, Pronk JT: Two-dimensional transcriptome analysis in chemostat cultures - Combinatorial effects of oxygen availability and macronutrient limitation in Saccharomyces cerevisiae. *Journal of Biological Chemistry* 2005, 280(1):437-447.
45. Brown CJ, Todd KM, Rosenzweig RF: Multiple Duplications of Yeast Hexose-Transport Genes in Response to Selection in a Glucose-Limited Environment. *Molecular Biology and Evolution* 1998, 15(#8):931-942.
46. Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D: Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* 2002, 99(25):16144-16149.
47. Souza MAA, Tropia MJ, Brandao RL: New aspects of the glucose activation of the H+-ATPase in the yeast Saccharomyces cerevisiae. *Microbiology-Sgm* 2001, 147:2849-2855.

48. Akashi H, Gojobori T: Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Proceedings of the National Academy of Sciences of the USA* 2002, 99(3695-3700).
49. Price N, Reed J, Palsson B: Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology* 2004, 2:886-897.
50. Fraenkel D: Carbohydrate metabolism. In: *The molecular biology of the yeast Saccharomyces Metabolism and gene expression.* Edited by Strathern J, Jones E, Broach J. Cold Spring Harbor, New York: CSHL press; 1982: 1-37.

Figure Legends

Figure 1a: Increased glucose information (bits, lower horizontal axis, filled circles) and reduced sensing error (percent, upper horizontal axis, open symbols) cause an increases in fitness (1-*s*) as estimated through biomass production in the yeast metabolic network via flux balance analysis. The vertical axis shows the selection coefficient *s*, the difference to the maximal biomass yield at perfectly accurate information. Results are nearly identical for the other four substrates, and are shown for O_2 and NH_3 in the inset. The dashed horizontal line demarcates a neutral zone, below which ($s<7.33\times10^{-8}$; see text) growth rate increases are selectively neutral. **b)** Dependency of the selection coefficient *s* on glucose information and ammonium information. **c)** Effect of glucose information (bits) on the selection coefficient *s*, if sensing accuracy for a varying number of other substrates (differently shaded bars) varies uniformly in the interval (0,16) bits. At zero bits glucose information, the lack of glucose information limits growth, regardless of how much information is available on other growth substrates. **d)** The selection coefficient *s* depends nonlinearly on the total amount of information available for all five substrates. Information about each substrate was varied in the interval (0,16) bits. All data are for a minimal, aerobic medium with glucose as the sole carbon source and the following five substrates: glucose, O_2, phosphate, sulfate, and NH_3.

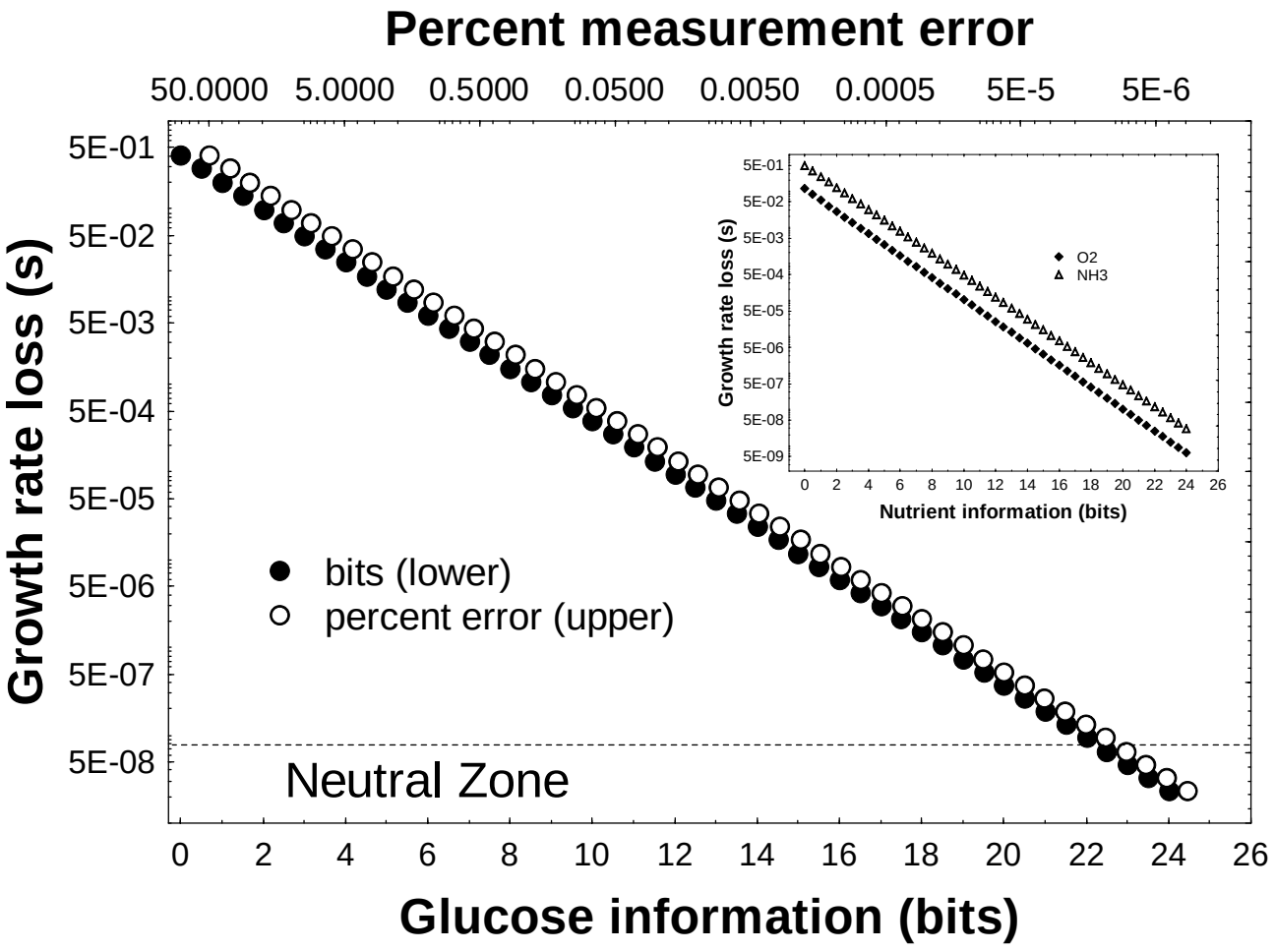
Percent measurement error
50.0000
5.0000
0.5000
0.0500
0.0050
0.0005
5E-5
5E-6
Growth rate loss (s)
5E-01
5E-02
5E-03
5E-04
5E-05
5E-06
5E-07
5E-08
bits (lower)
percent error (upper)
Neutral Zone
0
2
4
6
8
10
12
14
16
18
20
22
24
26
Glucose information (bits)
Growth rate loss (s)
5E-01
5E-02
5E-03
5E-04
5E-05
5E-06
5E-07
5E-08
5E-09
O2
NH3
0
2
4
6
8
10
12
14
16
18
20
22
24
26
Nutrient information (bits)

Figure 1A

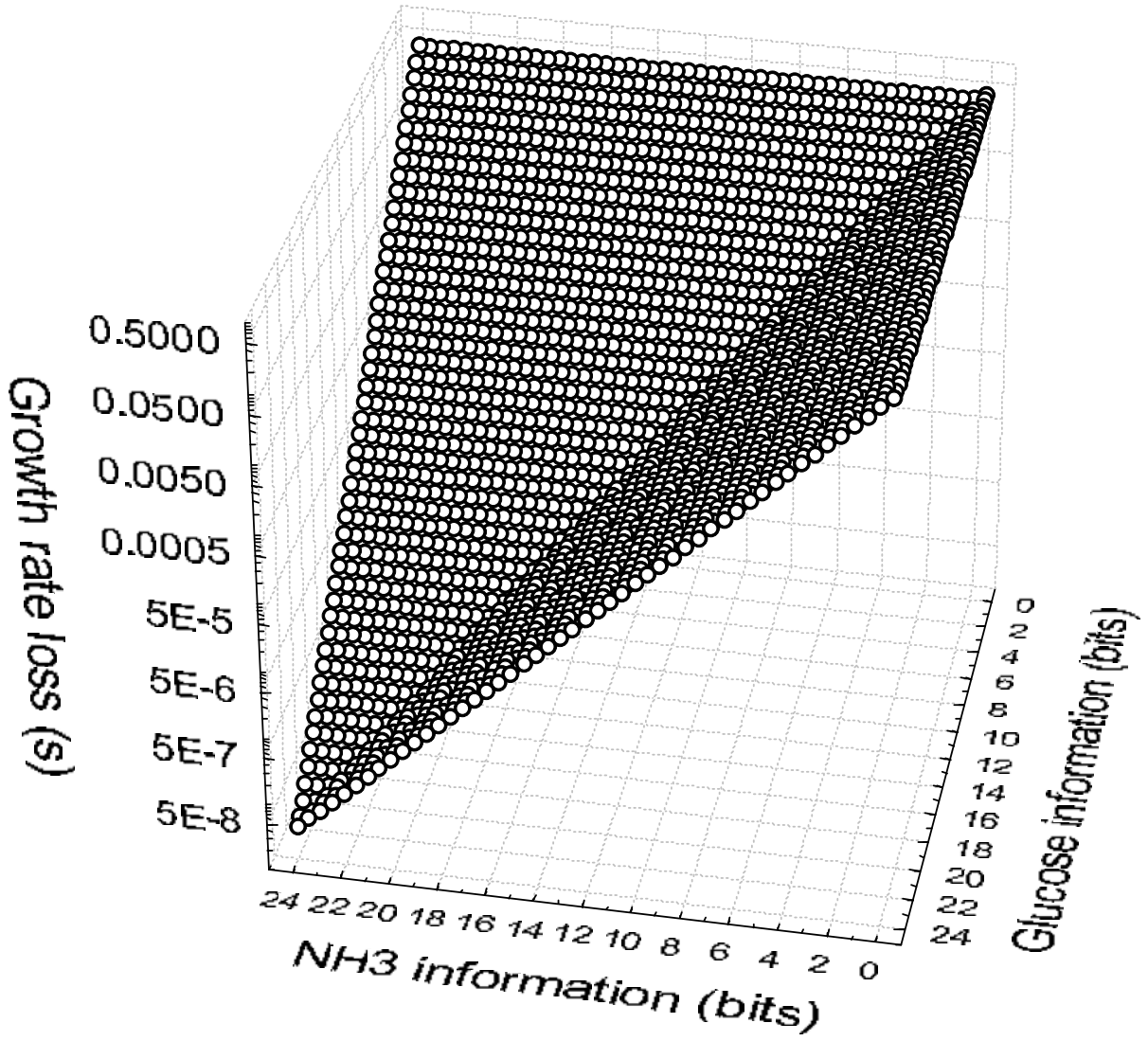
Growth rate loss (s)
0.5000
0.0500
0.0050
0.0005
5E-5
5E-6
5E-7
5E-8
NH3 information (bits)
24 22 20 18 16 14 12 10 8 6 4 2 0
Glucose information (bits)
0 2 4 6 8 10 12 14 16 18 20 22 24

Figure 1B

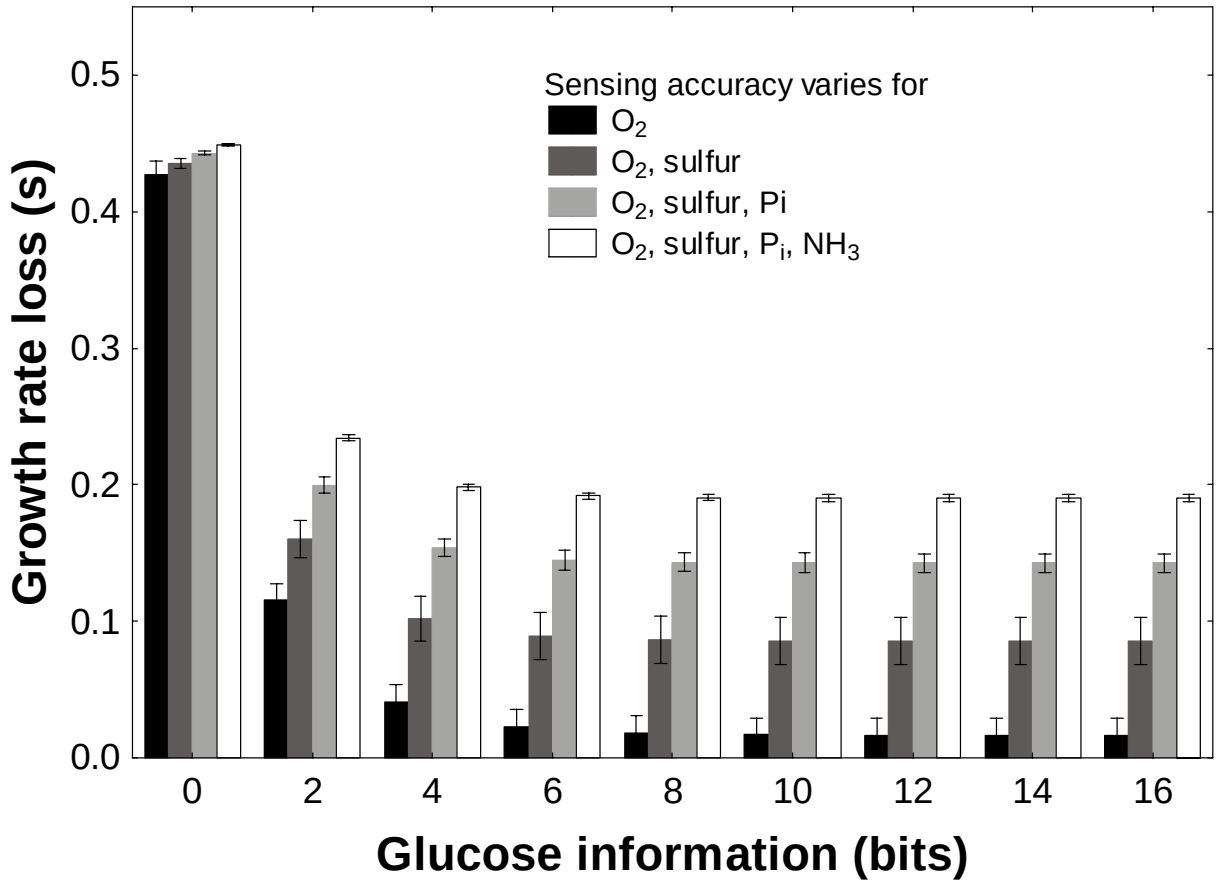
Sensing accuracy varies for
O_2
O_2, sulfur
O_2, sulfur, Pi
O_2, sulfur, P_i, NH_3
Growth rate loss (s)
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10
12
14
16
Glucose information (bits)

Figure 1C

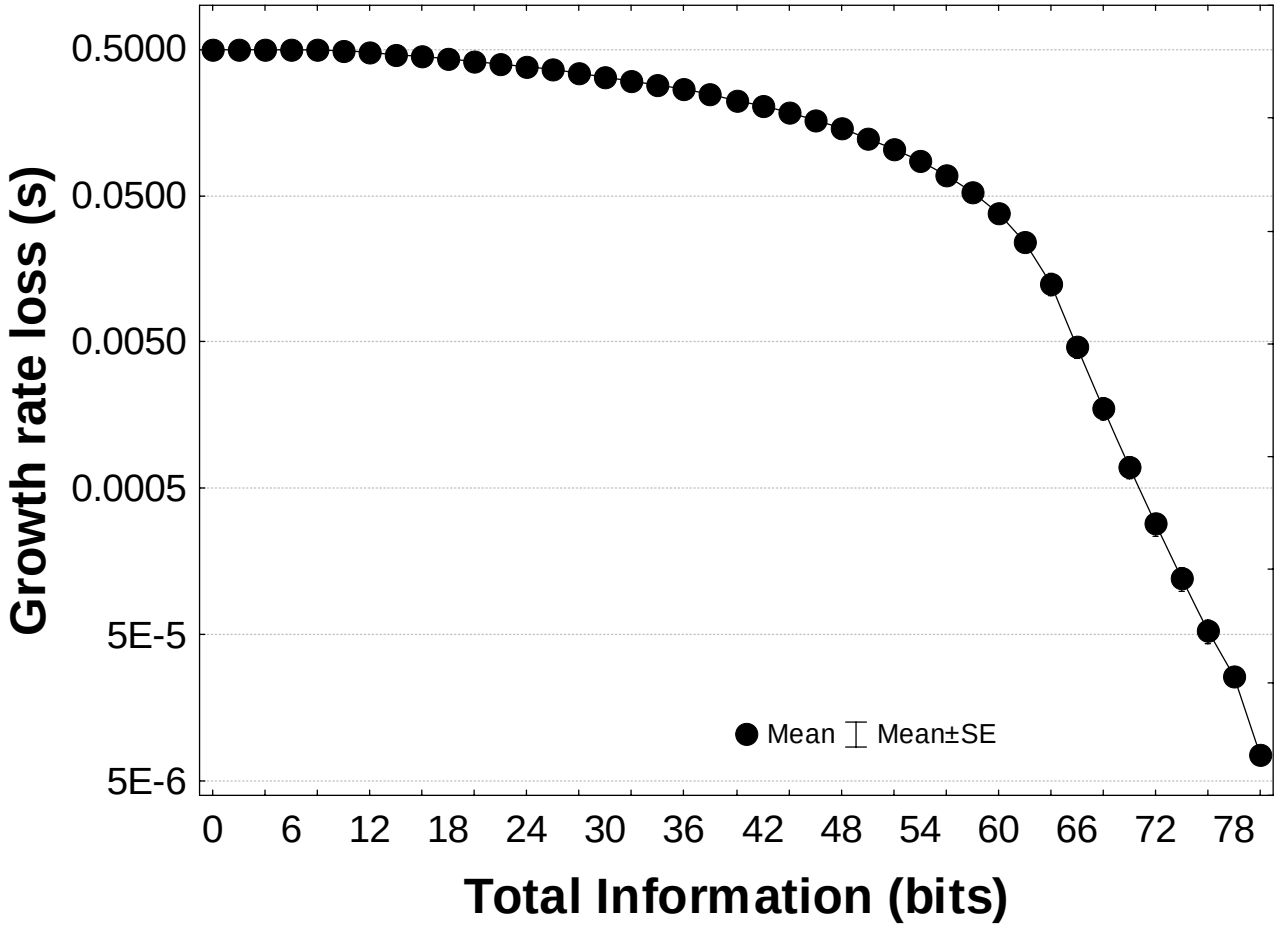
0.5000
0.0500
0.0050
0.0005
5E-5
5E-6
Growth rate loss (s)
0
6
12
18
24
30
36
42
48
54
60
66
72
78
Total Information (bits)
Mean
Mean±SE

Figure 1D