

# An atomic Boltzmann machine capable of on-chip learning

Brian Kiraly<sup>1,†</sup>, Elze J. Knol<sup>1,†</sup>, Werner M. J. van Weerdenburg<sup>1</sup>, Hilbert J. Kappen<sup>2</sup>, and Alexander A. Khajetoorians<sup>1,\*</sup>

<sup>1</sup>*Institute for Molecules and Materials, Radboud University, Nijmegen, the Netherlands*

<sup>2</sup>*Donders Institute for Neuroscience, Radboud University, Nijmegen, the Netherlands*

<sup>†</sup>These authors contributed equally.

\*corresponding author: [a.khajetoorians@science.ru.nl](mailto:a.khajetoorians@science.ru.nl)

**The Boltzmann Machine (BM) is a neural network composed of stochastically firing neurons that can learn complex probability distributions by adapting the synaptic interactions between the neurons<sup>1</sup>. BMs represent a very generic class of stochastic neural networks that can be used for data clustering, generative modelling and deep learning<sup>2</sup>. A key drawback of software-based stochastic neural networks is the required Monte Carlo sampling, which scales intractably with the number of neurons. Here, we realize a physical implementation of a BM directly in the stochastic spin dynamics of a gated ensemble of coupled cobalt atoms on the surface of semiconducting black phosphorus. Implementing the concept of orbital memory utilizing scanning tunnelling microscopy, we demonstrate the bottom-up construction of atomic ensembles whose stochastic current noise is defined by a reconfigurable multi-well energy landscape. Exploiting the anisotropic behaviour of black phosphorus, we build ensembles of atoms with two well-separated intrinsic time scales that represent neurons and synapses. By characterizing the conditional steady-state distribution of the neurons for given synaptic configurations, we illustrate that an ensemble can represent many distinct probability distributions. By probing the intrinsic synaptic dynamics, we reveal an autonomous reorganization of the synapses in response to external electrical stimuli. This self-adaptive architecture paves the way for on-chip learning directly in atomic-scale machine learning hardware.**

Boltzmann Machines<sup>1</sup> (BMs) have been leveraged in deep neural networks to perform dimensionality reduction utilizing both supervised and unsupervised learning approaches<sup>2,3</sup>. BMs are formally equivalent to an interacting Ising spin system with memory-bearing weighted interactions<sup>4</sup>. For many learning problems, the synaptic weights define a very complex, multi-modal energy landscape. To this end, the BM, as well as the Hopfield model, are strongly linked to the concept of spin glasses<sup>5-7</sup>. The BM learning problem is to adapt the synaptic weights such that the thermodynamic equilibrium distribution of the spins resembles a given data distribution as closely as possible. To compute the thermal equilibrium for a given BM is intractable, meaning that the computation time scales exponentially with the number of spins. Consequently, the thermal equilibrium is estimated through extensive Monte Carlo sampling, which is computationally expensive. It would therefore be advantageous to utilize spins in real materials<sup>8,9</sup> to emulate the BM, allowing the physics of the material to compute the equilibrium spin distribution. For such systems to be learnable, one needs materials that exhibit not only a degree of memory, but also an energy landscape that can be adapted in order to represent many possible spin distributions.

The design of neuromorphic architecture has been focused on implementing machine learning tasks utilizing electronic and/or magnetic phenomena directly in solid-state materials<sup>8,10-14</sup>. These devices can also mimic the biophysics of neurons and synapses using physical processes for use in energy-efficient brain-inspired computing<sup>15</sup>. To date, these approaches implement learning using hybrid schemes, combining the computational primitives of the materials and off-line computers to label data and implement training algorithms<sup>16,17</sup>. Yet, in addition to dealing with scalability concerns, one of the landmark goals of neuromorphic architecture is to create “on-chip” learning, eliminating the reliance on off-chip training schemes implemented via external computers. A fundamental shift toward this end requires the discovery and exploration of materials that autonomously adapt themselves on the basis of data, emphasizing the need for atomic-scale materials which are inherently reconfigurable in response to external stimuli and exhibit memory based on past experience.

Here, we demonstrate a Boltzmann machine capable of on-chip learning scaled to ensembles of coupled cobalt (Co) dopants on the surface of semiconducting black phosphorus (BP). Combining bi-stable atomic units, with gate-driven electric fields and substrate mediated interactions, we realize BM behaviour from configurations as small as three atoms. Taking a bottom-up approach, we start by representing stochastic, binary neurons by single Co atoms on BP, which exhibit bistable valencies and can be locally gated into a stochastic switching regime by the tip of a scanning tunnelling microscope (STM) (Fig. 1a)<sup>18,19</sup>. Utilizing atomic fabrication, we effectuate coupling between dopants to produce collective stochastic behaviour, which mimics the steady-state distributions resulting from a multi-well energy landscape for a BM. They further exhibit a rich and sensitive response to input electrical signals, introduced through a local gate. Strongly anisotropic substrate-driven interactions arising from the dielectric behaviour of BP<sup>18,19</sup> allow us to build atomic-scale synapses yielding memory-bearing, multi-valued distributions. Exploiting the separation of time-scales between neural and synaptic dynamics, we demonstrate a self-learning “on-chip” behaviour in which the synapses self-adapt their values in response to various input stimuli introduced by the external gating field. This result demonstrates, not only the scaling of a neuromorphic architecture to the atomic limit, but also provides a viable route for creating atomic-scale autonomous materials for on-chip learning.

The BM is a recurrent neural network where the fluctuating neurons are binary-valued ( $s = 0$  or  $1$ ), equivalent to stochastic Ising spins. BMs are closely related to finite temperature Ising spin models and derive their name from their equilibrium Boltzmann-Gibbs distribution. As in conventional feed-forward architectures, neurons are connected via synapses that are defined as continuously-valued, memory-bearing weights ( $w_{ij}$ ) with a bias (external field) for each individual neuron ( $b_i$ ). In the case of BM, the

couplings ( $w_{ij}$ ) are symmetric:  $w_{ij} = w_{ji}$ . Boltzmann machines belong to a class of neural networks called energy-based models, where the total energy of BM with a state defined by  $\mathbf{s} = (s_1, \dots, s_n)$  is defined as

$$E(\mathbf{s}|\mathbf{b}, \mathbf{w}) = -\sum_{i>j} w_{ij} s_i s_j - \sum_i b_i s_i \quad (1)$$

As the spins are continuously fluctuating, the energies ( $E(\mathbf{s}|\mathbf{b}, \mathbf{w})$ ) define a steady-state probability distribution for finding the system in a particular state  $\mathbf{s}$  according to the Boltzmann distribution:

$$\mathcal{P}(\mathbf{s}|\mathbf{b}, \mathbf{w}) = \frac{1}{Z(\mathbf{b}, \mathbf{w})} e^{-E(\mathbf{s}|\mathbf{b}, \mathbf{w})} \quad (2)$$

where  $Z(\mathbf{b}, \mathbf{w})$  is the partition function accounting for all possible states of the system. In this picture, learning is implemented by modifying or updating the weights  $w_{ij}$  and biases  $b_i$ . A material representation of the BM requires (i) stochastic neural elements  $s_i$ , and (ii) tunable and memory-bearing interactions  $w_{ij}$ ,  $b_i$  that are ideally self-adaptable to external stimuli.

The building blocks for all of the ensembles shown in this work are individual and identical Co atoms residing on the surface of semiconducting BP. It was recently demonstrated that a single Co atom exhibits orbital memory, based on the concept of its bistable valency<sup>20</sup>. We denote the two atomic states as  $s = 0$  and 1 (Fig. 1a,b), in analogy to the spins defined in the BM. The atomic valency is extremely sensitive to its environment: when probed with STM at applied voltages  $V_s < V_{th}$  (where  $V_{th}$  is the gate voltage threshold required to induce stochastic switching), it exhibits extended lifetimes which epitomize its utility as a memory. However, when  $V_s > V_{th}$  the atomic valency switches stochastically between its two states (seen in  $I_t(t)$  - Fig. 1e with the tip position marked “x” in the STM image in Fig. 1b) with a favourability governed by the local electrostatic environment. In this gated regime ( $V_s > V_{th}$ ), we identify the valency  $s = 0$  or 1 of an individual Co atom as a neuron in the BM representation (Fig. 1a). In order to correlate the state of  $s$  with a discrete and distinct current level measured in the stochastic noise, we reduce the gate voltage below  $V_{th}$  to freeze the given state before identifying it with constant-current STM

(Fig. 1b and Extended Data Figure 1). All subsequently displayed STM images were taken in this regime. To clearly distinguish different  $s$  configurations, we colour code each distinct level in  $I_t(t)$  (Fig. 1e). As the substrate is semiconducting, an applied potential between tip and substrate leads to a local voltage drop within the volume of the semiconductor underneath the tip (shaded area in Fig. 1a), implying the possibility of non-locally gating multiple atoms simultaneously (such as  $s_1$  and  $s_2$  in Fig. 1a).

A major step toward the realization of an atomic-scale BM was the effectuation of interatomic coupling facilitated by the controlled manipulation of Co atoms with the STM tip. When positioning two atoms with interatomic separation less than 3 nm along the BP  $x$  direction (armchair =  $x = [100]$ , zig-zag =  $y = [010]$ ) (Fig. 1c), we simultaneously gate both atoms with  $V_s > V_{th}$ , observing the emergence of four distinct and discrete states with appreciable lifetimes in current traces ( $I_t(t)$  - constant-height measurement shown in Fig. 1e with the tip position marked “x” in the STM image in Fig. 1c). Using constant-current STM to correlate  $I_t(t)$  with the possible state configurations, we map the four possible static  $s$  configurations onto each of these current levels (Fig. 1f colour coded in blue with  $s$  configurations shown in extended data Figure 1). Integrating over time, we can extract the steady-state probability distribution  $\mathcal{P}(s)$  (Fig. 1h) from the measured multi-state current ( $I_t(t)$ ). Coupled magnetic atoms at surfaces that exhibit stochastic switching have frequently been observed in the Ising limit<sup>21</sup>, where only two configurations show appreciable residence lifetimes, correlated to a favourable ferromagnetic or antiferromagnetic state<sup>22,23</sup>. By contrast,  $\mathcal{P}(s)$  shown in Fig. 1i exhibits non-zero populations of all four configurations, indicating that multiple configurations have appreciable lifetimes. This distribution illustrates the presence of multiple low-lying energy minima, suggesting a multi-well energy landscape of the form (1) with non-uniform couplings and biases. Moreover, the large difference of  $\mathcal{P}(s)$  compared to the probability distribution of an isolated atom confirms appreciable substrate-induced coupling between the atoms (see also extended data figure 2). This behaviour scales as we place a third atom in the  $x$

direction (Fig. 1d). Above the gate threshold, we observe 7 of 8 possible  $s$  configurations (Fig. 1g) and a more complex probability distribution (Fig. 1j). This illustration demonstrates that the number of minima increases with the system size.

While atoms coupled in the  $x$ -direction illustrate a complex probability distribution that we can associate with a BM distribution, the weights  $w_{ij}$  that define the probability distribution are fixed by the substrate-mediated couplings. In order to mimic learning, it is imperative to create memory-bearing and switchable synaptic weights  $w_{ij}$  interlinking the spins. To do this, we take advantage of the in-plane electronic anisotropy of BP<sup>18,19</sup>. As we show below, this anisotropy leads to a strong separation of time scales between atoms coupled in the  $x$  and  $y$  directions. In this way, we modify  $\mathcal{P}(s)$  of chains of Co atoms formed along the  $x$ -direction utilizing the memory of a coupled satellite Co atom separated in the  $y$ -direction (Fig. 2a, c-d). We define the orbital memory states of multiple satellite atoms by  $\mathbf{k}=(k_1, \dots, k_m)$  (Fig. 2a), quantifying their influence by examining the conditional probability distributions,  $\mathcal{P}(s|\mathbf{k})$ .

To illustrate this synaptic concept, we show a configuration where a single Co atom (labelled by  $k$ ) is placed approximately 2 nm in the  $y$ -direction from two coupled atoms,  $s_1$  and  $s_2$  (Fig. 2c). As seen in Fig. 2e ( $k=0$ ) and Fig. 2f ( $k=1$ ) for the two distinct states of the top Co atom, notable differences can be seen in the multi-state current ( $I_i(t)$ ) when the system is gated, clearly demonstrating that the  $k$  atom valency has a substantial impact on the steady-state conditional probability distributions  $\mathcal{P}(s|k=0)$  (Fig. 2g) and  $\mathcal{P}(s|k=1)$  (Fig. 2h). The distributions  $\mathcal{P}(s|k=0)$  and  $\mathcal{P}(s|k=1)$  can be modelled as distinct Boltzmann distributions (Eq. 1, 2) using the factor graph representation shown in Fig. 2b. We learn  $\mathbf{h}, \mathbf{w}$  for each value of  $k$  using the maximum likelihood approach, which is equivalent to minimizing the Kullback-Leibler divergence between the empirical probability distribution  $\mathcal{P}(s|\mathbf{k})$  of the states and the BM

distribution  $\mathcal{P}(s|\mathbf{b},\mathbf{w})$ . The results of this fitting for both  $k=0$  and  $k=1$  are shown in Table 1. The conditions  $k = 0$  or  $1$  result in significant modifications to the synaptic coupling ( $w_{12}$ ) and the biases ( $b_1, b_2$ ), directly influencing the BM energy landscape (1). Thus, the valency of the satellite Co atom acts as discrete parametrization for  $\mathbf{b}, \mathbf{w}$ . For different  $\mathbf{k}$ , a different distribution  $\mathcal{P}(s|\mathbf{k})$  is obtained, thus  $\mathbf{k}$  takes on the role of the parameters  $\mathbf{b}, \mathbf{w}$  of the BM. The lifetime of the  $\mathbf{k}$ -state is multiple orders of magnitude larger than the  $s$  states, as we detail later, providing the necessary memory element needed for synaptic weight. Therefore, this three-atom representation can be mapped to a BM (Fig. 2b, see also extended data figure 5) which contains two neurons and one binary synapse.

In order to be able to represent and learn a larger set of distributions, we introduce four  $k$  atoms coupled to three  $s$  atoms (Fig. 3a). With two states associated with each  $k$  atom, there are 16 possible  $\mathbf{k}$  configurations, so that we can represent 16 possible distributions  $\mathcal{P}(s|\mathbf{k})$  over the spins  $s$ . We show a subset of five possible  $\mathbf{k}$  configurations in Fig. 3. Figure 3a shows constant-current STM images, below the gate threshold, identifying the valency of each  $k$  atom for the five illustrated configurations. We employ the same characterization as in Fig. 2, measuring  $I_i(t)$  above the gate threshold (Fig. 3b), and extracting  $\mathcal{P}(s|\mathbf{k})$ , for each value of  $\mathbf{k}$ . These five distributions are plotted in Fig. 3c, illustrating highly varying changes in the distributions, depending on  $\mathbf{k}$ . Using the same modelling described above, for three spins with individual biases  $\mathbf{b}$ , pairwise couplings  $\mathbf{w}$ , and a three spin coupling  $w_{123}$ , we find different solutions  $\mathcal{P}(s|\mathbf{b}, \mathbf{w}, w_{123}) = \mathcal{P}(s|\mathbf{k})$  for different  $\mathbf{k}$ , indicating that different  $\mathbf{k}$  configurations correspond to different synapse and bias values and different state distributions. As seen in Table 2, the mapping of  $\mathbf{k}$  onto  $\mathbf{b}, \mathbf{w}, w_{123}$  is highly non-linear, likely related to the complex electrostatic environment of the atomic ensemble.

In Fig. 2 and 3, we showed that each  $\mathbf{k}$  configuration results in a different distribution  $\mathcal{P}(s|\mathbf{k})$  of the  $s$  configurations, and thus emulates synapses and biases of the BM, scaled to the atomic limit. Hence, changes in  $\mathbf{k}$  imply changes in the BM parameters and thus correspond to learning. Within conventional neuromorphic architecture, learning would be accomplished by implementing an external computer to modify the values of  $\mathbf{k}$  according to a predefined training algorithm. These values would then be fixed, unless they are modified externally. However, in biological systems, learning is a process identified with synaptic evolution, and it is an ongoing and constantly evolving process with an inherently different time scale compared to neural fluctuations. While neurons change their state on the order of milliseconds, learning, e.g. in the context of long-term potentiation, occurs over minutes to hours<sup>24</sup>.

The situation is similar in our atomic system. The  $\mathbf{k}$  configurations can dynamically change over a slow time scale under the influence of external stimuli applied electrically (Fig. 4c). The synaptic ( $\mathbf{k}$ ) energy landscape self-adapts in response to the stimuli, such that the various distributions  $\mathcal{P}(\mathbf{k})$  become environmentally dependent:  $\mathcal{P}(\mathbf{k}) \rightarrow \mathcal{P}(\mathbf{k}|\varepsilon)$ . As the time scales governing  $s$  and  $\mathbf{k}$  dynamics are separated by orders of magnitude, neural computation is equated with the  $s$  variables which exhibit fast dynamics, while learning is equated to the slower environmentally dependent evolution of the synaptic distribution ( $\mathcal{P}(\mathbf{k}|\varepsilon)$ ). Realizing both a separation of time scales and a self-adaptive/autonomous response provides a clear route toward on-chip learning<sup>25</sup>.

In order to demonstrate this on-chip learning concept, we studied the  $\mathbf{k}$  dynamics in a seven atom BM (shown experimentally in Fig. 4a and illustrated schematically in Fig 4b) over significantly longer time scales in order to allow for the spontaneous evolution of  $\mathbf{k}$ . We probed the response of  $\mathcal{P}(\mathbf{k}|\varepsilon)$  to changes

in the environmental stimuli in the form of small changes in an applied offset voltage ( $\varepsilon = V_{\text{off}} = V_s - V_{\text{th}}$ , where  $V_{\text{th}} \approx 400$  mV), in a regime where the total voltage is far above the stochastic switching threshold. We performed the measurements by initializing the ensemble into a random  $\mathbf{s}$  and  $\mathbf{k}$  configuration before measuring  $I_t(t)$  at a specified  $V_{\text{off}}$ . At each environmental condition ( $V_{\text{off}}$ ), measurements were conducted until the  $\mathcal{P}(\mathbf{k}|\varepsilon)$  converges to a representative steady-state distribution (Fig. 4d-f, see also extended data figure 7). Generally, the time scale for convergence of  $\mathcal{P}(\mathbf{k}|\varepsilon)$  for given  $\varepsilon$  is 1000-4000 times longer than for  $\mathcal{P}(\mathbf{s}|\mathbf{k})$  for given  $\mathbf{k}$ . As seen in Fig. 4d-f,  $\mathcal{P}(\mathbf{k}|\varepsilon)$  reaches a steady-state that is extremely sensitive to changes in the stimulus ( $V_{\text{off}}$ ) on the order of 10 mV. When  $V_{\text{off}} = 200$  mV,  $\mathbf{k} = (0000)$  is the most favourable synaptic configuration ( $\mathcal{P}(\mathbf{k}=0000|\varepsilon_3) = 0.43$ ), while at  $V_{\text{off}} = 160$  mV the favourability moves to  $\mathbf{k} = (0100)$  ( $\mathcal{P}(0100|\varepsilon_1) = 0.38$ ). In other words,  $\mathcal{P}(\mathbf{k}|\varepsilon)$  autonomously adapts in response to small variations in the gate voltage. In order to demonstrate that this response is not random, we did the following control experiment; we measured at (i)  $\varepsilon_1$ , (ii)  $\varepsilon_3$ , and again at (i)  $\varepsilon_1$ , while checking the conditional probability  $\mathcal{P}(\mathbf{k}|\varepsilon)$ . The measurements confirm that the  $\mathcal{P}(\mathbf{k}|\varepsilon_1)$  distribution remains the same before and after the intermediary stimulus  $\varepsilon_3$ . Furthermore, it is clear that the evolution of  $\mathcal{P}(\mathbf{k}|\varepsilon)$  is non-linear in  $V_{\text{off}}$ ; this is nicely exemplified when examining the  $\varepsilon$ -dependent probability for  $\mathbf{k} = (1000)$ :  $\mathcal{P}(\mathbf{k}=1000|\varepsilon_1) = 0.09$ ,  $\mathcal{P}(\mathbf{k}=1000|\varepsilon_2) = 0.33$ ,  $\mathcal{P}(\mathbf{k}=1000|\varepsilon_3) = 0.20$ , which is maximum at  $V_{\text{off}} = 180$  mV ( $\varepsilon_2$ ). Visualizing such non-linearity is aided by considering the additional complexity in the joint probability distribution  $\mathcal{P}(\mathbf{s}, \mathbf{k}|\varepsilon) = \mathcal{P}(\mathbf{s}|\mathbf{k}, \varepsilon)\mathcal{P}(\mathbf{k}|\varepsilon)$  (shown in blue in Fig. 4d-f). What is clearly evidenced in Fig. 4d-f, is that distinct steady-state  $\mathcal{P}(\mathbf{k}|\varepsilon)$  distributions are correlated with distinct environmental stimuli (also seen in extended data figure 6). In other words, the  $\mathbf{k}$  configurations self-adapt over a longer time scale, conditioning their state on the input stimulus according to a multi-modal energy landscape defined by the stimuli, analogous to the landscapes of the spins.

Having demonstrated the impact of both AC stimuli on  $\mathcal{P}(s)$  for a single atom, and DC stimuli on  $\mathcal{P}(s)$  for single and coupled atoms in extended data Fig. 3 and 4, we suggest a strategy for utilizing these Co atom ensembles on BP to achieve generic neural computation scaled to high dimensional problems. The basic concept is to encode high dimensional patterns such as images and their class labels into the spectral components of the input AC signal. Each image with its class label would be encoded in the state  $s$  of the BM. By repeated presentation of training patterns, the BM learns the image classifications, because the  $k$  configurations would converge to a value such that  $\mathcal{P}(s|k)$  encodes the statistical relation between the  $s$  representation of the images.

In conclusion, we have constructed an atomic-scale circuit which emulates a Boltzmann machine directly on the surface of the semiconductor black phosphorus. When controllably coupling individual Co atoms, we observed the onset of multi-well behaviour upon gating which is equivalent to the complex energy landscape observed in recurrent neural networks. Utilizing the anisotropic behaviour of black phosphorus, we were able to create atomic-scale synapses in which the synaptic weights were stored in a network of orbital memories of individual cobalt atoms. We have further shown that a separation of neural and synaptic time scales larger than three orders of magnitude naturally arises; this allowed us to confirm that the synaptic, memory-bearing atoms autonomously reorganize in response to an input stimulus. In addition to this self-adaption, we showed that Co atoms on BP exhibit rich and complex phase spaces in response to both DC and AC voltage signals, which could allow the network to solve higher dimensional problems. Using scanning tunnelling microscopy to expand the ensembles to 10's of atoms could significantly increase the number of encodable distributions, which scales with  $2^k$ . These ensembles represent the ultimate scaling limit in a physical system, comprised of single atoms with a total footprint of less than  $50 \text{ nm}^2$ . It remains an open question how the Boltzmann behaviour of these cobalt ensembles relates to the physical properties of the system, such as the underlying substrate-mediated interaction

mechanism and the dielectric response of the black phosphorus. Likewise, the concept of orbital memory is not necessarily limited to the surface and the methodology used here, suggesting that the concepts shown here can be extended to three-dimensional ensembles. Finally, as each cobalt atom has an unexplored spin degree of freedom<sup>26</sup>, these model systems provide a rich platform toward creating on-chip learning algorithms directly in hardware.

## Methods

### *Scanning Tunnelling Microscopy*

STM measurements were performed under ultrahigh vacuum ( $< 1 \times 10^{-10}$  mbar) conditions with an Omicron low-temperature STM at a base temperature of 4.4 K. The DC bias was provided by the STM control unit (SPECS Nanonis RC4/SC4) and applied to the sample. Additionally, for the AC experiments, a sinusoidal or triangular waveform was provided by an arbitrary waveform generator (Keysight 33622A) and added to the bias line with an active adder. The typical time resolution for data acquisition was approximately 1 ms. All STM images were acquired by means of constant-current feedback. All  $I_t(t)$  measurements were acquired with the tip at a constant height and the feedback loop turned off. To ensure that the tip height was equal in different experiments, the tip was stabilized with constant-current feedback on the bare black phosphorus, at  $I_t = 20$  pA,  $V_s = -400$  mV. For the AC experiments, the tip was stabilized on the centre of the Co atom, as defined by the maximum apparent height of state  $s = 0$  found using atom tracking. Sequentially, the tip position was offset by 1 nm with respect to the atom centre, as marked in the constant-current images. Electrochemically etched W tips were used for the measurements; each tip was treated in situ by electron bombardment, field emission, as well as dipped and characterized on a clean Au(111) surface. Black phosphorus crystals were purchased from HQ graphene and subsequently stored in vacuum ( $< 1 \times 10^{-8}$  mbar). The crystals were cleaved under ultrahigh vacuum conditions at pressures below  $2 \times 10^{-10}$  mbar, and immediately transferred to the microscope for in-situ

characterization. Cobalt was evaporated directly into the STM chamber with  $T_{\text{STM}} < 5$  K for the entire duration of the dosing procedure. Atomic manipulation of the cobalt atoms was done by dragging the atoms in constant-current feedback mode with  $-150 \text{ mV} < V_s < -80 \text{ mV}$  and  $1 \text{ nA} < I_t < 20 \text{ nA}$ .

### *Computing probability distributions*

To identify each  $s$  switching event in the  $I_t(t)$  data, an algorithm was used that determines the locations of abrupt changes (steps) in data. The algorithm split the data into segments for which the difference between the residual error and the mean of the data in the segment is minimal<sup>27,28</sup>. Where necessary, the data was pre-processed by smoothing and corrected for the z-drift between tip and sample. After identifying all the switches in the data, the (local) mean of all the segments ( $I_{\text{mean},i}$ ) was calculated between switching events. Discrete maxima in the histogram of the mean and/or raw current signal were used to identify a target current ( $I_{s,i}$ ) for each  $s$  configuration; individual segments in  $I_t(t)$  were assigned to  $s$  configurations based on the smallest value for  $|I_{\text{mean},i} - I_{s,i}|$ . To compute  $\mathcal{P}(s)$ , the histogram of the discretized data was computed and the total values for each  $s$  configuration were normalized to the total length of the measurement. To acquire distributions  $\mathcal{P}(k)$ , each  $k$  switch was identified manually. For both  $\mathcal{P}(s)$  and  $\mathcal{P}(k)$ , all data is shown after acquiring at minimum two times longer than the convergence time. The error bars of the distributions  $\mathcal{P}(s)$  and  $\mathcal{P}(k)$  are negligible and were therefore not plotted.

### *Boltzmann Machine Modelling*

The Boltzmann machine is a stochastic neural network of binary neurons. Each neuron  $i$  takes binary values  $s_i = 1, 0$  that denote firing and non-firing, respectively. The neural dynamics is given by sequential Glauber dynamics: At each time step, a neuron  $i$  is selected at random and its state is set to the new value  $s_i'$  with probability  $p(s_i' | \mathbf{s}) = \sigma(s_i' h_i(\mathbf{s}))$  with  $\mathbf{s} = (s_1, \dots, s_n)$  the current state of the network.

$$h_i(\mathbf{s}) = \sum_{j=1}^n w_{ij} s_j + b_i$$

is the summed input of all other neurons on neuron  $i$  with  $w_{ij}$  the synaptic coupling from neuron  $j$  to neuron  $i$  and  $-b_i$  is a threshold value.  $\sigma$  is the sigmoid non-linearity  $\sigma(x) = (1 + e^{-2x})^{-1}$ .

For finite values of the couplings and thresholds, the Glauber dynamics defines an ergodic Markov process that has a unique stationary distribution<sup>29</sup>. When the coupling matrix  $w$  is symmetric ( $w_{ij} = w_{ji}$ ) the Glauber dynamics satisfies detailed balance and the stationary distribution is the Boltzmann distribution Eq. 2 with energy Eq. 1. Learning the Boltzmann machine is defined as a procedure to find the parameters  $w, b$  such that the Boltzmann distribution Eq. 2 is as close as possible to a given distribution  $q$ .  $q$  is typically given in terms of a data set of  $N$  samples  $s^\mu, \mu = 1, \dots, N$ , where each sample  $s^\mu$  is a binary vector of length  $n$ . The data set defines the empirical probability distribution

$$q(\mathbf{s}) = \frac{1}{N} \sum_{\mu=1}^N \delta_{\mathbf{s}, s^\mu}.$$

In the simplest case with no hidden units, the 'distance' between  $p$  and  $q$  is defined as the relative entropy or Kullback-Leibler divergence

$$KL(q, p) = \sum_{\mathbf{s}} q(\mathbf{s}) \log \frac{q(\mathbf{s})}{p(\mathbf{s})}.$$

Minimizing  $KL(q, p)$  with respect to  $w, b$  is equivalent to maximizing the log likelihood

$$L = \sum_{\mathbf{s}} q(\mathbf{s}) \log p(\mathbf{s}).$$

The maximization can be performed by gradient ascent on  $L$ :

$$\Delta w_{ij} \propto \frac{\partial L}{\partial w_{ij}} = \langle s_i s_j \rangle_q - \langle s_i s_j \rangle_p$$

$$\Delta b_i \propto \frac{\partial L}{\partial b_i} = \langle s_i \rangle_q - \langle s_i \rangle_p$$

where  $\langle \dots \rangle_{p,q}$  is expectation with respect to the Boltzmann distribution  $p$  and the empirical distribution  $q$ , respectively. Learning stops when the gradients are zero, i.e. when the statistics under  $p$  and  $q$  are equal. This is the well-known Boltzmann Machine learning rule<sup>30</sup>.

## References

- 1 Hinton, G. & Sejnowski, T. Optimal perceptual inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 448-453 (1983).
- 2 Hinton, G. E. & Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science* **313**, 504 (2006).
- 3 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444 (2015).
- 4 Hertz, J., Krogh, A. & Palmer, R. G. *Introduction to the theory of neural computation*. (1991).
- 5 Edwards, S. F. & Anderson, P. W. Theory of spin glasses. II. *Journal of Physics F: Metal Physics* **6**, 1927 (1976).
- 6 Edwards, S. F. & Anderson, P. W. Theory of spin glasses. *Journal of Physics F: Metal Physics* **5**, 965 (1975).
- 7 Sherrington, D. & Kirkpatrick, S. Solvable Model of a Spin-Glass. *Physical Review Letters* **35**, 1792 (1975).
- 8 Grollier, J., Querlioz, D. & Stiles, M. D. Spintronic Nanodevices for Bioinspired Computing. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers* **104**, 2024-2039 (2016).
- 9 Kolmus, A., Katsnelson, M. I., Khajetoorians, A. A. & Kappen, H. J. Atom-by-atom construction of attractors in a tunable finite size spin array. *New Journal of Physics* **22**, 023038 (2020).
- 10 Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. The missing memristor found. *Nature* **453**, 80-83 (2008).
- 11 Romera, M. *et al.* Vowel recognition with four coupled spin-torque nano-oscillators. *Nature* **563**, 230-234 (2018).
- 12 Torrejon, J. *et al.* Neuromorphic computing with nanoscale spintronic oscillators. *Nature* **547**, 428-431 (2017).
- 13 Chen, T. *et al.* Classification with a disordered dopant-atom network in silicon. *Nature* **577**, 341-345 (2020).
- 14 Bose, S. K. *et al.* Evolution of a designless nanoparticle network into reconfigurable Boolean logic. *Nature Nanotechnology* **10**, 1048-1052 (2015).
- 15 Indiveri, G. & Liu, S. Memory and Information Processing in Neuromorphic Systems. *Proceedings of the IEEE* **103**, 1379-1397 (2015).
- 16 Prezioso, M. *et al.* Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61-64 (2015).
- 17 Burr, G. W. *et al.* Neuromorphic computing using non-volatile memory. *Advances in Physics: X* **2**, 89-124 (2017).
- 18 Kiraly, B. *et al.* Anisotropic Two-Dimensional Screening at the Surface of Black Phosphorus. *Physical Review Letters* **123**, 216403 (2019).
- 19 Prishchenko, D. A., Mazurenko, V. G., Katsnelson, M. I. & Rudenko, A. N. Coulomb interactions and screening effects in few-layer black phosphorus: a tight-binding consideration beyond the long-wavelength limit. *2D Materials* **4**, 025064 (2017).
- 20 Kiraly, B. *et al.* An orbitally derived single-atom magnetic memory. *Nature Communications* **9**, 3904 (2018).
- 21 Khajetoorians, A. A. *et al.* Atom-by-atom engineering and magnetometry of tailored nanomagnets. *Nature Physics* **8**, 497-503 (2012).
- 22 Khajetoorians, A. A. *et al.* Current-Driven Spin Dynamics of Artificially Constructed Quantum Magnets. *Science* **339**, 55-59 (2013).
- 23 Loth, S., Baumann, S., Lutz, C. P., Eigler, D. M. & Heinrich, A. J. Bistability in Atomic-Scale Antiferromagnets. *Science* **335**, 196 (2012).

- 24 Purves, D. *et al. Neuroscience*. (2019).
- 25 Heskes, T. M. & Kappen, B. Learning processes in neural networks. *Physical Review A* **44**, 2718-2726 (1991).
- 26 Badrtdinov, D. I., Rudenko, A. N., Katsnelson, M. I. & Mazurenko, V. V. Control of magnetic interactions between surface adatoms via orbital repopulation. *arXiv e-prints*, arXiv:2002.09354 (2020).
- 27 Lavielle, M. Using penalized contrasts for the change-point problem. *Signal Processing* **85**, 1501-1510 (2005).
- 28 Killick, R., Fearnhead, P. & Eckley, I. A. Optimal Detection of Changepoints With a Linear Computational Cost. *Journal of the American Statistical Association* **107**, 1590-1598 (2012).
- 29 Kappen, H. J. in *Handbook of Biological Physics* Vol. 4 (eds F. Moss & S. Gielen) 517-552 (North-Holland, 2001).
- 30 Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. A Learning Algorithm for Boltzmann Machines\*. *Cognitive Science* **9**, 147-169 (1985).

## Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 818399). This research was funded in part by ONR Grant N00014-17-1-2569. A.A.K. and E.J.K. acknowledge the NWO-VIDI project "Manipulating the interplay between superconductivity and chiral magnetism at the single-atom level" with project number 680-47-534. B.K. acknowledges NWO-VENI project "Controlling magnetism of single atoms on black phosphorus" with project number 016.Veni.192.168.

## Author Contributions

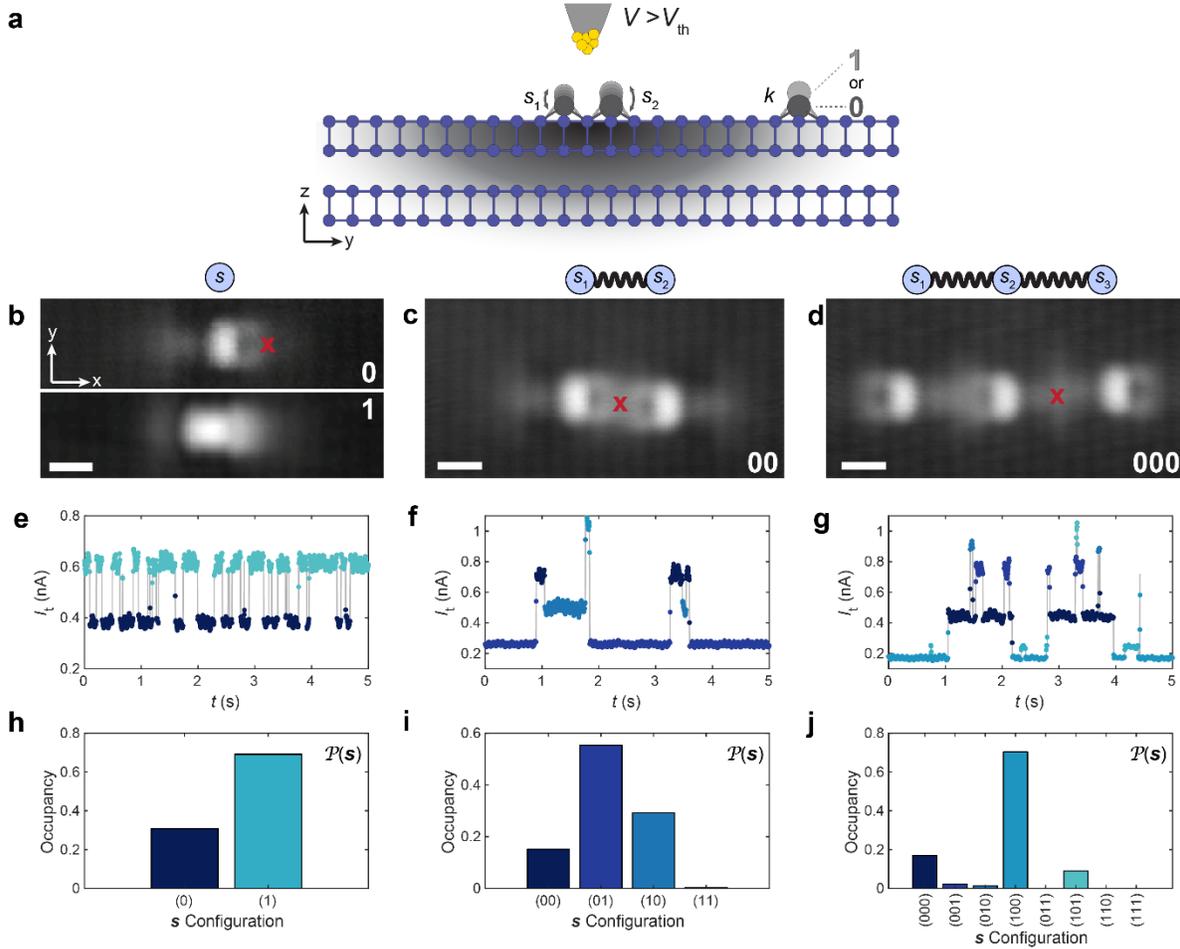
B.K. and E.J.K. performed the experiments under the direction and supervision of A.A.K. B.K. and E.J.K. developed the data analysis, while B.K., E.J.K., H.J.K. and A.A.K. participated in the scientific analysis. W.M.J.v.W. developed the AC experimental setup. H.J.K. performed the Boltzmann machine modelling. A.A.K. and H.J.K. designed the experiments. The manuscript was written by all B.K., E.J.K., H.J.K. and A.A.K.

## Competing Interests

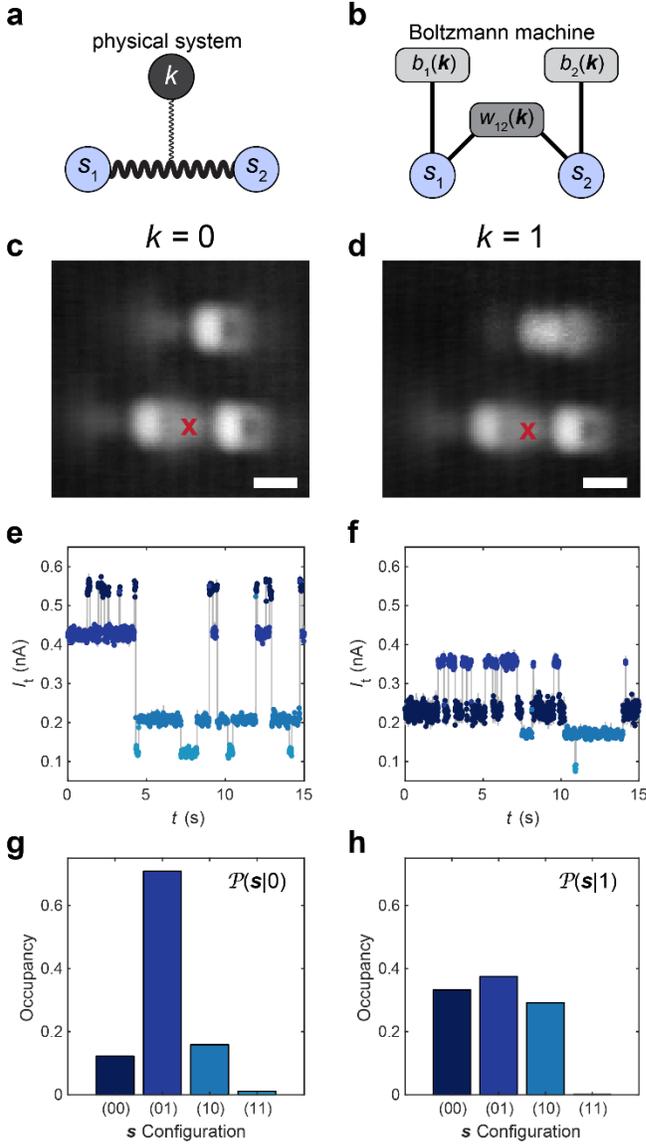
The authors declare no competing interests.

## Materials and Correspondence

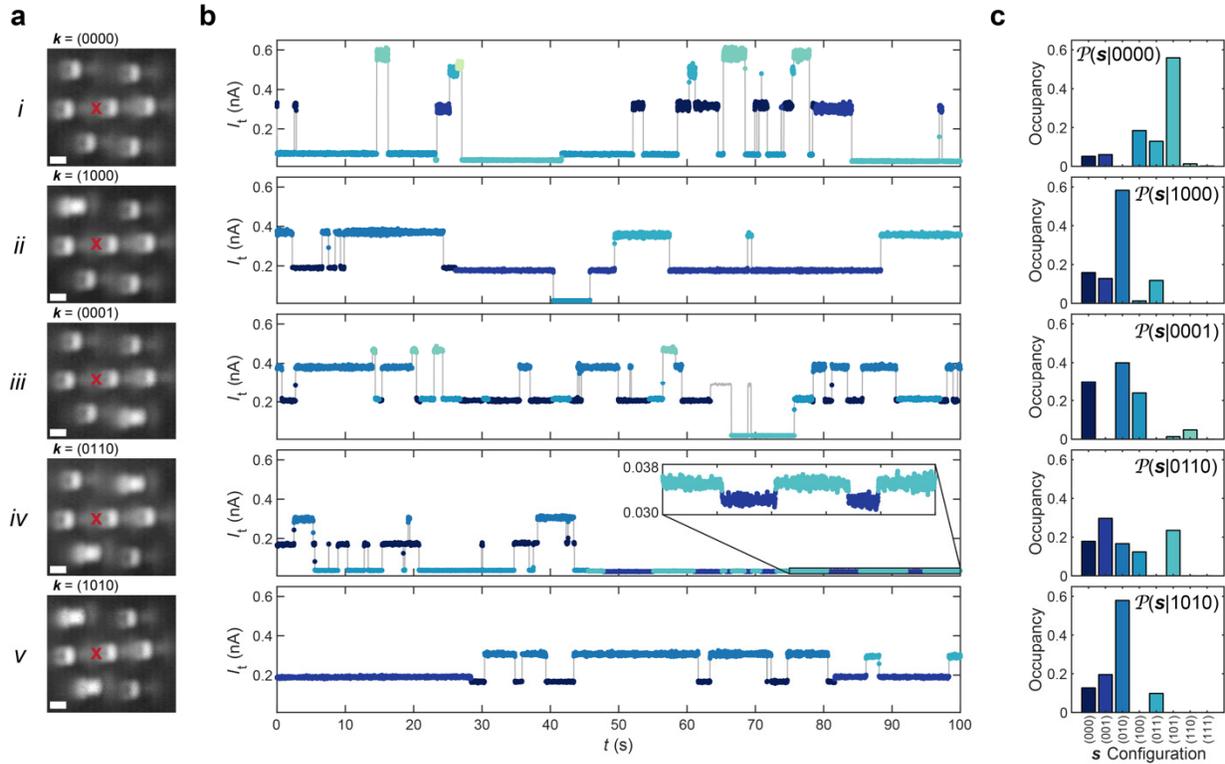
Correspondence to Alexander Khajetoorians.



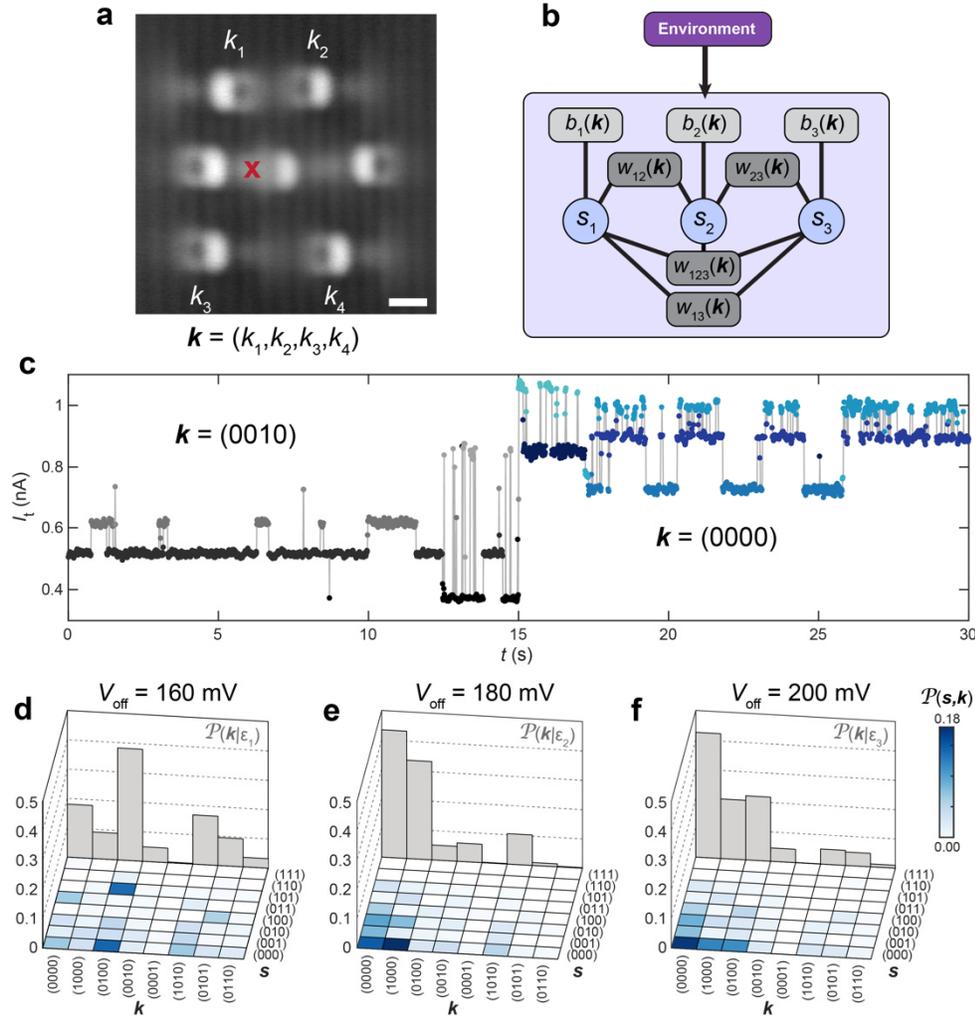
**Figure 1. Neural dynamics from coupled cobalt atoms on black phosphorus.** **a**, Schematic illustrating experimental setup with STM tip gating the system above the switching threshold. Here, due to differences in local band bending (grey)  $s_1$  and  $s_2$  atoms switch stochastically, while  $k$  remains static. **b**, Isolated cobalt atom in two valency configurations ( $s = 0$  and  $1$ ) with armchair =  $x = [100]$  and zig-zag =  $y = [010]$  crystallographic directions defined ( $V_s = -400$  mV, scale bar = 1 nm). **c**, Two cobalt atoms in state  $(s_1, s_2) = (0,0)$  and **d**, three cobalt atoms in state  $(s_1, s_2, s_3) = (0,0,0)$  separated by approximately 5-6 unit cells along the armchair direction ( $V_s = -400$  mV, scale bar = 1 nm). **e**, Two-state current signal  $I_t(t)$  observed in constant-height when gating the single cobalt atom at the red “x” in **a** with  $V_s = 550$  mV. **f-g**, Multi-state current signal  $I_t(t)$  measured at the red “x” in **c** and **d** with  $V_s = 550$  mV. **h**, Time-integrated probability distribution ( $\mathcal{P}(s)$ ) for each valency ( $s = 0$  or  $1$ ) in the isolated Co atom. **i-j**,  $\mathcal{P}(s)$  for dimer and trimer systems, respectively.



**Figure 2. Construction of a binary atomic synapse via anisotropic coupling.** **a**, Schematic representation of the three atom system in **c** and **d**, where the two atoms  $s_1$  and  $s_2$  are strongly coupled and the  $k$  atom is weakly coupled to  $s_1$  and  $s_2$ . **b**, Factor-graph representation of the Boltzmann machine mapped to the experimental configuration depicted in **a** and shown in **c** and **d**. **c-d**, Constant-current STM topography with  $k=0$  and  $k=1$ , respectively ( $V_s = -400$  mV, scale bar = 1 nm). **e-f**,  $I_t(t)$  measured at the red “x” in **c** and **d** with  $V_s = 500$  mV. **g-h**, Conditional, time-integrated probability distributions  $\mathcal{P}(s|k)$  for  $k=0$  and  $k=1$ , respectively.



**Figure 3. Multi-valued synapses.** **a**, Constant-current STM images of a seven atom cobalt ensemble with three *s* atoms and four *k* atoms ( $V_s = -400$  mV, scale bar = 1nm). The *k* configurations are: *i* = *k* = (0000), *ii* = *k* = (1000), *iii* = *k* = (0001), *iv* = *k* = (0110), and *v* = *k* = (1010). **b**,  $I_t(t)$  measured at the red “x” positions in **a** with  $V_s = 500$  mV. **c**, Conditional probability distributions,  $\mathcal{P}(s|\mathbf{k})$ .



**Figure 4. Synaptic dynamics and self-adaption.** **a**, Constant-current STM image of a seven atom cobalt ensemble with three  $s$  atoms and four  $k$  atoms ( $V_s = -400$  mV, scale bar = 1nm). **b**, Factor graph representation of the Boltzmann machine used to model the atomic ensemble in **a**, with a diagram illustrating the influence of the environment on the BM. **c**,  $I_i(t)$  taken at the red “x” in **a** with  $V_{\text{off}} = 160$  mV showing a spontaneous modification of the system’s  $k$  (before in grey and after in blue), which is observed through the distinct current levels and stochastic dynamics. **d-f**, Probability distributions  $\mathcal{P}(s, k | \epsilon)$  shown in blues and  $s$ -integrated probability distribution ( $\mathcal{P}(k | \epsilon)$  – grey histograms), given the environmental conditions  $\epsilon_1 \equiv V_{\text{off}} = 160$  mV,  $\epsilon_2 \equiv V_{\text{off}} = 180$  mV,  $\epsilon_3 \equiv V_{\text{off}} = 200$  mV.

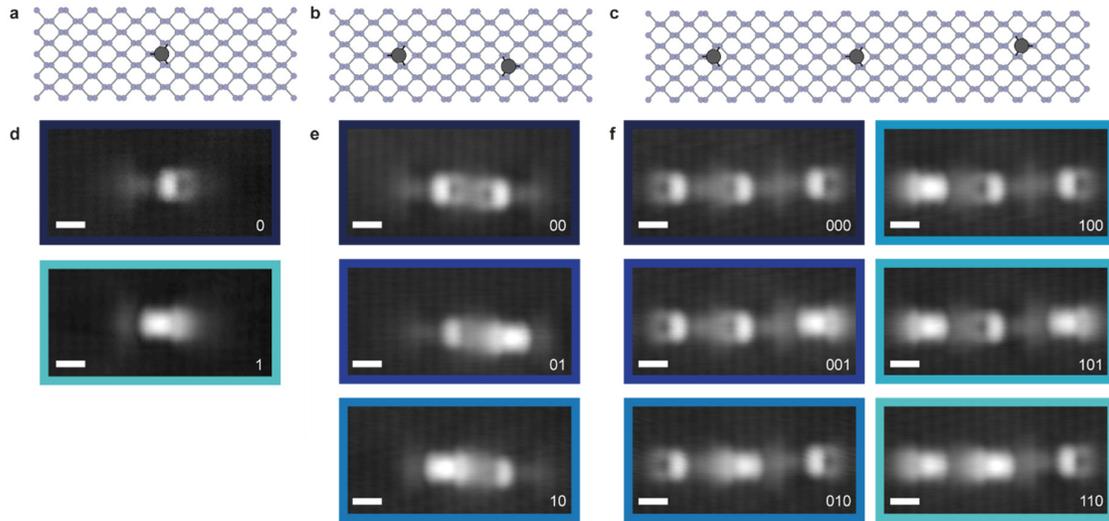
	$k = 0$	$k = 1$
$w_{12}$	-4.48	-6.2
$b_1$	0.28	-0.14
$b_2$	1.78	0.1

**Table 1. Weights and Biases for the  $k$  configurations shown in Fig. 2.**

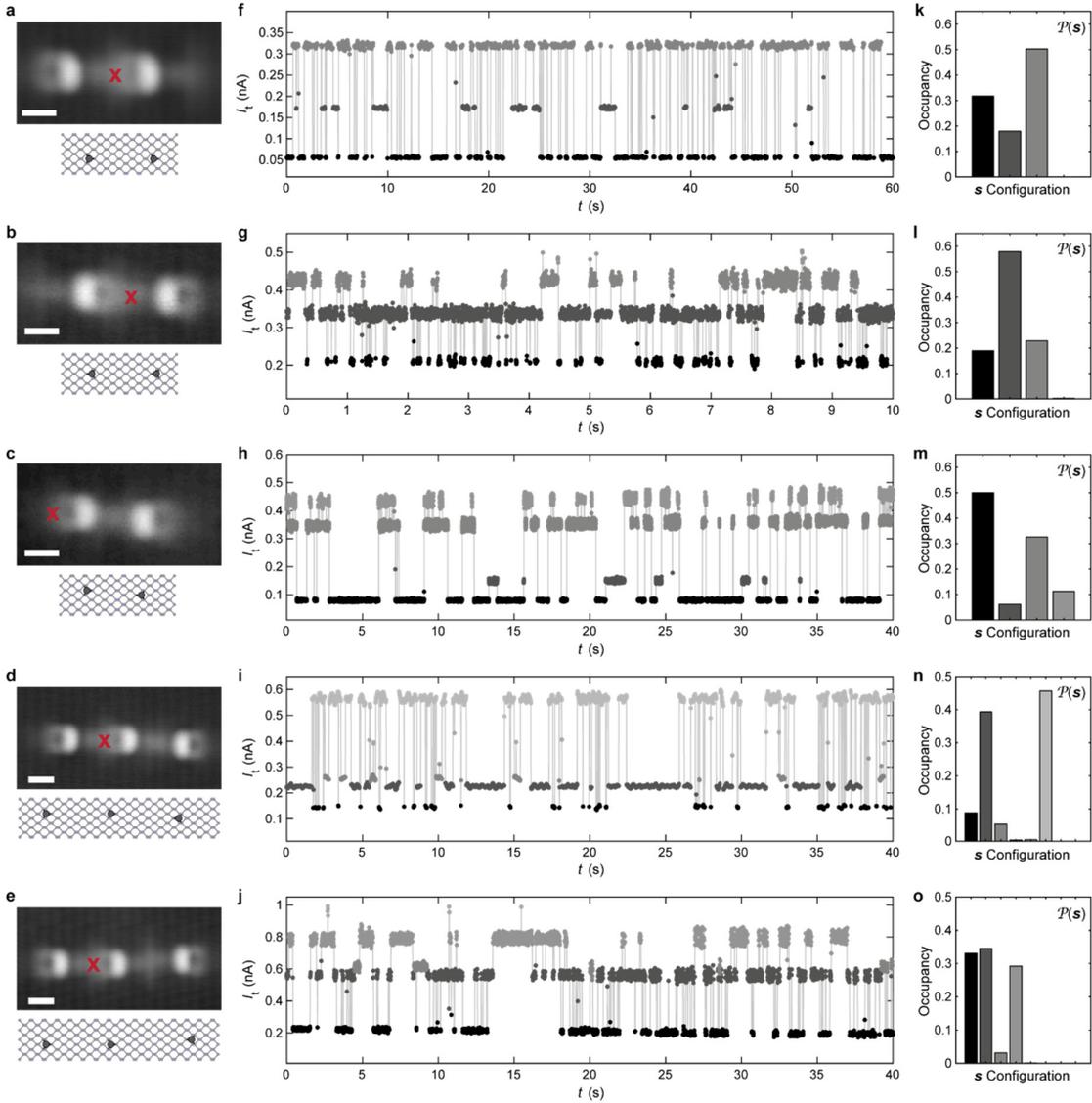
	$k=(0000)$	$k=(1000)$	$k=(0001)$	$k=(0110)$	$k=(1010)$
$w_{12}$	17.88	-16.66	6.68	-17.88	-2.44
$w_{23}$	13.14	-15.9	13.76	-10.72	-1.9
$w_{13}$	19.58	-10.72	7.48	-18.32	-2.8
$w_{123}$	-16.88	13.36	-10.16	14.4	2.24
$b_1$	-17.84	2.3	-1.58	18	-11.94
$b_2$	-19.3	16.52	-13.62	3.1	1.9
$b_3$	-12.52	9.82	-22.78	11.28	2.12

**Table 2. Weights and Biases for the  $k$  configurations shown in Fig. 3.**

## Extended Data

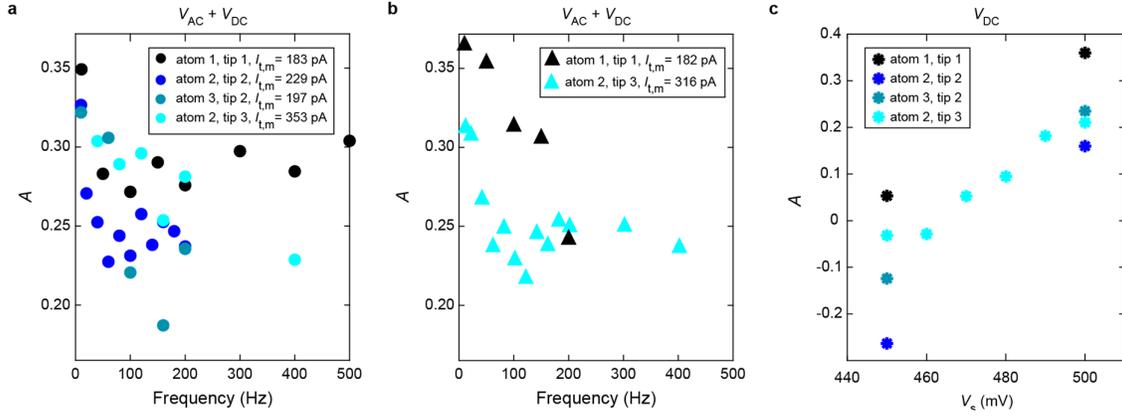


**Extended Data Figure 1. Schematics and valency configurations of coupled Co atoms on BP.** In order to correlate the atomic valency (seen in **d-f**) to a specific level in the current signal  $I_t(t)$  (see Figure 1) the following procedure was used. Stochastic current noise was measured by applying a voltage with  $V_s > V_{th}$  for a specified period of time. Immediately after this period of time, the applied voltage was quickly (less than 100 ms) reduced to  $V_s < V_{th}$  in order to freeze the final atomic valency. The state of the atom or atoms ( $s_i$ ) could then be measured using constant-current STM with  $V_s < V_{th}$ . This procedure was repeated 2-6 times for each state in order to ensure no additional switches occurred during the voltage ramp to  $V_s < V_{th}$ . For extremely short-lived states, the  $I_t(t)$  curve was manually terminated after the system switched into the target current level. The subsequent identification procedure was kept the same. **a-c**, Schematic of the BP lattice structure with the atomic positions of the coupled Co atoms indicated. **d-f**, STM topographies of the different configurations that were observed in the  $I_t(t)$  traces of the configurations seen in Figure 1 ( $V_s = -400$  mV). For the case of two and three coupled atoms **e-f**, the (11) and (111) configurations could not be trapped due to their short lifetime.

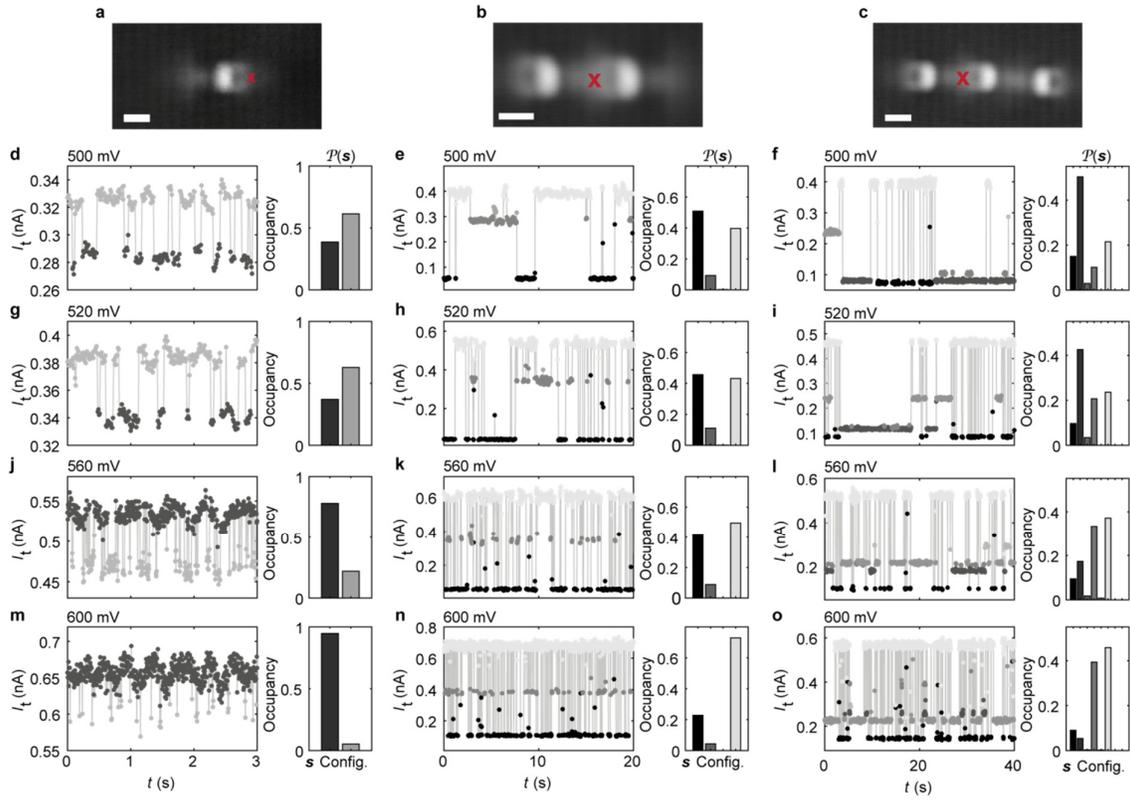


**Extended Data Figure 2. Neural dynamics of various configurations of coupled Co atoms on BP.**

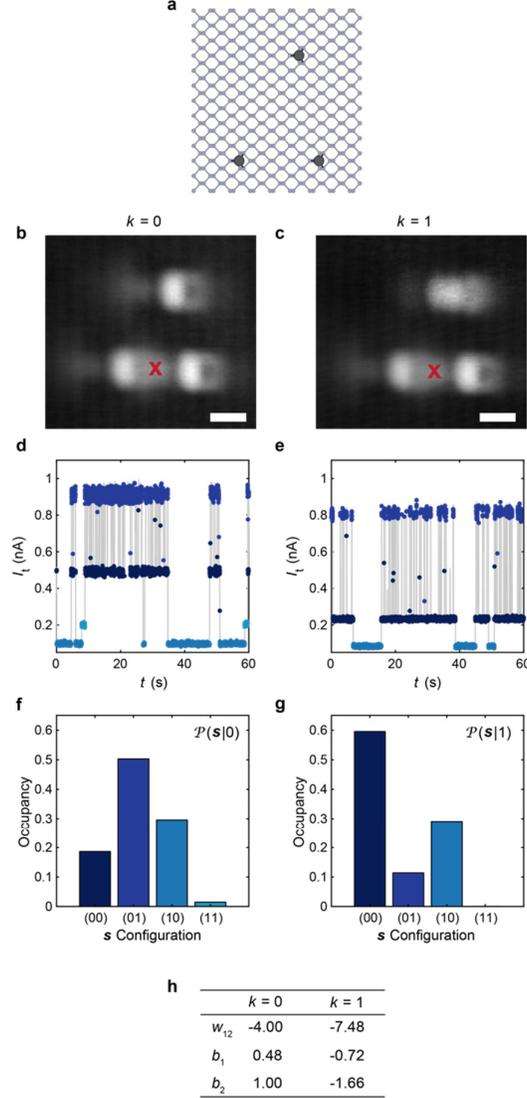
As seen in the atomic models shown in **a-e**, the relevant regime of coupling between  $s$  atoms in this work is reached when the interatomic separation is between four and eight BP lattice sites in the  $x$  direction. As this coupling is quite long-range, it is relatively robust to small changes in the precise atomic positions of the individual  $s$  atoms. Panels **a-e** show STM topographies of several configurations of two and three coupled Co atoms on BP ( $V_s = -400$  mV, scale bar = 1 nm) and models of the atom positions on the BP. **f-j**  $I_t(t)$  measured at the red X in **a-e**. Note that the configuration in **e** is the same as in Figure 1, but the  $I_t(t)$  trace is measured at a different position. The tip position does affect the probability distribution  $\mathcal{P}(s)$ , but it does not affect the observation of stochastic and multi-state behaviour. Bias voltage for **f**:  $V_s = 540$  mV, **g, h, j**:  $V_s = 550$  mV, **i**:  $V_s = 600$  mV. **k-o**, Time-integrated probability distribution  $\mathcal{P}(s)$ . To determine the probability distributions, at least 200 switching events were measured. The various  $\mathcal{P}(s)$  distributions reveal that qualitative differences in each of the energy landscapes exist. However, the overall presence of a multi-well potential is present in all configurations. We observe that the multi-well behaviour persists for coupled Co atoms whose interatomic separation is between 1.6 nm and 3.0 nm. The data in this figure was collected with different tip apices.



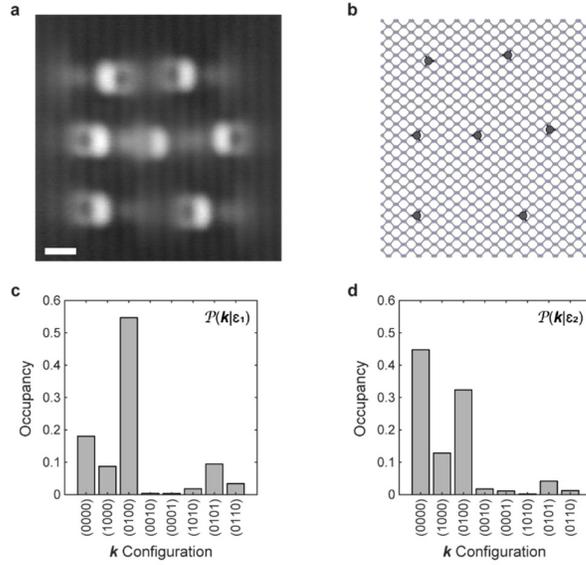
**Extended Data Figure 3. Frequency dependent response of a single cobalt atom on BP. a-b,** Asymmetry of the probability distribution  $\mathcal{P}(s)$  as a function of frequency of the applied AC voltage for sine and triangle waveforms, respectively. The asymmetry is defined as  $A = \mathcal{P}(0) - \mathcal{P}(1)$ . The experiment was performed on different atoms with different microtips (see legends). For each plotted data point multiple  $I_t(t)$  traces were taken. Small tip variations led to differences in the mean current  $I_{t,m}$  for each  $I_t(t)$  measurement (see legends). Individual current traces with a mean current within 14 pA or roughly twice the standard deviation of the overall mean  $I_{t,m}$  were used to calculate  $\mathcal{P}(s)$ ,  $A$ . The data was taken with a constant DC offset voltage below the switching threshold ( $V_s = 350$  mV) and an AC voltage was added with peak-to-peak amplitude  $V_{AC,pp} = 300$  mV. As the DC voltage was set below the gate threshold  $V_{th}$ , the cumulative DC + AC voltage periodically exceeded the gate threshold at intervals determined by the waveform and frequency. We confirmed the fidelity of the ingoing waveform at various amplitudes and frequencies. From **a** and **b**, the asymmetry increases with decreasing frequency below approximately 200 Hz. Thus, the single-atom probability distribution  $\mathcal{P}(s)$  can be influenced by AC voltages in a non-linear fashion. **c**, Asymmetry of the probability distribution  $\mathcal{P}(s)$  as a function of  $V_s$ , for the different microtips that were used in the AC experiment in **a** and **b**. For all microtips used in this experiment, the asymmetry increases with increasing  $V_s$ .



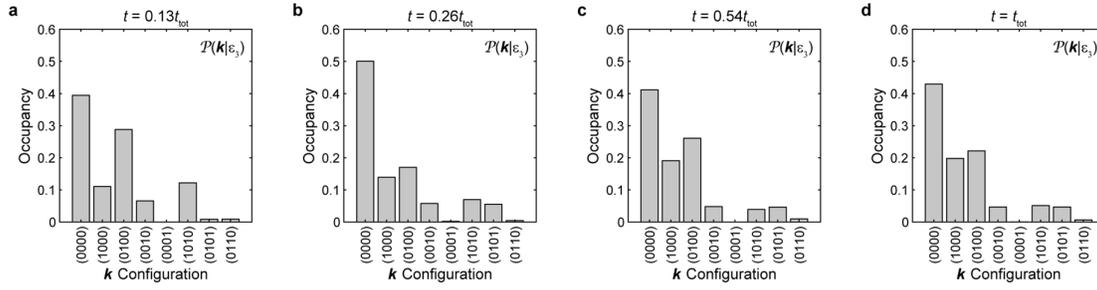
**Extended Data Figure 4. Response of coupled cobalt atoms on BP to varying DC voltages.** **a-c**, STM topographies of a single Co atom on BP, two coupled atoms and three coupled atoms, respectively ( $V_s = -400$  mV, scale bar = 1 nm). **d-o**,  $I_t(t)$  traces (left) and time-integrated probability distributions  $\mathcal{P}(s)$  (right) at different DC biases: **d-f**  $V_s = 500$  mV, **g-i**  $V_s = 520$  mV, **j-l**  $V_s = 560$  mV and **m-o**  $V_s = 600$  mV. The  $I_t(t)$  traces were measured at the red X's in **a-c**, all with the same tip apex. To determine the time-integrated probability distributions, at least 300 switching events were measured. **d, g, j, m** correspond to the configuration in **a**; **e, h, k, n** correspond to the configuration in **b** and **f, i, l, o** correspond to the configuration in **c**. From the  $I_t(t)$  traces, it is clear that for higher DC voltages ( $V_s$ ) the switching rate is higher. Furthermore, when increasing the number of coupled atoms, the switching rate decreases. As seen in the probability distributions  $\mathcal{P}(s)$  for the single atom (**d, g, j, m**), the state favourability is strongly modified by the applied DC voltage ( $V_s$ ). For the coupled dimer and trimer, the steady-state distribution  $\mathcal{P}(s)$  evolves in a complex, non-linear manner with increasing  $V_s$ . All the data in this figure was collected with the same tip apex.



**Extended Data Figure 5. Influence of the microtip on the binary atomic synapse from Figure 2.** In order to understand the influence of the STM tip on the  $I_t(t)$  measurement and subsequently  $\mathcal{P}(s|k)$  distributions, we studied the exact same atomic ensemble from Figure 2 with a different microtip. In general, after preparing the tip on Au(111) as described in the methods, small tip modifications were performed on nearby cobalt atoms, by gently moving the tip into contact with the atom. Small reconstructions of the tip apex were characterized using constant-current STM images of the top-site Co atoms. We understand such changes to the apex of the STM tip as small changes to the local band bending profile (schematically illustrated in Figure 1). All the data in this figure was collected with the same tip apex and demonstrates that the small modifications to the local band bending modify the  $w_i$  and  $b_i$  of the BM, but do not destroy the presence of the tunable, multi-well energy landscape. **a**, Schematic of the BP lattice structure with the atomic positions of the coupled Co atoms indicated. **b-c**, Constant-current STM topography with  $k=0$  and  $k=1$ , respectively ( $V_s = -400$  mV, scale bar = 1 nm). **d-g**,  $I_t(t)$  traces and time-integrated probability distributions  $\mathcal{P}(s|k)$  measured with a slightly different tip apex compared to the data shown in Figure 2. **h**, Weights and biases for the two  $k$  configurations.



**Extended Data Figure 6. Influence of the microtip on  $\mathcal{P}(k|\varepsilon)$  in the Co atom ensemble in Figure 4.** In order to further confirm that the specific termination of the STM tip was not responsible for the self-adaption demonstrated in Figure 4, we studied the response  $\mathcal{P}(k|\varepsilon)$  for the exact same atomic ensemble in Figure 4 with a different microtip. Again, interpreting the modification of the tip apex as a shift in the local band bending, it is clear that the qualitative change in  $\mathcal{P}(k|\varepsilon)$  observed in Figure 4 is well reproduced with a different microtip. In this case, the applied offset biases might be slightly shifted with respect to the data in Figure 4, consistent with a slight offset in the tip's work function. **a**, Constant-current STM topography of the seven Co atom ensemble in Figure 4 ( $V_s = -400$  mV). **b**, Schematic of the BP lattice structure with the atomic positions of the seven Co atoms indicated. **c-d**, s-Integrated probability distribution  $\mathcal{P}(k|\varepsilon)$  for  $\varepsilon_1 \equiv V_{\text{off}} = 180$  mV,  $\varepsilon_2 \equiv V_{\text{off}} = 200$  mV for a slightly different microtip compared to the distributions shown in Figure 4.



**Extended Data Figure 7. Time-evolution of the probability distribution  $\mathcal{P}(k|\epsilon_3)$  in Figure 4.** To quantitatively understand the length of time necessary for  $\mathcal{P}(k|\epsilon)$  to reach a steady-state distribution, we consider the evolution of  $\mathcal{P}(k|\epsilon_3)$  at selected time intervals from Figure 4. **a**,  $\mathcal{P}(k|\epsilon_3)$  after 92 minutes in which 61  $k$ -switching events occurred (13% of the total measurement time). **b**,  $\mathcal{P}(k|\epsilon_3)$  after 180 minutes in which 101  $k$ -switching events occurred (26% of the total measurement time). **c**,  $\mathcal{P}(k|\epsilon_3)$  after 372 minutes in which 156  $k$ -switching events occurred (54% of the total measurement time). **d**,  $\mathcal{P}(k|\epsilon_3)$  after 684 minutes in which 340  $k$ -switching events occurred (100% of the total measurement time). It is seen that after 13% of the total measurement time, the two most favourable states are already clearly identified. After 54% of the time (**c**) the distribution no longer changes significantly. The data in this figure was collected with the same tip as in Figure 4.