

Self-similarity of complex networks

Chaoming Song¹, Shlomo Havlin² & Hernán A. Makse¹

¹Levich Institute and Physics Department, City College of New York, New York, New York 10031, USA

²Minerva Center and Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel

Complex networks have been studied extensively owing to their relevance to many real systems such as the world-wide web, the Internet, energy landscapes and biological and social networks^{1–5}. A large number of real networks are referred to as ‘scale-free’ because they show a power-law distribution of the number of links per node^{1,6,7}. However, it is widely believed that complex networks are not invariant or self-similar under a length-scale transformation. This conclusion originates from the ‘small-world’ property of these networks, which implies that the number of nodes increases exponentially with the ‘diameter’ of the network^{8–11}, rather than the power-law relation expected for a self-similar structure. Here we analyse a variety of real complex networks and find that, on the contrary, they consist of self-repeating patterns on all length scales. This result is achieved by the application of a renormalization procedure that coarse-grains the system into boxes containing nodes within a given ‘size’. We identify a power-law relation between the number of boxes needed to cover the network and the size of the box, defining a finite self-similar exponent. These fundamental properties help to explain the scale-free nature of complex networks and suggest a common self-organization dynamics.

Two fundamental properties of real complex networks have attracted much attention recently: the small-world and the scale-free properties. Many naturally occurring networks are ‘small world’ because we can reach a given node from another one, following the path with the smallest number of links between the nodes, in a very small number of steps. This corresponds to the so-called ‘six degrees of separation’ in social networks¹⁰. It is mathematically expressed by the slow (logarithmic) increase of the average diameter of the network, $\bar{\ell}$, with the total number of nodes N , $\bar{\ell} \approx \ln N$, where ℓ is the shortest distance between two nodes and defines the distance metric in complex networks^{6,8,9,11}. Equivalently, we obtain:

$$N \approx e^{\bar{\ell}/\ell_0} \tag{1}$$

where ℓ_0 is a characteristic length.

A second fundamental property in the study of complex networks arises with the discovery that the probability distribution of the number of links per node, $P(k)$ (also known as the degree distribution), can be represented by a power-law (‘scale-free’) with a degree exponent γ that is usually in the range $2 < \gamma < 3$ (ref. 6):

$$P(k) \approx k^{-\gamma} \tag{2}$$

These discoveries have been confirmed in many empirical studies of diverse networks^{1–4,6,7}.

With the aim of providing a deeper understanding of the underlying mechanism that leads to these common features, we need to probe the patterns within the network structure in more detail. The question of connectivity between groups of interconnected nodes on different length scales has received less attention. But many examples exhibit the importance of collective behaviour, such as interactions between communities within social networks, links between clusters of websites of similar subjects, and the highly modular manner in which molecules interact to keep a cell alive. Here we show that real complex networks, such as the world-wide web (WWW), social, protein–protein interaction networks (PIN) and cellular networks are invariant or self-similar under a length-scale transformation.

This result comes as a surprise, because the exponential increase in equation (1) has led to the general understanding that complex networks are not self-similar, since self-similarity requires a power-law relation between N and ℓ .

How can we reconcile the exponential increase in equation (1) with self-similarity, or (in other words) an underlying length-scale-invariant topology? At the root of the self-similar properties that we unravel in this study is a scale-invariant renormalization procedure that we show to be valid for dissimilar complex networks.

To demonstrate this concept we first consider a self-similar

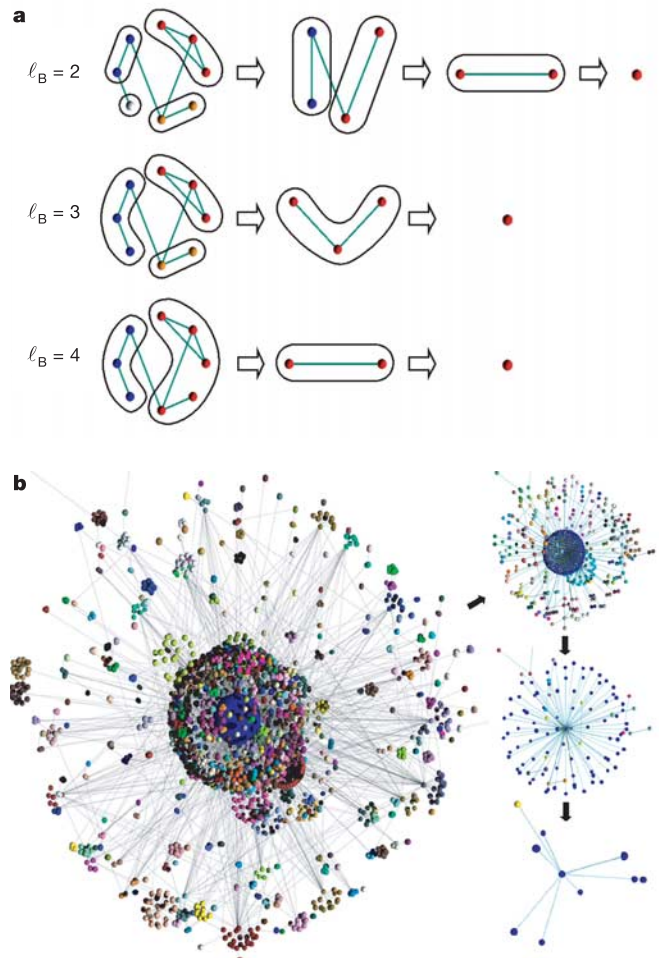


Figure 1 The renormalization procedure applied to complex networks. **a**, Demonstration of the method for different ℓ_B . The first column depicts the original network. We tile the system with boxes of size ℓ_B (different colours correspond to different boxes). All nodes in a box are connected by a minimum distance smaller than the given ℓ_B . For instance, in the case of $\ell_B = 2$, we identify four boxes that contain the nodes depicted with colours red, orange, white and blue, each containing 3, 2, 1 and 2 nodes, respectively. Then we replace each box by a single node; two renormalized nodes are connected if there is at least one link between the unrenormalized boxes. Thus we obtain the network shown in the second column. The resulting number of boxes needed to tile the network, $N_B(\ell_B)$, is plotted in Fig. 2 versus ℓ_B to obtain d_B as in equation (3). The renormalization procedure is applied again and repeated until the network is reduced to a single node (third and fourth columns for different ℓ_B). **b**, The stages in the renormalization scheme applied to the entire WWW. We fix the box size to $\ell_B = 3$ and apply the renormalization for four stages. This corresponds, for instance, to the sequence for the network demonstration depicted in the second row in panel **a**. We colour the nodes in the web according to the boxes to which they belong. The network is invariant under this renormalization, as explained in the legend of Fig. 2d and the Supplementary Information.

network embedded in euclidean space, of which a classical example would be a fractal percolation cluster at criticality¹². To unfold the self-similar properties of such clusters we calculate the fractal dimension using a ‘box-counting’ method and a ‘cluster-growing’ method¹³.

In the first method we cover the percolation cluster with N_B boxes of linear size ℓ_B . The fractal dimension or box dimension d_B is then given by¹⁴:

$$N_B \approx \ell_B^{-d_B} \quad (3)$$

In the second method, the network is not covered with boxes. Instead one seed node is chosen at random and a cluster of nodes centred at the seed and separated by a minimum distance ℓ is calculated. The procedure is then repeated by choosing many seed nodes at random and the average ‘mass’ of the resulting clusters ($\langle M_c \rangle$, defined as the number of nodes in the cluster) is calculated as a function of ℓ to obtain the following scaling:

$$\langle M_c \rangle \approx \ell^{d_f} \quad (4)$$

defining the fractal cluster dimension d_f ¹⁴. Comparing equations (4) and (1) implies that $d_f = \infty$ for complex small-world networks.

For a homogeneous network characterized by a narrow degree distribution (such as a fractal percolation cluster) the box-counting method of equation (3) and the cluster-growing method of equation (4) are equivalent, because every node typically has the

same number of links or neighbours. Equation (4) can then be derived from equation (3) and $d_B = d_f$, and this relation has been regularly used.

The crux of the matter is to understand how we can calculate a self-similar exponent (analogous to the fractal dimension in euclidean space) in complex inhomogeneous networks with a broad degree distribution such as equation (2). Under these conditions equation (3) and (4) are not equivalent, as will be shown below. The application of the proper covering procedure in the box-counting method (equation (3)) for complex networks unveils a set of self-similar properties such as a finite self-similar exponent and a new set of critical exponents for the scale-invariant topology.

Figure 1a illustrates the box-covering method using a schematic network composed of eight nodes. For each value of the box size ℓ_B , we search for the number of boxes needed to tile the entire network such that each box contains nodes separated by a distance $\ell < \ell_B$.

This procedure is applied to several different real networks: (1) a part of the WWW composed of 325,729 web pages that are connected if there is a URL link from one page to another⁶ (<http://www.nd.edu/~networks>); (2) a social network where the nodes are 392,340 actors linked if they were cast together in at least one film¹⁵; (3) the biological networks of protein–protein interactions found in *Escherichia coli* (429 proteins) and *Homo sapiens* (946 proteins) linked if there is a physical binding between them (database available via the Database of Interacting Proteins^{16,17}, other PINs are discussed in the Supplementary Information), and

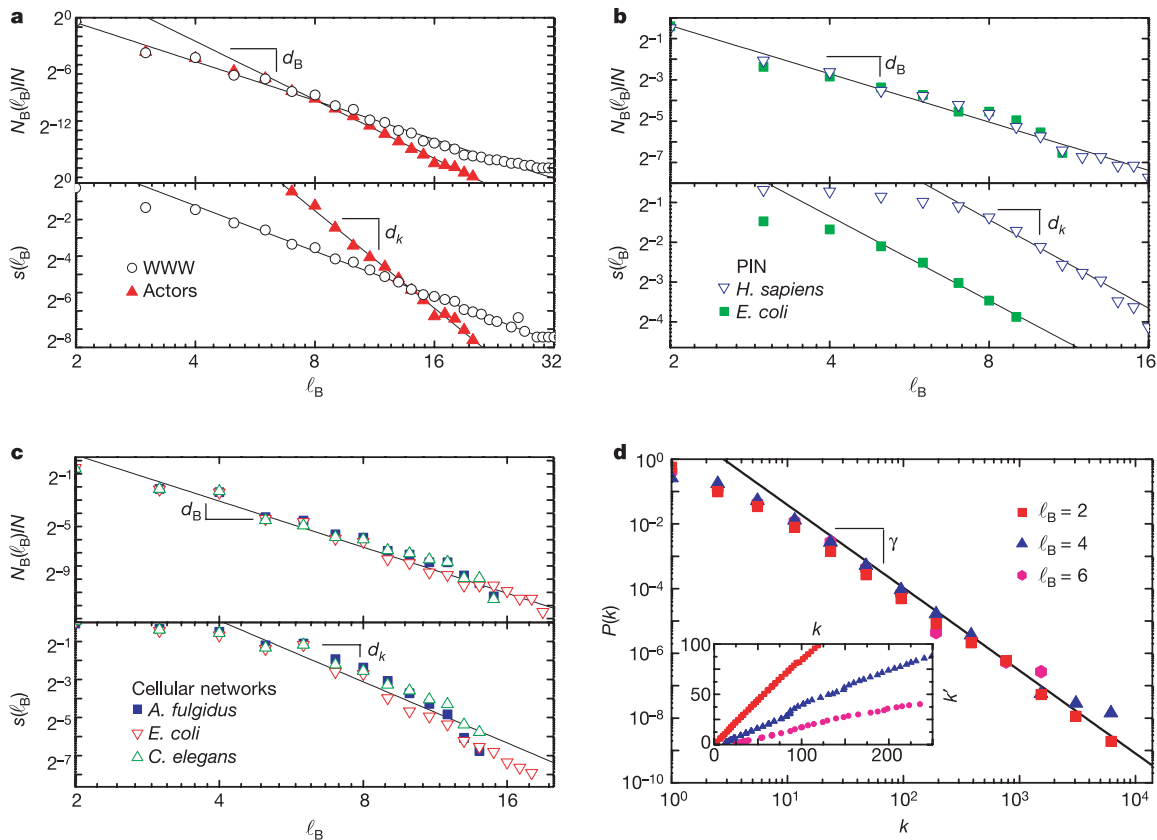


Figure 2 Self-similar scaling in complex networks. **a**, The upper panel shows a log-log plot of N_B versus ℓ_B , revealing the self-similarity of the WWW and actor network according to equation (3). The lower panel shows the scaling of $s(\ell_B)$ versus ℓ_B according to equation (9). The error bars are of the order of the symbol size. **b**, Same as **a** but for two PINs: *H. sapiens* and *E. coli*. Results are analogous to **b** but with different scaling exponents. **c**, Same as **a** for the cellular networks of *A. fulgidus*, *E. coli* and *C. elegans*. **d**, Invariance of the degree distribution of the WWW under the renormalization for different

box sizes, ℓ_B . We show the data collapse of the degree distributions, demonstrating the self-similarity at different scales. The inset shows the scaling of $K = s(\ell_B)k$ for different ℓ_B , whence we obtain the scaling factor $s(\ell_B)$. Moreover, we also apply the renormalization for a fixed box size, for instance $\ell_B = 3$ as shown in Fig. 1b for the WWW, until the network is reduced to a few nodes, and find that $P(k)$ is invariant under these multiple renormalizations as well, for several iterations (see Supplementary Information).

(4) the cellular networks compiled by ref. 18 using a graph-theoretical representation of all the biochemical pathways based on the WIT integrated-pathway genome database¹⁹ (<http://igweb.integratedgenomics.com/IGwit>) of 43 species from Archaea, Bacteria and Eukarya. Here we show the results for *Archaeoglobus fulgidus*, *E. coli* and *Caenorhabditis elegans*¹⁸; the full database is analysed in the Supplementary Information. It has been previously determined that the WWW and actor networks are small-world and scale-free, characterized by equation (2) with $\gamma = 2.6$ and 2.2, respectively¹. For the PINs of *E. coli* and *H. sapiens* we find $\gamma = 2.2$ and 2.1, respectively. All cellular networks are scale-free with average exponent $\gamma = 2.2$ (ref. 18). We confirm these values and show the results for the WWW in Fig. 2.

Figure 2a and b shows the results of $N_B(\ell_B)$ according to equation (3). They reveal the existence of self-similarity in the WWW, actors and *E. coli* and *H. sapiens* PINs with self-similar exponents $d_B = 4.1$, $d_B = 6.3$, and $d_B = 2.3$ and $d_B = 2.3$, respectively. The cellular networks are shown in Fig. 2c and have $d_B = 3.5$.

We now elaborate on the apparent contradiction between the two definitions of self-similar exponents in complex networks. After performing a renormalization at a given ℓ_B , we calculate the mean mass of the boxes covering the network, $\langle M_B(\ell_B) \rangle$, to obtain:

$$\langle M_B(\ell_B) \rangle \equiv N/N_B(\ell_B) \approx \ell_B^{d_B} \quad (5)$$

which is corroborated by direct measurements for all the networks, and shown in Fig. 3a for the WWW.

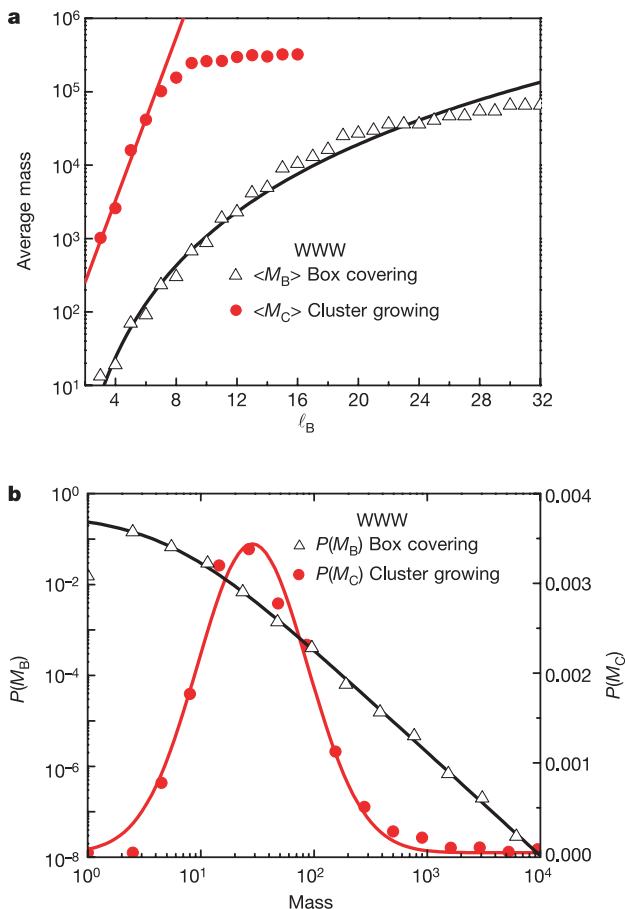


Figure 3 Different averaging techniques lead to qualitatively different results. **a**, Mean value of the box mass in the box-counting method, $\langle M_B \rangle$, and the cluster mass in the cluster growing method, $\langle M_C \rangle$, for the WWW. The solid lines represent the power-law fit for $\langle M_B \rangle$ and the exponential fit for $\langle M_C \rangle$ according to equations (5) and (6), respectively. **b**, Probability distribution of M_B and M_C for $\ell_B = 4$ for the WWW. The curves are fitted by a power-law and a log-normal distribution, respectively.

On the other hand, the average obtained from the cluster-growing method (for this calculation we average over single boxes without tiling the system) gives rise to an exponential growth of the mass:

$$\langle M_C(\ell_B) \rangle \approx \ell_B^{s/\ell_1} \quad (6)$$

with $\ell_1 \approx 0.78$ in accordance with the small-world effect equation (1), as seen in Fig. 3a.

The topology of scale-free networks is dominated by several highly connected hubs—the nodes with the largest degree—implying that most of the nodes are connected to the hubs via one or very few steps. Therefore the average obtained in the cluster-growing method is biased; the hubs are overrepresented in equation (6) because almost every node is a neighbour of a hub. By choosing the seed of the clusters at random, there is a very large probability of including the hubs in the clusters. On the other hand, the box-covering method is a global tiling of the system, providing a flat average over all the nodes: that is, each part of the network is covered with an equal probability. Once a hub (or any node) is covered, it cannot be covered again. We conclude that equations (3) and (4) are not equivalent for inhomogeneous networks with topologies dominated by hubs with a large degree.

The biased sampling of the randomly chosen nodes is clearly demonstrated in Fig. 3b. We find that the probability distribution of the mass of the boxes for a given ℓ_B is very broad and can be approximated by a power-law: $P(M_B) \approx M_B^{-2.2}$ in the case of the WWW and $\ell_B = 4$. On the other hand, the probability distribution of M_C is very narrow and can be fitted by a log-normal distribution (see Fig. 3b). In the box-covering method there are many boxes with very large and very small masses, in contrast to the peaked distribution in the cluster-growing method, thus showing the biased nature of the latter method in inhomogeneous networks. This biased average leads to the exponential growth of the mass in equation (6) and it also explains why the average distance is logarithmic with N , as in equation (1).

The box-counting method provides a powerful tool for further investigations of network properties because it enables a renormalization procedure, revealing that the self-similar properties and the scale-free degree distribution persist irrespectively of the amount of coarse-graining of the network.

Subsequent to the first step of assigning the nodes to the boxes we create a new renormalized network by replacing each box by a single node. Two boxes are then connected, provided that there was at least one link between their constituent nodes. The second column of the panels in Fig. 1a shows this step in the renormalization procedure for the schematic network, while Fig. 1b shows the results for the same procedure applied to the entire WWW for $\ell_B = 3$.

The renormalized network gives rise to a new probability distribution of links, $P(k')$, which is invariant under the renormalization:

$$P(k) \rightarrow P(k') \approx (k')^{-\gamma} \quad (7)$$

Figure 2d supports the validity of this scale transformation by showing a data collapse of all distributions with the same γ according to equation (7) for the WWW.

Further insight arises from relating the scale-invariant properties (equation (3)) to the scale-free degree distribution (equation (2)). Plotting (see inset in Fig. 2d for the WWW) the number of links k' of each node in the renormalized network versus the maximum number of links k in each box of the unrenormalized network exhibits a scaling law:

$$k \rightarrow k' = s(\ell_B)k \quad (8)$$

This equation defines the scaling transformation in the connectivity distribution. Empirically we find that the scaling factor s (< 1) scales with ℓ_B with a new exponent d_k :

$$s(\ell_B) \approx \ell_B^{-d_k} \quad (9)$$

as shown in Fig. 2a for the WWW and actor networks (with $d_k = 2.5$ and $d_k = 5.3$, respectively), in Fig. 2b for the protein networks ($d_k = 2.1$ for *E. coli* and $d_k = 2.2$ for *H. sapiens*) and in Fig. 2c for the cellular networks with $d_k = 3.2$.

Equations (8) and (9) shed light on how families of hierarchical sizes are linked together. The larger the families, the fewer links exist. Surprisingly, the same power-law relation exists for large and small families represented by equation (2).

From equation (7) we obtain $n(k)dk = n'(k')dk'$, where $n(k) = NP(k)$ is the number of nodes with links k and $n'(k') = N'P(k')$ is the number of nodes with links k' after the renormalization (N' is the total number of nodes in the renormalized network). Using equation (8), we obtain $n(k) = s^{1-\gamma}n'(k)$. Then, upon renormalizing a network with N total nodes we obtain a smaller number of nodes N' according to $N' = s^{\gamma-1}N$. The total number of nodes in the renormalized network is the number of boxes needed to cover the unrenormalized network at any given ℓ_B , so we have $N' = N_B(\ell_B)$. Then, from equations (3) and (9) we obtain the relation between the three indexes:

$$\gamma = 1 + d_B/d_k \quad (10)$$

Equation (10) is confirmed for all the networks analysed here (see Supplementary Information). In all cases the calculation of d_B and d_k and equation (10) gives rise to the same γ exponent as that obtained in the direct calculation of the degree distribution. The significance of this result is that the scale-free properties characterized by γ can be related to a more fundamental length-scale invariant property, characterized by the two new indexes d_B and d_k . □

Received 4 August; accepted 30 November 2004; doi:10.1038/nature03248.

1. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
2. Dorogovtsev, S. N. & Mendes, J. F. F. *Evolution of Networks: From Biological Nets to the Internet and the WWW* (Oxford Univ. Press, Oxford, 2003).
3. Pastor-Satorras, R. & Vespignani, A. *Evolution and Structure of the Internet: a Statistical Physics Approach* (Cambridge Univ. Press, Cambridge, 2004).
4. Newman, M. E. J. The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
5. Amaral, L. A. N. & Ottino, J. M. Complex networks—augmenting the framework for the study of complex systems. *Eur. Phys. J. B* **38**, 147–162 (2004).
6. Albert, R., Jeong, H. & Barabási, A.-L. Diameter of the World Wide Web. *Nature* **401**, 130–131 (1999).
7. Faloutsos, M., Faloutsos, P. & Faloutsos, C. On power-law relationships of the Internet topology. *Comput. Commun. Rev.* **29**, 251–262 (1999).
8. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–61 (1960).
9. Bollobás, B. *Random Graphs* (Academic, London, 1985).
10. Milgram, S. The small-world problem. *Psychol. Today* **2**, 60–67 (1967).
11. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
12. Bunde, A. & Havlin, S. *Fractals and Disordered Systems* Ch. 2 (eds Bunde, A. & Havlin, S.) 2nd edn (Springer, Heidelberg, 1996).
13. Vicsek, T. *Fractal Growth Phenomena* 2nd edn, Part IV (World Scientific, Singapore, 1992).
14. Feder, J. *Fractals* (Plenum, New York, 1988).
15. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
16. Xenarios, I. et al. DIP: the database of interacting proteins. *Nucleic Acids Res.* **28**, 289–291 (2000).
17. *Database of Interacting Proteins (DIP)* (<http://dip.doe-mbi.ucla.edu>) (2000).
18. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
19. Overbeek, R. et al. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acid Res.* **28**, 123–125 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to J. Bruijic for many discussions. This work is supported by the National Science Foundation, Materials Theory. S.H. thanks the Israel Science Foundation and ONR for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.A.M. (makse@mailaps.org).

Strong polarization enhancement in asymmetric three-component ferroelectric superlattices

Ho Nyung Lee, Hans M. Christen, Matthew F. Chisholm, Christopher M. Rouleau & Douglas H. Lowndes

Condensed Matter Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

Theoretical predictions—motivated by recent advances in epitaxial engineering—indicate a wealth of complex behaviour arising in superlattices of perovskite-type metal oxides. These include the enhancement of polarization by strain^{1,2} and the possibility of asymmetric properties in three-component superlattices³. Here we fabricate superlattices consisting of barium titanate (BaTiO₃), strontium titanate (SrTiO₃) and calcium titanate (CaTiO₃) with atomic-scale control by high-pressure pulsed laser deposition on conducting, atomically flat strontium ruthenate (SrRuO₃) layers. The strain in BaTiO₃ layers is fully maintained as long as the BaTiO₃ thickness does not exceed the combined thicknesses of the CaTiO₃ and SrTiO₃ layers. By preserving full strain and combining heterointerfacial couplings, we find an overall 50% enhancement of the superlattice global polarization with respect to similarly grown pure BaTiO₃, despite the fact that half the layers in the superlattice are nominally non-ferroelectric. We further show that even superlattices containing only single-unit-cell layers of BaTiO₃ in a paraelectric matrix remain ferroelectric. Our data reveal that the specific interface structure and local asymmetries play an unexpected role in the polarization enhancement.

Oxide heterostructures with atomically abrupt interfaces, defined by atomically flat surface terraces and single-unit-cell steps, can now be grown on well-prepared single-stepped substrates^{4–7}. This advance has encouraged theoretical investigations that have led to predictions of new artificial materials^{1–3,8–10}. The atomic-scale control of the combining of dissimilar materials is expected to produce striking property enhancements as well as new combinations of desired properties. Here we discuss the experimental realization of one of these predictions, the strain enhancement of ferroelectric polarization. The challenge associated with fabricating such strained structures—the deliberate and controlled deposition of up to hundreds of individual layers—remains a formidable task, for which the principal technique used has been high-vacuum molecular beam epitaxy^{5,11}. However, many insulators do not yield the correct oxide stoichiometry (or expected resulting physical properties) when grown by molecular beam epitaxy. Furthermore, a shortage of electrically conducting oxide substrates and our still-limited understanding of the stability and growth mechanisms of conducting-film electrodes have hindered the electrical characterization of oxide superlattices.

To address these challenges, we have recently shown that atomically flat, electrically conducting SrRuO₃ electrodes can be grown with a surface quality that mimics that of the substrate (Fig. 1a)⁷. Pulsed laser deposition (PLD) has long been regarded as an effective method for synthesizing various oxide heterostructures^{12–15}, but obtaining atomically sharp interfaces has been difficult in the comparatively high-pressure processes needed to maintain oxygen stoichiometry. Here we demonstrate the growth by a high-pressure PLD technique of hundreds of individual perovskite layers of BaTiO₃, SrTiO₃ and CaTiO₃. These superlattices were grown with layer-by-layer control, yielding as-grown samples with compositionally abrupt interfaces, atomically smooth surfaces, and excellent ferroelectric behaviour that indicated oxygen stoichiometry.