A superconducting associative memory that is defect tolerant

2001 J. Phys.: Condens. Matter 13 L697

(http://iopscience.iop.org/0953-8984/13/29/101)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **13** (2001) L697–L703 PII: S0953-8984(01)24685-2

LETTER TO THE EDITOR

A superconducting associative memory that is defect tolerant

P Chandra[1] **and L B Ioffe**[2,3]

[1] NEC Research Institute, 4 Independence Way, Princeton, NJ 08540, USA
[2] Landau Institute for Theoretical Physics, Moscow, Russia
[3] Department of Physics, Rutgers University, Piscataway, NJ 08855, USA

Received 10 May 2001, in final form 18 June 2001
Published 6 July 2001
Online at stacks.iop.org/JPhysCM/13/L697

Abstract
We present a content-addressable memory whose central component is a superconducting crossbar array with $2N$ elements connected by N^2 junctions. Because multiple pathways exist between any two elements, this storage device is tolerant to physical defects in the interconnections. Furthermore, each pattern of N bits is stored non-locally in the N^2 junctions, so information access and retrieval are tolerant to input errors. This superconducting memory should exhibit picosecond single-bit acquisition times with negligible energy dissipation during switching and multiple non-destructive read-outs of the stored data.

The demand for high-density data storage with ultrafast accessibility motivates the search for new memory implementations. Ideally such storage devices should be robust against input error (fault tolerant) and unreliability of individual elements (defect tolerant). In conventional memory devices, bits are stored in distinct physical elements at specific locations; they are addressed by their spatial coordinates. The corruption of any input (bit address) or storage element results in the loss of a particular portion of the stored pattern. By contrast, in an associative memory each image can be addressed by a part of its content, even if somewhat corrupted, thus making it tolerant to input errors [1]. Furthermore, if it is constructed such that p patterns each of N pixels are stored non-locally in $N_{eff} > Np$ physical elements, it might also be robust against the unreliability of device components. A similar situation is familiar in holography, where an image is reconstructed from recorded interference fringes; this process is not significantly affected by partial corruption of the holographic plate.

The primary component of our defect-tolerant associative memory is a superconducting crossbar network (figure 1), where each bit is represented by a wire. More specifically, this array consists of a stack of two perpendicular sets of N parallel wires separated by a thin oxide layer [2–5]. At low temperatures a Josephson junction exists at each node of this array; each pattern of N bits is stored non-locally in these N^2 interconnections. The redundant crossbar design ensures that each pair of bits (wires) is physically linked by many paths, thereby allowing

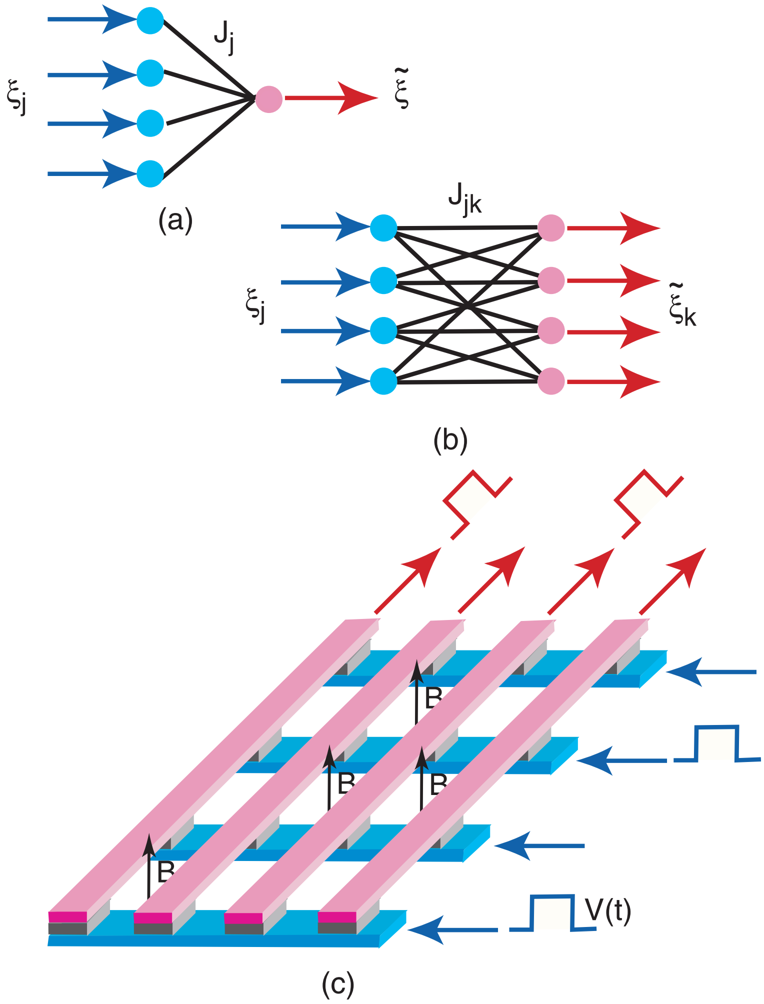

Figure 1. A logical representation of (a) a McCulloch–Pitts neuron and (b) a McCulloch–Pitts neural network. Here $\xi_j = \pm 1$ and $\tilde{\xi}_j = \pm 1$ are the inputs (blue) and the outputs (red) respectively, and J_{jk} are the couplings (black). The non-linear elements (neural cells) are magenta. (c) A schematic diagram of the crossbar Josephson memory cell where the inputs (blue) and outputs (red) are voltage pulses. The role of the logical couplings (black) is played by the Josephson junctions, whose phases are controlled by local applied fields B.

uninterrupted interbit contact despite some faulty couplings; a related architecture has been used to explore defect-tolerant computation [6].

In this content-addressable memory, each image of N bits is coded by $2N$ superconducting phases, $\{\phi_j, \phi_k\}$ with $1 \leqslant (j, k) \leqslant N$, associated with the $2N$ wires. More specifically, the desired patterns are stored in the N^2 array junctions and correspond to stable configurations of the $2N$ superconducting phases. Each stored image can be addressed/retrieved using voltage pulses and the Josephson phase–voltage relation

$$\Delta\phi_j = \frac{2\pi}{\Phi_0} \int V \, \mathrm{d}t$$

where $\Delta\phi_j$ is the phase change of the jth wire, and Φ_0 is the flux quantum. Thus the inputs/outputs of this superconducting memory are short voltage pulses with fixed area (cf. figure 1), in contrast to the constant voltage levels used in conventional semiconductor electronics.

Superconducting memory devices show promise for ultrafast high-density information storage with low power dissipation. In many early designs, digital information was stored locally in trapped magnetic fluxes that were switched between single-flux quantum (SFQ) states, $\Phi \in \{0, \Phi_0\}$, by Josephson junctions [7, 8]. In such static SFQ circuitry, the bit information was retrieved by voltage levels; the required reset of the latching junction after each retrieval event limited the clock speed of the cell. In order to maintain their advantage in speed compared to other technologies, Josephson-junction devices must use SFQ for both information storage and retrieval. Dynamical SFQ technology, where information is passed between circuit elements with SFQ voltage pulses ($\int V \, dt = \Phi_0$), offers this possibility [9]. Our superconducting memory cell incorporates this approach in a defect-tolerant design to ensure operation despite ever-present unreliable junctions; furthermore, its content addressability is appealing for high-density applications.

The important energy scales of the crossbar array are those associated with the superconducting wires and the Josephson junctions. Each superconducting wire is characterized by a macroscopic phase which is constant in equilibrium; here we assume that phase slips in each wire are energetically unfavourable. Application of a magnetic field results in the rotation of this phase along each wire, where the rotation angle is determined by the amplitude of the applied field [10]. The interaction energy of a Josephson junction is determined by the phase difference across its insulating layer; thus patterns can be written into the couplings of the superconducting network by tuning local applied fields.

The coexistence of multiconnectivity and non-linear elements (Josephson-coupled superconducting wires) in this array is a crucial feature that it shares with prototypical associative memories [1]. Theoretical studies indicate that this crossbar network has long-range temporal correlations (memory) and an extensive number of metastable states [5, 11]; furthermore, it has also been fabricated and studied in the laboratory [12, 13]. Thus it is natural to explore its possible use for content-addressable information storage, specifically pursuing its realization of a simple neural network. Here one would like to store p patterns, $\zeta_j^\mu = \pm 1$ $(1 \leqslant \mu \leqslant p)$, so that if the memory is exposed to a key (ξ_j) that has a significant overlap with a stored image:

$$q = \frac{1}{N} \sum_j^N \xi_j \zeta_j \geqslant \frac{1}{\sqrt{N}}$$

then its output is the desired pattern ($\tilde{\xi}_k = \zeta_k$). For convenience, we use the variables $\xi = \pm 1$ instead of the usual binary notation $n = 0, 1$; they are related by $\xi = 2n - 1$. A simple model for such a memory is an array of McCulloch–Pitts neurons (figure 1). The patterns are stored in the couplings, J_{jk}, which can be both positive and negative. Each non-linear element has multiple inputs, and the output is a non-linear function of the weighted sum of the inputs:

$$\tilde{\xi}_k = \mathrm{sgn}\left(\sum_j J_{jk} \xi_j \right)$$

where $\mathrm{sgn}(x) = \{+1, -1\}$ for $\{x \geqslant 0, x < 0\}$. Clearly the output is robust against input errors due to the multiple connections present.

In order to ensure that stored patterns can be retrieved by small keys, the couplings must be chosen such that these images correspond to stable configurations of the network. Hopfield has proposed an algorithm [14] where the desired patterns are local minima of an energy function, e.g.

$$H = -\frac{1}{2}\sum_{jk} J_{jk} S_j \tilde{S}_k$$

where $S_j, \tilde{S}_k \in \{-1, +1\}$. Here all possible states of the network are represented as a rough energy landscape whose stable minima correspond to stored patterns. Release of the system in any part of this landscape leads to its search for the nearest valley. In this algorithm, the couplings are chosen such that the energy is minimized for maximal overlap of $\tilde{S}_k$, the array configuration, and the desired output, $\tilde{\xi}_k$ $(=\zeta_k)$. For one pattern, this condition is satisfied for $J_{jk} = (1/N)\zeta_j\zeta_k$ where N is the number of pattern bits; then

$$H = -\frac{1}{2N}\left(\sum_k \tilde{S}_k \zeta_k\right)^2.$$

With this choice of J_{jk}, the output is identical to that of the McCulloch–Pitts network since $\tilde{S}_k \approx \zeta_k$ minimizes H. We also note that, with these J_{jk}, a real input $S_j \approx \zeta_j$ yields the desired output if it has errors in almost all its bits. In the Hopfield model, the couplings associated with several stored images are simple superpositions of the one-pattern case such that

$$J_{jk} = \frac{1}{N}\sum_{\mu=1}^{p} \zeta_j^{\mu}\zeta_k^{\mu}$$

where μ labels each pattern. The total pattern storage capacity of the network, p_{max}, is dependent on the acceptable error rate; in general $p_{max} = \alpha N$ where $\alpha \leqslant 0.138$ if the allowed bit-error probability in each pattern is $P_{error} < 0.01$ for Ising spins; for xy-spins [1] with binary couplings ($J_{jk} = \pm 1$), the case relevant for the superconducting array, $\alpha \sim 0.1$. Here we discuss the Hopfield algorithm because of its simplicity, but we note that other more efficient algorithms [1] can also be implemented in this network.

The crossbar Josephson array (figure 1) can be adapted to become a superconducting analogue of a McCulloch–Pitts network. It is described by the Hamiltonian

$$\mathcal{H}_0 = \mathrm{Re}\sum_{jk} S_j^* J_{jk} \tilde{S}_k \tag{1}$$

with $1 \leqslant (j,k) \leqslant N$ where j (k) labels the horizontal (vertical) wires respectively. S_j and $\tilde{S}_k$ are effective complex spins with unit amplitude, e.g. $S_j = \mathrm{e}^{\mathrm{i}\phi_j}$ where ϕ_j is the phase of the jth superconducting wire. In this implementation S_j and $\tilde{S}_k$ code the stored images. Their relative change with respect to a reference pattern ($S_j = \tilde{S}_k = 1$ for all j, k) is accessed using voltage pulses and the Josephson relation previously discussed. More specifically a key of voltage pulses can be applied to a small subset ($>\sqrt{N}$) of the horizontal wires, thereby altering the phase differences at the associated nodes. The phases of the vertical wires must readjust in order for the system to settle into a stable configuration, a process which results in the absence/presence of a voltage pulse. We re-emphasize that the array phases must be reset to a reference state with $\phi_j = \phi_k = 0$ $(1 \leqslant j,k \leqslant N)$ before subsequent image acquisition; this will be discussed further below. Each stored image of N output voltage pulses can therefore be addressed/retrieved by a particular key. Half-integer SFQ voltage pulses ($\Delta\phi = \int V\,\mathrm{d}t = \Phi_0/2$) may be used in direct analogy with the McCulloch–Pitts array where inputs $n_k \in \{0, 1\}$ now refer to the absence/presence of a SFQ pulse.

The couplings of the superconducting array:

$$J_{jk} = \frac{E_J}{\sqrt{N}} \exp\frac{2\pi\mathrm{i}\Phi_{jk}}{\Phi_0}$$

are determined by the magnetic flux, Φ_{jk}, and the energy scale of an individual junction, E_J. In order to determine J_{jk}, we consider the closed rectangle formed by horizontal (vertical) wires j (k) and completed by the sample edges passing through the origin. The flux enclosed in this loop contributes a phase Φ_{jk}/Φ_0 to the coupling. For a uniform magnetic field H, $\Phi_{jk} = H(jk)l^2$ where l is the interwire spacing. We emphasize that the sign of J_{jk} depends on the value of Φ_{jk}/Φ_0. In complete analogy with the McCulloch–Pitts network, patterns are stored in this superconducting associative memory by appropriate choice of the weights, J_{jk}. Because the J_{jk} are functions of the enclosed fluxes, Φ_{jk}, they can be set to their desired values by appropriately tuning the local applied field H_{jk}. From a practical standpoint, these couplings are 'written' by local fields applied to individual plaquettes of area l^2. For an elementary cell with edges defined by $\{j, j+1, k, k+1)\}$, the product of the couplings at the four corners, $P_{j,j+1,k,k+1}$, determines the plaquette flux, $\Phi^{jk}_{\mathrm{plaquette}}$—if $P_{j,j+1,k,k+1} \geqslant 0$, then $\Phi^{jk}_{\mathrm{plaquette}} = \Phi_0/2$; otherwise, $\Phi^{jk}_{\mathrm{plaquette}} = 0$. We note that other algorithms can be used to determine the couplings with more complicated applied fluxes; here we use the Hopfield model as an illustrative example.

In practice, this writing procedure can be accomplished by a conventional planar array of superconducting quantum inference device (SQUID) loops superimposed on the crossbar network. The latter, unlike the former, traps flux; the desired plaquette in the superconducting associative memory can be addressed using a combination of voltage pulses and current biasing [9]. In order to ensure that such externally imposed fluxes control the superconducting phases of the crossbar array, we require that the self-induced flux associated with its area, $A = (Nl)^2$, is less than a single flux quantum. This condition for weak supercurrents and negligible induced fields puts limits on the normal-state resistance of each junction (R_0) and N [12, 13]. We note that an added advantage of weak supercurrents is the lack of unwanted cross-talk between adjacent elements; this problem often occurs when information is stored in flux, and the associated supercurrents are large. We emphasize that in our memory, the Josephson junctions switch fluxoids while the applied magnetic fluxes remain fixed; it is the supercurrents, not the local fields, that code the information that is retrieved.

The tolerance of the crossbar Josephson memory cell to defective elements and to input error results from the non-local nature of its data storage both at the physical and the logical levels. In conventional planar superconducting arrays there are $O(N^2)$ individual *short* superconducting wires, and the fluxoids are spatially confined to areas $A \sim l^2$ where l is the internode spacing. By contrast, in the multiconnected network the phases reside on the $2N$ wires of length Nl; thus the fluxoids here are extended to the entire system. Data are coded non-locally in configurations of these superconducting phases, similarly to the situation in an optical holographic storage device [15]. There the stored patterns are independent of the input and an analogous superconducting holographic memory can be constructed. For example, let us consider the stored configuration $\zeta_k^p = \exp(2\pi \mathrm{i} kp/N)$ where k and p are indices labelling the horizontal wires and the stored patterns respectively. Then the input voltage pulses would be

$$\int V\,\mathrm{d}t = \left[\frac{kp}{N}\right]\Phi_0$$

where [ ] refers to the fractional part. Using the clipped Hopfield algorithm [1], we have

$$J_{jk} = \frac{1}{N}\,\mathrm{sgn}\sum_{\mu}^{p}\zeta_j^{\mu}\zeta_k^{\mu}$$

which yields the desired output

$$\tilde{S}_k^{\mu} = \sum_j J^*_{jk}\zeta_k^{\mu} \approx \zeta_k^{\mu}.$$

We note that any orthogonal basis for the inputs will work; therefore the crossbar Josephson network can be used as a key component of both an associative and a holographic memory.

The associative memory described here codes the images using analogue variables, the superconducting phases of the $2N$ wires, which are vulnerable to error; thus in any practical realization these continuous degrees of freedom must be clipped using a double-well potential. Furthermore, an additional field must be imposed to ensure that the reference state ($\phi_j = \phi_k = 0$ for all j, k) is stable. One of the simplest ways to incorporate both of these features into our system is to add two terms to our original $\mathcal{H}_0$ resulting in the Hamiltonian

$$\mathcal{H} = \mathcal{H}_0 - h_1 \sum_j \cos 2\phi_j - h_2 \sum_k \cos \phi_k \tag{2}$$

where $\mathcal{H}$ is defined in equation (1), and h_1 and h_2 are parameters that are chosen to optimize performance by minimizing error. In figure 2, we display numerical results with typical values of h_1 and h_2. Here the crucial point is that one can tune these parameters such that there is error-free recovery of a stored N-bit image for a key larger than ηN bits where $\eta = 0.3$ for $N = 128$. Thus, consistent with our expectations, we have a clear demonstration that this array is content addressable and fault tolerant.

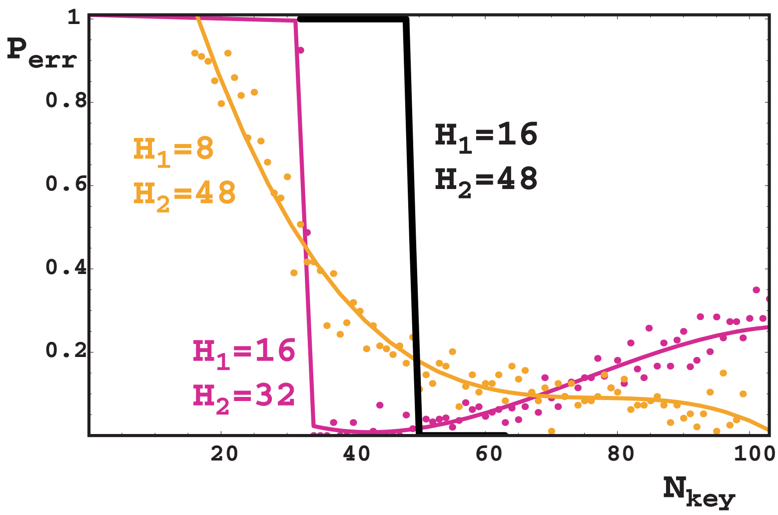

Figure 2. The probability of a single error as a function of the key length in the system of $N = 128$ horizontal/vertical wires and 16 stored images for three different choices of threshold fields. Clearly, the choice $H_1 = 16$, $H_2 = 48$ gives very reliable recovery for any key exceeding 48. The optimum (high reliability and large minimum key length) choice for this number of wires is $H_1 \approx 16$, $H_2 \approx 40$.

Practically, this superconducting memory consists of superimposed conventional and crossbar Josephson networks for writing and reading respectively, and a phase-reset circuit. The latter can be built from a set of planar double-junction SQUID loops connected to each horizontal wire and frustrated by a local magnetic field; they provide variable grounding to the circuit. A large coupling to ground locks the relevant phase into a reference state, whereas a weak one allows the next data retrieval process to be performed [16]. Similarly the first clipping field can be implemented by connecting each horizontal wire to a four-junction SQUID loop with a π-flux where the strength of the junctions determines h_1. The second field, h_2, is realized by coupling each vertical wire to ground via a single Josephson junction.

The resulting memory cell can then be embedded in an environment with known input/output SFQ circuitry that includes DC/SFQ voltage pulse converters and SFQ transmission/amplification lines [9]. The network parameters, the individual junction resistance (R_0), capacitance (C_0), critical current (I_c), and the coupling (E_J) and the charging (E_C) energy scales, should be chosen to optimize performance by minimizing access and retrieval times. The read-out procedure in the proposed associative memory is limited by the access time of each image which, in turn, is determined by the system's relaxation time, τ_R, to an energy minimum.

Since the phase differences at each junction evolve simultaneously, the single-junction switching time (τ_0) provides a natural timescale for the equilibration of the whole array. Thus this memory is highly parallelized from a computational standpoint. Naturally one expects much longer relaxation times near a critical point, which here corresponds to the situation where the maximum number of patterns (p_{max}) is stored in the crossbar network. Indeed our numerics indicate that for the number of stored patterns $p < p_{max}$, the relaxation occurs very fast; specifically for the network of $N = 128$ wires and $p = 16$ (where $p_{max} \sim 20$), we find that $\tau_R \leqslant 10\tau_0$. Thus to optimize performance, we want to minimize τ_0, the single-junction timescale that determines the image access time:

$$\tau_0 \sim \max\left(\frac{1}{\Delta}, \sqrt{\frac{R_0}{N^2 E_J R_Q}}\right)$$

where Δ is the superconducting gap, $E_J = I_c\Phi_0/(2\pi c)$ and $R_Q = \hbar/e^2$ is the quantum of resistance.

There are two additional conditions for E_C and R_0, $E_C \leqslant 0.01N^2 E_J$ and $(1/N)R_0 \ll R_Q$, that are required to minimize errors by ensuring weak phase fluctuations; here $E_C = e^2/(2C_0)$. The energy dissipated in each read-out process is approximately the total Josephson energy per wire ($\sim NE_J$), which is roughly $NR_Q/(R_0\Delta)$ for the optimal parameters. For aluminium wires this minimal dissipation per bit is 10^{-21} joules for aluminium wires in contrast to its value of $\sim 10^{-15}$ joules for conventional semiconducting electronics [17]. As an aside, we note that thermal noise plays a negligible role here for reasonable system sizes and temperatures since for the optimal choice of parameters ($E_J \sim (R_Q/R_0)\Delta$) the operating temperature $T \leqslant \Delta \ll E_{IM} \sim NE_J$ is significantly less than the typical energy scale associated with each image/wire, E_{IM}.

In summary, we have proposed an associative memory device that is a superconducting analogue of a McCulloch–Pitts network. Because this memory is intrinsically parallel due to its crossbar design, an image of N bits can be retrieved in a time (per bit) $\tau_{DT} \sim \tau_A/N$ where τ_A is the single-bit access time; by contrast $\tau_{DT} = \tau_A$ in a conventional local memory. An array of $N = 1000$ wires with $l = 0.5$ μm, which satisfies the self-induced flux criteria [12], has a bit capacity of $B_C = 0.1 \times N^2 = 10^5$ bits; a set of such arrays on a typical 1 cm^2 chip would then have a capacity of one gigabyte with an image access time (per bit) of $\tau_{DT} = 10^{-15}$ seconds. By contrast, current state-of-the-art nonvolatile memories (e.g. flash EEPROMs and ferroelectric RAMs) [18] have capacities of 100 kilobytes with 100 nanosecond single-bit access times; for a 1000-bit image the comparable image access time (per bit) is $\tau_{DT} = 10^{-7}$ seconds due to serial bit retrieval. The fault tolerance of the proposed superconducting memory enhances its appeal as a candidate for ultrafast high-density information storage without conventional problems of volatility, power dissipation, and subsequent heat removal. In contrast to most previous hardware implementations of associative memories, this design is tolerant to *both* physical and logical imperfections; this situation is reminiscent of the biological networks that inspired the original studies of content-addressable data storage [1]. It also shares this physical robustness with optical holographic memories, though its access time is significantly faster and its read-out is non-destructive. Such use of non-local degrees of freedom for information storage has also been proposed for fault-tolerant quantum computation since the effects of noise should be small [19]. We end by noting that we can field-tune the long-range array such that its stored images are maximally distant from each other in phase space. In this case the matrix elements associated with external noise will be negligible, and these patterns will have long decoherence times. Such orthogonal configurations could be promising as basis states for quantum logic processes.

References

[1] Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Redwood City, CA: Addison-Wesley)

[2] Vinokur V M, Ioffe L B, Larkin A I and Feigelman M V 1987 *Sov. Phys.–JETP* **66** 198–210

[3] Chandra P, Ioffe L B and Sherrington D 1995 *Phys. Rev. Lett.* **75** 713–16

[4] Chandra P, Feigelman M V and Ioffe L B 1996 *Phys. Rev. Lett.* **76** 4805–8

[5] Chandra P, Feigelman M V, Ioffe L B and Kagan D M 1997 *Phys. Rev.* B **56** 11 553–65

[6] Heath J R, Kuekes P J, Snider G S and Williams R S 1998 *Science* **280** 1716–21

[7] Fulton T A, Dynes R and Anderson P W 1973 *Proc. IEEE* **61** 28–35

[8] Likharev K K 1986 *Dynamics of Josephson Junctions and Circuits* (Amsterdam: Gordon and Breach)

[9] Likharev K K and Semenov V K 1991 *IEEE Trans. Appl. Supercond.* **1** 1–28

[10] Tinkham M 1980 *Introduction to Superconductivity* (New York: McGraw-Hill)

[11] Chandra P, Ioffe L B and Sherrington D 1998 *Phys. Rev.* B **58** 14 669–72

[12] Chandra P, Feigelman M V, Gershenson M E and Ioffe L B 1997 *Complex Behaviour of Glassy Systems* ed M Rubi and C Perez-Vicente (Berlin: Springer) pp 376–84

[13] Shea H R and Tinkham M 1997 *Phys. Rev. Lett.* **79** 2324–7

[14] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2254–558

[15] Buse K, Adibi A and Psaltis D 1998 *Nature* **393** 665–7

[16] Chandra P and Ioffe L B 1997 *US Patent number* 5,629,889

[17] Keyes R W 1988 *IBM J. Res. Dev.* **32** 24–8

[18] Scott J F 2000 *Ferroelectric Memories* (Berlin: Springer)

[19] Kitaev A Y 1997 *arXive e-print* quant-ph/970721
Nielson M A and Chung I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)