Abstract—In this paper, we analyze four typical sequential Hopfield neural network (HNN) based algorithms for image restoration and reconstruction, which are the modified HNN (PK) algorithm, the HNN (ZCVJ) algorithm with energy checking, the eliminating-highest-error (EHE) algorithm, and the simulated annealing (SA) algorithm. A new measure, the correct transition probability (CTP), is proposed for performance of iterative algorithms and is used in this analysis. The CTP measures the correct transition probability for a neuron transition at a particular time and reveals the insight of the performance at each iteration. The general properties of the CTP are discussed. Derived are the CTP formulas of these four algorithms. The analysis shows that the EHE algorithm has the highest CTP in all conditions of the severity of blurring, the signal-to-noise ratio (SNR) of a blurred noisy image, and the regularization term. This confirms the result in many previous simulations that the EHE algorithms can converge to more accurate images with much fewer iterations, have much higher correct transition rates than other HNN algorithms, and suppress streaks in restored images. The analysis also shows that the CTP’s of all these algorithms decrease with the severity of blurring, the severity of noise, and the degree of regularization, which also confirms the results in previous simulations. This in return suggests that the correct transition probability be a rational performance measure.

Index Terms—Image restoration, neural network, nonlinear detection.

I. INTRODUCTION

A. Motivation

Z HOU et al. [2] were the first who proposed the use of the HNN for image restoration. The modified Hopfield neural network (MHNN) models were proposed by Paik and Katsaggelos [3] for gray image restoration and by Sun and Yu [4] for binary image restoration and reconstruction. The algorithms based on the MHNN ensure network stability without checking energy change iteration by iteration. Sun and Yu [4] are the first who proposed a generalized framework for the MHNN algorithms with group updating. Sun [8] publicized a generalized updating rule (GUR) for the binary MHNN, which can operate in any sequence of updating modes with guaranteed stability.

In order to properly choose neurons to update at each iteration, Sun and Yu [5] proposed the eliminating-highest-error (EHE) criterion. Simulation results in [5] show that if the EHE criterion is used, the network can converge to a more accurate solution in much fewer iterations. In [6], a combination of an analog HNN and the MHNN is applied to blind image restoration. Simulation results show that only if the EHE criterion is applied, can the networks obtain an accurate image and accurate estimate of blur parameters. In [7] it is shown that the EHE criterion can suppress streaks in the restored image, which is a problem commonly existing in the HNN based algorithms and in the conventional algorithms. The EHE criterion is successfully applied to the code-division multiple-access (CDMA) multiuser detection in [13], [14] with promising results.

Motivated by revealing the mystery why the EHE criterion can help the MHNN restore high-quality images in much fewer iterations than other MHNN algorithms, in the companion paper [16], we find via simple analysis that the EHE criterion can achieve the highest correct transition probability at each neural transition. A new criterion, the fastest-energy-descent (FED) criterion, is proposed. It is shown that the FED criterion achieves the largest amount of energy descent. In the case of image restoration, the FED and EHE criteria are equivalent. Two new versions of the EHE based algorithm and three new algorithms based on the FED criterion are developed. A new performance measure, the correct transition rate (CTR), is proposed for the performance evaluation of iterative algorithms for simulation. Via simulation for gray image restoration, performance of six HNN based algorithms is compared, which are the SA algorithm [2], ZCVJ algorithm [2], PK algorithm [3, Algorithm 2], and three EHE algorithms (equivalent to the FED based algorithms for image restoration). In the simulation, all EHE algorithms demonstrated superior performance over the other algorithms in terms of the SNR and CTR. The goal of this paper is to analyze the performance of these algorithms.

In the literature of image restoration and reconstruction, the signal-to-noise ratio (SNR) of recovered image in simulation is usually used as a measure for the performance of an algorithm and for the quality of restored image. The visual quality of the recovered image is also considered as a subjective measure. However, the SNR and the visual quality of finally recovered image are only the output of the black box—the iterative algorithm, which can not reveal the insight of the performance of the algorithm. Because of this, the comparison of SNR does not insightfully benefit the development of an algorithm. For example, for any iterative algorithm existing in literature, it is unknown
In terms of the SNR how good a particular iteration is and how this iteration influences finally recovered image quality. This is apparently due to the lack of a proper performance measure for each particular iteration. Since the goal of all algorithms for image recovery is to reduce the difference between the recovered image and the original image, the probability that an iteration correctly reduces this difference can serve as a measure for the performance of the algorithm and serve as a predictor for the quality of finally recovered image. As the statistics used in an iteration are always ready, this probability is feasible to calculate. Since this probability depends on the operation of the algorithm at a particular iteration, it must reveal the insightful performance of the algorithm. This probability also depends on the degradation model, statistical parameters of noise distribution, and other statistics, and so reveals how these parameters influence the performance of the algorithm and the recovered image quality. Moreover, because this probability is also directly related to the energy function used in the algorithm, this probability may also help us find a better energy function more suitable for a particular algorithm and image degradation model.

In this paper, we propose such a new performance measure called the correct transition probability (CTP). The correct transition probability has a statistically different meaning from the correct transition rate proposed in [16]. The CTP is a measure for each particular iteration and is hard to estimate in simulation but suitable for theoretical derivation. In contrast, the CTR accounts for the average correct transition probability over all transitions in a long run of image recovery and is suitable for estimation in simulation. In this paper, the properties of the CTP are discussed. The formulas of the correct transition probability for the four sequential HNN based algorithms are derived and analyzed, which are the PK algorithm, ZCVJ algorithm, EHE algorithm, and SA algorithm. The analysis results confirm many phenomena observed in previous simulations.

B. Problem Formulation

In image restoration and reconstruction, a degraded image is formulated by the following linear equation [9]

\[
y = Ax + n
\]

(1)

where \(x' \in \{0, 1, \ldots, L - 1\}^n\), \(n, y \in \mathbb{R}^n\), and \(A \in \mathbb{R}^{m \times n}\). \(L\) is the total number of levels of image intensity, \(x'\) is the original image, \(A\) represents a spatially-shift-invariant linear system or a projection matrix, \(n\) is additive noise, and \(y\) is the observed blurred noisy image. In this paper, we assume that \(n\) is white Gaussian with zero mean and \(n\) by \(n\) covariance matrix \(\sigma^2 I\), where \(I\) is an identity matrix and \(\sigma^2\) is the noise variance per pixel. We assume that \(m = n\).

Given an \(A\), a solution of (1) can be obtained by the minimization of a squared error plus a regularization term [9]

\[
x: \min_{x \in \{0, \ldots, L-1\}^n} \|Ax - y\|^2 + \lambda \|Dx\|^2
\]

(2)

where \(D\) is a difference operator and \(\lambda \geq 0\). Usually, if \(x_i = \text{constant}\) for all \(i \in \{1, \ldots, n\}\), \(Dx = 0\). The purpose to add the regularization term is to compromise smoothness and fineness of recovered image when the noise is severe.

By defining an energy function

\[
E(x) = -\frac{1}{2} x^T (W + \lambda G)x + \Theta^T x
\]

(3)

where \(W = -A^TA, G = -D^TD\) and \(\Theta = -A^T y\), (2) can be rewritten as

\[
x: \min_{x \in \{0, \ldots, L-1\}^n} E(x).
\]

(4)

To solve (4) by using the HNN, let \(x, W + \lambda G\) and \(\Theta\) represent the network state, the interconnection matrix and the neural bias vector of the HNN, respectively. The negative gradient of the energy function \(E\) at time \(k\) is given by

\[
\frac{\partial E(x)}{\partial x} = -(W + \lambda G)x - \Theta.
\]

(5)

The correct transition probability \(h_i(k)\) is also the input of the \(i\)th neuron. In [16], a generalized updating rule (GUR) for images of \(L\)-level intensity is proposed.

**GUR:** Given a sequence of updating modes \(L(k) \subseteq \{1, \ldots, n\}\) for \(k \geq 0\), \(x(k)\) is updated by

\[
x_i(k + 1) = \begin{cases} x_i(k) + 1, & \text{if } i \in L(k), \ h_i(k) < t_i(k) \\ x_i(k), & \text{if } i \notin L(k), \ h_i(k) > t_i(k) \\ x_i(k) - 1, & \text{if } i \notin L(k), \ h_i(k) > t_i(k) \\ x_i(k), & \text{otherwise}, \end{cases}
\]

(6)

where \(t_i(k)\) is the \(i\)th-neuron threshold at time \(k\). The update stops at \(k_e\) when no neuron is updateable according to (6) for any \(k \geq k_e\).

The update stops at \(k_e\) when no neuron is updateable according to (6) for any \(k \geq k_e\).

In the sequential updating mode, \(L(k)\) at any time \(k\) contains only one element of \(\{1, \ldots, n\}\). The \(i\)th neural threshold in the sequential mode becomes

\[
t_i(k) = \frac{1}{2} \sum_{j \in \Omega(k)} |w_{ij} + \lambda g_{ij}|.
\]

(7)

In terms of the GUR, the state transition of the neurons whose indices are in \(J(k)\) reduces the energy \(E\) [16].

A different \(L(k)\) for \(k \geq 0\) defines a different algorithm and guides the network to converge to fixed points of different image quality. \(L(k)\) can be either pre-specified, or determined based on the current state of the network. The sequential MHNN algorithm proposed by Paik and Katsaggelos [3, Algorithm 2] up-
In statistical sense, this is equivalent to update neurons in a random choice. The PK algorithm can be as follows.

**PK Algorithm:** At time $k$, equiprobably choose an $i$ from $\{1, \ldots, n\}$ to form $L(k) = \{i\}$. Update the neural state by

$$x_i(k+1) = \begin{cases} x_i(k) - 1, & \text{if } i \notin L(k), \quad h_i(k) < -t \\ x_i(k) + 1, & \text{if } i \notin L(k), \quad h_i(k) > t \\ x_i(k), & \text{otherwise.} \end{cases}$$

As a special instance of the GUR [16], the PK algorithm ensures the energy descent at each nonzero update.

The ZCVJ algorithm proposed by Zhou et al. [2] equivalently uses the same updating rule of the PK algorithm and also ensures monotonic energy descent but has much more complex computations. The correct transition probability derived in this paper for the PK algorithm is also applicable to the ZCVJ algorithm.

The SA algorithm proposed by Zhou et al. [2] allows energy increase with a probability decreasing with time. Note [16] that the energy change due to the $i$th-neuron state transition $\Delta x_i(k) \neq 0$ is

$$\Delta E_i(k) = -\Delta x_i(k) h_i(k) - \frac{1}{2} (w_{ii} + \lambda g_i) = -\Delta x_i(k) h_i(k) + t. \quad (10)$$

Then the SA algorithm has the following form.

**SA Algorithm:** At time $k$, equiprobably choose an $i$ from $\{1, \ldots, n\}$ to form $L(k) = \{i\}$. Denote $p_\leftarrow^k = \exp[-\Delta E_i(k)/T]$ with $\Delta E_i(k)$ produced by $\Delta x_i(k) = -1$ and $p_\rightarrow^k = \exp[-\Delta E_i(k)/T]$ with $\Delta E_i(k)$ produced by $\Delta x_i(k) = 1$ where $T \geq 0$ is the temperature. Update the neural state by

$$x_i(k+1) = \begin{cases} x_i(k) - 1, & \text{if } i \in L(k) \text{ and } p_\leftarrow^k > \gamma, \text{ or } \\ x_i(k) + 1, & \text{if } i \in L(k) \text{ and } p_\rightarrow^k > \gamma \\ x_i(k), & \text{otherwise.} \end{cases} \quad (11)$$

where $\gamma$ is a random variable of uniform distribution over $[0, 1]$. \hfill \square

Sun and Yu [5] proposed two forms of the EHE criterion for the sequential and partially-simultaneous updating modes, respectively. In the sequential EHE based algorithm, $L(k)$ is determined by

$$L(k) = \{i \in J(k) \mid h_i(k) = \max_{j \in J(k)} |h_j(k)| \}. \quad (12)$$

Without loss of generality, we assume that only one $i$ satisfies the condition in (12). The sequential EHE algorithm [16] has the following form.

**EHE Algorithm:** At time $k$, determine $L(k)$ by (12). Update the neural state by

$$x_i(k+1) = \begin{cases} x_i(k) - 1, & \text{if } i \in L(k), \quad h_i(k) < -t \\ x_i(k) + 1, & \text{if } i \in L(k), \quad h_i(k) > t \\ x_i(k), & \text{otherwise.} \end{cases} \quad (13)$$

The difference among these PK, ZCVJ, SA, and EHE algorithms is that they apply different criterion in determination of a neuron to update at time $k$. All the PK, ZCVJ, and EHE algorithms only change the state of an updateable neuron whose index is in $J(k)$. However, the PK and ZCVJ algorithms change state of any updateable neuron whose index is in $J(k)$ without any selection, while the EHE algorithm changes the state of the unique updateable neuron not only whose index is in $J(k)$ but also whose absolute input value $|h_i(k)|$ is the largest among all the updateable neurons. As shown in [16], the EHE criterion implies the largest amount of energy descent at each iteration when applied to image restoration.

We note that although any nonzero neural state transition in the PK, ZCVJ, and EHE algorithms guarantees energy descent ($\Delta E(k) = E(k+1) - E(k) < 0$), it does not guarantee the decrease of $||\Delta x^i(k)||$ [more precisely, the decrease of $||\Delta x^i(k)||$] that is what we really want to achieve in image restoration and reconstruction. On the other hand, although the decrease of $||\Delta x^i(k)||$ is not guaranteed (i.e., with probability one), a neural state transition still has some probability to decrease $||\Delta x^i(k)||$. This probability is called correct transition probability in this paper. Obviously, the higher the correct transition probability is, the better the performance is. We will show in this paper that by using the EHE criterion, the EHE algorithm can achieve a much higher correct transition probability than other three algorithms.

The rest of the paper is organized as follows. In Section II, the correct transition probability is defined, and the statistics used in analysis are discussed. In Section III, the correct transition probability for the PK and ZCVJ algorithms is derived. In Section IV, the correct transition probability for the EHE algorithm is derived and compared with that of the PK and ZCVJ algorithms. The correct transition probability of the SA algorithm is derived and discussed in Section V. Section VI demonstrates numerical evaluations. Conclusions are made in Section VII and proofs are included in the Appendix.

**II. DEFINITION OF CORRECT TRANSITION PROBABILITY AND USEFUL STATISTICS**

A. Definition of Correct Transition Probability

Let $\Delta x^i(k) = x^i(k) - x^i(k)$ denote the $i$th element of $\Delta x^i(k) = x(k) - x(k)$. At any time $k$, the $n$ neural indices can be classified into three sets: $C(k) = \{i \mid \Delta x^i(k) = 0\}$, $\Omega^-(k) = \{i \mid i \in \{1, \ldots, n\}, \Delta x^i(k) < 0\}$ and $\Omega^+(k) = \{i \mid i \in \{1, \ldots, n\}, \Delta x^i(k) > 0\}$. We call $C(k)$ the index set of correct neural states, and $\Omega^-(k) = \Omega^-(k) \cup \Omega^+(k)$ the index set of the wrong neural states. Clearly, $C(k) \cup \Omega^-(k) \cup \Omega^+(k) = \{1, \ldots, n\}$.

A neural state transition $\Delta x_i(k) \neq 0$ is said a correct transition if it decreases $||\Delta x^i(k)|| = ||x^i(k) - x^i(k)||$, or a wrong transition otherwise. It is clear that 1) if $i \in C(k)$, the intensity $x^i(k)$ of the $i$th pixel of the original image is equal to the state $x_i(k)$ of the $i$th neuron. Either increasing or decreasing $x_i(k)$ by one is a wrong transition; 2) if $i \in \Omega^-(k)$, the intensity $x^i(k)$ of the $i$th pixel of the original image is smaller than the state $x_i(k)$ of the $i$th neuron. Decreasing $x_i(k)$ by one is a correct transition, and increasing $x_i(k)$ by one is a wrong transition; 3) if $i \in \Omega^+(k)$, the intensity $x^i(k)$ of the $i$th pixel of the original image is larger than the state $x_i(k)$ of the $i$th neuron. Increasing $x_i(k)$ by one
is a correct transition, and decreasing \( x_i(k) \) by one is a wrong transition.

Consider an algorithm that at time \( t \) chooses the \( i \)-th neuron to change its state. The correct transition probability of the \( i \)-th neuron is defined as the probability that conditioned with the occurrence of the \( i \)-th-neuron state transition, the transition decreases the difference between \( x_i(k) \) and \( x_i(k) \) (i.e., the transition is a correct transition)

\[
\eta_i = Pr \left( i \in \Gamma^-(k) \mid \Delta x_i(k) = -1, \, \text{or} \, i \in \Gamma^+(k), \right) \\
\Delta x_i(k) = 1 | \Delta x_i(k) \neq 0 \\
= Pr \left( i \in \Gamma^-(k) \right) Pr \left( \Delta x_i(k) = -1 | \Delta x_i(k) \neq 0 \right) \\
+ Pr \left( i \in \Gamma^+(k) \right) Pr \left( \Delta x_i(k) = 1 | \Delta x_i(k) \neq 0 \right).
\]  

\[ (13) \]

For the notation simplicity, we denote \( u_i^-(k) \equiv Pr(\Delta x_i(k) = -1 | \Delta x_i(k) \neq 0) \), \( u_i^+(k) \equiv Pr(\Delta x_i(k) = 1 | \Delta x_i(k) \neq 0) \), and \( p_i^{(g)}(k) \equiv Pr(\Delta x_i^g(k) = l) \). By Bayes’ theorem, (13) can be rewritten as

\[
\eta_i = Pr \left( i \in \Gamma^-(k) \mid \Delta x_i(k) = -1 \right) u_i^-(k) \\
+ Pr \left( i \in \Gamma^+(k) \mid \Delta x_i(k) = 1 \right) u_i^+(k)
\]

\[
= \sum_{l=-L+1}^{-1} Pr \left( \Delta x_i^g(k) = l | \Delta x_i(k) = -1 \right) u_i^-(k) \\
+ \sum_{l=1}^{L-1} Pr \left( \Delta x_i^g(k) = l | \Delta x_i(k) = 1 \right) u_i^+(k) \\
= \sum_{l=-L+1}^{-1} Pr \left( \Delta x_i(k) = -1 | \Delta x_i^g(k) = l \right) p_i^{(g)}(k) \\
+ \sum_{l=1}^{L-1} Pr \left( \Delta x_i(k) = 1 | \Delta x_i^g(k) = l \right) p_i^{(g)}(k)
\]

\[ (14) \]

The correct transition probability is defined as

\[
\eta = \sum_{i=1}^{n} \eta_i Pr(L(k) = \{i\}).
\]

\[ (15) \]

In (14), \( \{\Delta x_i(k) = -1 | \Delta x_i^g(k) = l \} \) for \( l < 0 \) and \( \{\Delta x_i(k) = 1 | \Delta x_i^g(k) = l \} \) for \( l > 0 \) are correct transitions at each difference level \( \Delta x_i^g(k) = l \). On the other hand, \( \{\Delta x_i(k) = -1 | \Delta x_i^g(k) = l \} \) for \( l \geq 0 \) and \( \{\Delta x_i(k) = 1 | \Delta x_i^g(k) = l \} \) for \( l \leq 0 \) are wrong transitions. The probabilities \( Pr[\Delta x_i(k) = -1 | \Delta x_i^g(k) = l] \) and \( Pr[\Delta x_i(k) = 1 | \Delta x_i^g(k) = l] \) are determined by the algorithm whose performance is measured. \( Pr(L(k) = \{i\}) \) in (15) also depends on the algorithm.

The correct transition probability has many properties. The following is the most interesting.

Property 1: The larger are the probabilities \( Pr[\Delta x_i(k) = -1 | \Delta x_i^g(k) = l] \) for \( l < 0 \) and \( Pr[\Delta x_i(k) = 1 | \Delta x_i^g(k) = l] \) for \( l > 0 \), or the smaller are the probabilities \( Pr[\Delta x_i(k) = -1 | \Delta x_i^g(k) = l] \) for \( l \geq 0 \) and \( Pr[\Delta x_i(k) = 1 | \Delta x_i^g(k) = l] \) for \( l \leq 0 \), the larger is the correct transition probability \( \eta \).

Property 2: For \( \forall i \), if \( Pr[\Delta x_i(k) = -1 | \Delta x_i^g(k) = l] = \alpha \neq 0 \) and \( Pr[\Delta x_i(k) = 1 | \Delta x_i^g(k) = l] = \beta \neq 0 \) are constants independent of \( l \), the algorithm has no restoration capability. If further the distribution of \( \Delta x_i^g(k) = l \) is symmetric to zero, \( \eta \leq 1/2 \) where the equality holds if and only if \( Pr(L(k) = \{i\}) = 1 \).

Property 4 is proved in the Appendix.

Property 5: If \( Pr[\Delta x_i(k) = -1 | \Delta x_i^g(k) = l] = \eta \), i.e., \( Pr[\Delta x_i(k) = 1 | \Delta x_i^g(k) = l] = \eta \), for all \( i \) and \( l \), \( x(k) \) is a fixed point of the algorithm with probability one.

Clearly, the closer is \( \eta \) to one, the fewer are the iterations and the more accurate is the restored image, and so the better is the performance of an algorithm; the closer is \( \eta \) to 0.5, the more the iterations are for an algorithm to converge, and the worse is the performance. If \( \eta < 0.5 \) for all \( k \), the algorithm has no restoration capability in statistic sense.

The popularly used SNR measures the performance of an algorithm and image quality based on only the finally-restored image. In contrast, the correct transition probability measures the performance of the algorithm at each neural transition. The correct transition probability also suggests the convergent rate and predicts the quality of finally restored image.

B. Useful Statistics

All the PK, ZCVJ, EHE, and SA algorithms make a nonzero neural transition \( \Delta x_i(k) \neq 0 \) based on the current network input vector \( h(k) \), the negative gradient of energy function. To derive the correct transition probabilities for these algorithms, we must know the properties of the statistics \( h_i(k) \) for \( j = 1, \ldots, n \). By noticing \( \Theta = -A^T y = W x^0 - A^T n \), we can rewrite the neural input vector (5) as

\[ h(k) = (W + \lambda G)x(k) = Wx^0 + A^T n \]

\[ = -W \Delta x(k) + \lambda Gx(k) + z \]

where \( z = A^T n \) can be shown a Gaussian random vector with zero mean and covariance matrix \( R_z = -\sigma^2 W \). We assume that \( \Delta x(k) \), \( x(k) \), and \( z \) are independent. When the network state is close to a fixed point after a number of iterations, the difference \( ||\Delta x^g(k)|| = ||x^0 - x(k)|| \) is small and the difference at each pixel can be uncorrelated. For this reason, we assume that \( \Delta x^g(k) \)'s are i.i.d. random variables with mean \( m_{\Delta x^g} \) and variance \( \sigma_{\Delta x^g}^2 \).

Since \( A \) and \( D \) usually represent operators of convolution over small windows, \( W \) and \( G \) are 2-D local operators. By \( \Psi_{w, i} = \{j | w_{ij} \neq 0\} \) and \( \Psi_{g, i} = \{j | g_{ij} \neq 0\} \), we define the sets of the indices of neurons that are located in the small windows determined by \( W \) and \( G \), respectively. In image restoration and reconstruction, the numbers of elements in \( \Psi_{w, i} \) and \( \Psi_{g, i} \) are usu-
ally much smaller than the number \( n \) of pixels (neurons) of the image. Because \( D \) is spatially invariant, in the notation of spatial location, the set \( \{ i | j \} \) for \( j \in \Psi_{m} \) is independent of \( i \). Similarly, in image restoration, because \( A \) is spatially invariant, the set \( \{ i | j \} \) for \( j \in \Psi_{m} \) is independent of \( i \). Since \( A \) and \( D \) are local operators on the image plane \( h(i,k) \) and \( h(j,k) \) are uncorrelated if the 2-D spatial distance between \( i \) and \( j \) is large.

We denote the random variable \( h_{m}(k) \) conditioned with \( \Delta x_{i}^{m}(k) = l \) by

\[
X_{m|i} \equiv \{ h_{m}(k) | \Delta x_{i}^{m}(k) = l \} = \{-u_{m} - \sum_{j \in \Psi_{m}, j \neq i} u_{m,j} \Delta x_{j}^{m}(k) + \lambda \sum_{j \in \Psi_{m}} g_{m,j} x_{j}(k) + z_{m}, \}
\]

By the law of large numbers, the first summation term in (17) can be approximated as a Gaussian random variable with mean \( m_{\Delta x_{m}}(k) = \sum_{j \in \Psi_{m}} g_{m,j} m_{x_{j}}(k) \) and variance \( \sigma_{\Delta x_{m}}^{2}(k) = \sum_{j \in \Psi_{m}} g_{m,j}^{2} \sigma_{m_{x_{j}}}(k)^{2} \). The second summation term in (17) can be approximated as a Gaussian random variable with mean \( m_{x_{m}}(k) = \sum_{j \in \Psi_{m}} g_{m,j} x_{j}(k) \) and variance \( \sigma_{x_{m}}^{2}(k) = \sum_{j \in \Psi_{m}} g_{m,j}^{2} \sigma_{x_{j}}^{2}(k) \). Finally, in (17), \( z_{m} = \sum_{j=1}^{n} a_{m,j} n_{j} \) is a Gaussian random variable with mean zero and variance \( \sigma_{z}^{2} = \sum_{j=1}^{n} a_{m,j}^{2} \sigma_{m_{x_{j}}}(k)^{2} = |w_{m}| \sigma_{z}^{2} \) that is independent of \( m \).

In the rest of the paper, we denote \( m_{\Delta x_{m}} = m_{\Delta x_{m}}(k), \sigma_{\Delta x_{m}} = \sigma_{\Delta x_{m}}(k), m_{x_{m}} = m_{x_{m}}(k), \sigma_{x_{m}} = \sigma_{x_{m}}(k), m_{n_{m}} = m_{n_{m}}(k), \sigma_{n_{m}} = \sigma_{n_{m}}(k), \) and \( \sigma_{z} = \sigma_{z}(k) \). Based on the above discussion, we have the following results:

**Property 6:**

1. \( X_{m|i} \) is approximately a Gaussian random variable with mean

\[
m_{m|i} = -u_{m|i} - m_{\Delta x_{m}} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j} + \lambda m_{x_{m}} \]

and variance

\[
\sigma_{m|i}^{2} = \sigma_{\Delta x_{m}}^{2} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j}^{2} + \lambda^{2} \sigma_{x_{m}}^{2} + |w_{m}| \sigma_{z}^{2}.
\]

2. In particular, \( X_{m|i} \equiv \{ h_{m}(k) | \Delta x_{i}^{m}(k) = l \} \) has mean

\[
m_{m|i} = -u_{m|i} - m_{\Delta x_{m}} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j} + \lambda m_{x_{m}} \]

and variance

\[
\sigma_{m|i}^{2} = \sigma_{\Delta x_{m}}^{2} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j}^{2} + \lambda^{2} \sigma_{x_{m}}^{2} + |w_{m}| \sigma_{z}^{2}.
\]

3. If the spatial distance between \( j \) and \( m \) is large, \( X_{m|i} \) and \( X_{m|i} \) are independent.

**III. CORRECT TRANSITION PROBABILITY OF PK AND ZCVJ ALGORITHMS**

The PK and ZCVJ algorithms have the same correct transition probability.

**Theorem 1:** For the PK and ZCVJ algorithms,

\[
F_{i}^{-}(l,k) \equiv \Pr (\Delta x_{i}^{m}(k) = -1 | \Delta x_{i}^{m}(k) = l) = \Phi \left( \frac{-m_{m|i} + t}{\sigma_{m|i}} \right),
\]

and

\[
F_{i}^{+}(l,k) \equiv \Pr (\Delta x_{i}^{m}(k) = 1 | \Delta x_{i}^{m}(k) = l) = \Phi \left( \frac{m_{m|i} - t}{\sigma_{m|i}} \right)
\]

where \( \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^{2}/2} \, dx \).

The PK and ZCVJ algorithms update \( \eta \) neurons equally likely in random or one by one in order. Hence, usually \( \Pr(L(k) = \{ i \}) = 1/n \). Theorem 1 results in the following corollaries.

**Corollary 1:** \( F_{i}^{-}(l,k) \) and \( F_{i}^{+}(l,k) \) are monotonically decreasing and increasing functions of \( l \), respectively.

**Corollary 2:** Assume \( m_{\Delta x_{m}} = m_{\Delta x_{m}}(k), \sigma_{\Delta x_{m}} = \sigma_{\Delta x_{m}}(k), m_{x_{m}} = m_{x_{m}}(k), \sigma_{x_{m}} = \sigma_{x_{m}}(k), m_{n_{m}} = m_{n_{m}}(k), \sigma_{n_{m}} = \sigma_{n_{m}}(k), \) and \( \sigma_{z} = \sigma_{z}(k) \). Based on the above discussion, we have the following results:

**Property 6:**

1. \( X_{m|i} \) is approximately a Gaussian random variable with mean

\[
m_{m|i} = -u_{m|i} - m_{\Delta x_{m}} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j} + \lambda m_{x_{m}} \]

and variance

\[
\sigma_{m|i}^{2} = \sigma_{\Delta x_{m}}^{2} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j}^{2} + \lambda^{2} \sigma_{x_{m}}^{2} + |w_{m}| \sigma_{z}^{2}.
\]

2. In particular, \( X_{m|i} \equiv \{ h_{m}(k) | \Delta x_{i}^{m}(k) = l \} \) has mean

\[
m_{m|i} = -u_{m|i} - m_{\Delta x_{m}} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j} + \lambda m_{x_{m}} \]

and variance

\[
\sigma_{m|i}^{2} = \sigma_{\Delta x_{m}}^{2} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j}^{2} + \lambda^{2} \sigma_{x_{m}}^{2} + |w_{m}| \sigma_{z}^{2}.
\]

3. If the spatial distance between \( j \) and \( m \) is large, \( X_{m|i} \) and \( X_{m|i} \) are independent.

**Corollary 1:** \( F_{i}^{-}(l,k) \) and \( F_{i}^{+}(l,k) \) are monotonically decreasing and increasing functions of \( l \), respectively.

**Corollary 2:** Assume \( m_{\Delta x_{m}} = m_{\Delta x_{m}}(k), \) and \( \sigma_{\Delta x_{m}} = \sigma_{\Delta x_{m}}(k), \sigma_{x_{m}} = \sigma_{x_{m}}(k), m_{n_{m}} = m_{n_{m}}(k), \sigma_{n_{m}} = \sigma_{n_{m}}(k), \) and \( \sigma_{z} = \sigma_{z}(k) \). Based on the above discussion, we have the following results:

**Property 6:**

1. \( X_{m|i} \) is approximately a Gaussian random variable with mean

\[
m_{m|i} = -u_{m|i} - m_{\Delta x_{m}} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j} + \lambda m_{x_{m}} \]

and variance

\[
\sigma_{m|i}^{2} = \sigma_{\Delta x_{m}}^{2} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j}^{2} + \lambda^{2} \sigma_{x_{m}}^{2} + |w_{m}| \sigma_{z}^{2}.
\]

2. In particular, \( X_{m|i} \equiv \{ h_{m}(k) | \Delta x_{i}^{m}(k) = l \} \) has mean

\[
m_{m|i} = -u_{m|i} - m_{\Delta x_{m}} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j} + \lambda m_{x_{m}} \]

and variance

\[
\sigma_{m|i}^{2} = \sigma_{\Delta x_{m}}^{2} \sum_{j \in \Psi_{m}, j \neq i} u_{m,j}^{2} + \lambda^{2} \sigma_{x_{m}}^{2} + |w_{m}| \sigma_{z}^{2}.
\]

3. If the spatial distance between \( j \) and \( m \) is large, \( X_{m|i} \) and \( X_{m|i} \) are independent.
Corollary 4: As $\sigma \to \infty$, the PK and ZCVJ algorithms have no restoration capability. If further the distribution of $\Delta x_\ell^\mu(k) = l$ is symmetric to zero, $\eta_{PK} = \eta_{ZCVJ} \leq 1/2$ where the equality holds if and only if $P_{\hat{F}}(C(k) = \emptyset) = 1$. 

Corollary 4 shows that if the noise is severe, the PK and ZCVJ algorithms are not capable to restore the image.

Corollary 5: As $\lambda \to \infty$, 1) the PK and ZCVJ algorithms have no restoration capability; 2) if the distribution of $\Delta x_\ell^\mu(k) = l$ is symmetric to zero, $\eta_{PK} = \eta_{ZCVJ} \leq 1/2$ where the equality holds if and only if $P_{\hat{F}}(C(k) = \emptyset) = 1/3$; if $x_i(k) = \text{constant}$ for all $i \in \{1, \ldots, n\}$, $\hat{x}(k)$ is a fixed point of the PK and ZCVJ algorithms with probability one.

Corollary 5 shows that if $\lambda$ is large, the state transition is independent of the original image and therefore the PK and ZCVJ algorithms have no restoration capability. Furthermore, a constant image is a fixed point. This confirms the fact that as $\lambda \to \infty$, any constant image is a minimum point of the energy function in (2).

Corollary 6: If $A = I$, $\sigma = 0$ and take $\lambda = 0$, then $\eta_{PK} = \eta_{ZCVJ} = 1$. 

Corollary 6 means that if there is neither the blurring nor the noise, given any initial image and set $\lambda = 0$, the PK and ZCVJ algorithms converge to the original image without error. 

Corollary 7: Assume $a_{ij} = (A)_{ij} = 1/n$ for $i$, $j \in \{1, \ldots, n\}$ and $\sigma = 0$. If $\Delta x_\ell^\mu(k)$ is uniformly distributed on $\{-L + 1, \ldots, -1, 1, \ldots, L - 1\}$ and take $\lambda = 0$, then $\eta_{PK} = \eta_{ZCVJ} \leq 1/2$.

In the case of Corollary 7, the observed data is a severely blurred noise-free image and contains only the average information of the original image. The PK and ZCVJ Algorithms are incapable to restore the image for $\eta_{PK} = \eta_{ZCVJ} \leq 1/2$. Corollaries 6 and 7 confirm the same results obtained in [7] by analyzing an upper bound of residual about the PK algorithm.

IV. CORRECT TRANSITION PROBABILITY OF EHE ALGORITHM

The EHE algorithm applies the EHE criterion in determination of the neuron to update at each iteration. We will show that the EHE criterion improves the performance of the EHE algorithm in terms of the correct transition probability.

Theorem 2: For the EHE algorithm, assume $X_{j\ell}$’s are independent,

$$Q_i^-(l, k) \equiv P_r(\Delta x_i(k) = -1 | \Delta x_i^-(k) = l) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-(m_i + l)/\sigma_{ij}} \exp \left( -\frac{x^2}{2} \right) \prod_{j=1, j \neq i}^n \Phi \left( -\frac{\sigma_{ij} x - m_{ij} - m_{i\ell}}{\sigma_{ji}} \right) dx,$$

$$Q_i^+(l, k) \equiv P_r(\Delta x_i(k) = 1 | \Delta x_i^+(k) = l) = \frac{1}{\sqrt{2\pi}} \int_{(m_i + l)/\sigma_{ij}}^{\infty} \exp \left( -\frac{x^2}{2} \right) \prod_{j=1, j \neq i}^n \Phi \left( \frac{\sigma_{ij} x + m_{ij} - m_{i\ell}}{\sigma_{ji}} \right) dx,$$

and the correct transition probability is

$$\eta_{EHE} = \frac{\sum_{l=-l+1}^{-1} Q_i^-(l, k) y_i^-(k)}{\sum_{l=-l+1}^{-1} Q_i^-(l, k) y_i^-(k) + \sum_{l=-l+1}^{-1} Q_i^+(l, k) y_i^+(k)}. \tag{27}$$

In parallel with Theorem 1, Theorem 2 results in the following corollaries. For simplicity, the proofs of some corollaries are omitted.

Corollary 8: $Q_i^-(l, k)$ and $Q_i^+(l, k)$ are monotonically decreasing and increasing functions of $l$, respectively.

Corollary 9: $\eta_{EHE}$ monotonically increases with $[|w_i|]$.

Corollary 10: If $m_{\Delta x} = 0$ and $\lambda = 0$, $\eta_{EHE}$ monotonically decreases with $\sum_{i=1}^n w_i^2$.

Corollary 11: $Q_i^-(l, k)$ and $Q_i^+(l, k)$ are monotonically decreasing and increasing functions of $m_{\Delta x}$, respectively.

Corollary 12: As $\sigma \to \infty$, the EHE algorithm has no restoration capability. If further the distribution of $\Delta x_i^\mu(k) = l$ is symmetric to zero, $\eta_{EHE} \leq 1/2$ where the equality holds if and only if $P_{\hat{F}}[C(k) = \emptyset] = 1$.

Corollary 13: As $\lambda \to \infty$, 1) the EHE algorithm has no restoration capability; 2) if the distribution of $\Delta x_i^\mu(k) = l$ is symmetric to zero, $\eta_{EHE} \leq 1/2$ where the equality holds if and only if $P_{\hat{F}}[C(k) = \emptyset] = 1/3$; if $x_i(k) = \text{constant}$ for all $i \in \{1, \ldots, n\}$, $\hat{x}(k)$ is a fixed point of the EHE algorithm with probability one.

Corollary 14: If $A = I, \sigma = 0$ and take $\lambda = 0$, then $\eta_{EHE} = 1$.

Corollary 15: Assume $a_{ij} = (A)_{ij} = 1/n$ for $i$, $j \in \{1, \ldots, n\}$ and $\sigma = 0$. If $\Delta x_i^\mu(k)$ is uniformly distributed on $\{-L + 1, \ldots, -1, 1, \ldots, L - 1\}$ and take $\lambda = 0$, then $\eta_{EHE} = 1/2$.

Corollaries 14 and 15 are similar to Corollaries 6 and 7. As stated in Corollaries 14 and 15, if the image is not blurred or is severely blurred, the performance of the PK (or ZCVJ) and EHE algorithms is the same. The reason is that the EHE criterion is invalid in these two extreme cases.

Corollary 16: $Q_i^-(l, k) \leq F_i^-(l, k)$ and $Q_i^+(l, k) \leq F_i^+(l, k)$ for all $l$.

Corollary 17: Assume that 1) the error pixels are isolated such that if $\Delta x_i^\mu(k) \neq 0$, then $\Delta x_i^\mu(k) = 0$ for all $j \in \Psi_{ij}$; 2) $|w_i| < 2|w_{ij}|$ for at least one $j \neq i$; 3) $|w_i| > |w_{ij}|$ for all $j \neq i$; and 4) $\sigma = 0$ and take $\lambda = 0$. Then $\eta_{EHE} = 1$, if $\eta_{PK} = \eta_{ZCVJ} < 1$.

To end this section, we give the following proposition.

Proposition 1: In general, $\eta_{EHE} > \eta_{PK} = \eta_{ZCVJ}$.

If there were no product term in $Q_i^-(l, k)$ and $Q_i^+(l, k)$, then $Q_i^-(l, k) = F_i^-(l, k)$ and $Q_i^+(l, k) = F_i^+(l, k)$ and we would
have $\eta_{EHE} = \eta_i$ that is given in the parenthesis of (24). Meanwhile, the product term in $Q^-_i(l, k)$ is a decreasing function of $l$ while the product term in $Q^+_i(l, k)$ is an increasing function of $l$. This means that relatively to say, $Q^-_i(l, k)$ is a faster decreasing function of $l$ than $F^-_i(l, k)$ while $Q^+_i(l, k)$ is a faster increasing function of $l$ than $F^+_i(l, k)$. Due to Property 1, this implies that $\eta_{EHE} > \eta_i$, and therefore Proposition 1 is true.

V. Correct Transition Probability of SA Algorithm

Similarly to the PK and ZCVJ algorithms, the SA algorithm randomly chooses a neuron to update. However, the SA algorithm allows energy increase with certain probability.

Theorem 3: For the SA algorithm,

$$R^-_i(l, k) \equiv \Pr(\Delta x_i^c(k) = -1 | \Delta x_i^c(k) = l) = \int_0^1 \Phi \left( \frac{m_{j_i} + t + T \ln \gamma}{\sigma_{j_i}} \right) d\gamma \tag{28}$$

$$R^+_i(l, k) \equiv \Pr(\Delta x_i^c(k) = 1 | \Delta x_i^c(k) = l) = \int_0^1 \Phi \left( \frac{m_{j_i} - t - T \ln \gamma}{\sigma_{j_i}} \right) d\gamma \tag{29}$$

where

$$\eta_{SA} = \sum_{i=1}^{n} \left[ \sum_{l=-L+1}^{-1} \left( \sum_{k=-L+1}^{L-1} R^-_i(l, k) p_i^{(l)}(k) \right) u_i^c(k) \right. \left. + \sum_{l=L+1}^{L-1} \left( \sum_{k=L+1}^{L-1} R^+_i(l, k) p_i^{(l)}(k) \right) u_i^c(k) \right] P_i(L(k) = \{i\}) \tag{30}$$

It is interesting to compare the SA algorithm with the PK and ZCVJ algorithms in terms of the correct transition probability. Corollaries 1–6, which are derived from the PK and ZCVJ algorithms, are also applicable to the SA algorithm. Furthermore, the following corollaries are yielded from Theorem 3.

Corollary 18: For all $l$, $R^-_i(l, k) \geq F^-_i(l, k)$ and $R^+_i(l, k) \geq F^+_i(l, k)$ where equalities hold if and only if $T = 0$.

Corollary 19: When the temperature becomes $T = 0$, $\eta_{SA} = \eta_{PK} = \eta_{ZCVJ}$.

As expected, Corollaries 18 and 19 mean that the PK and ZCVJ algorithms are special instances of the SA algorithm as $T = 0$.

Corollary 20: As $T \to \infty$, the SA algorithm has no restoration capability. If further the distribution of $\Delta x_i^c(k) = l$ is symmetric to zero, $\eta_{EHE} \leq 1/2$ where the equality holds if and only if $P_i[C(k) = 0] = 1$.

Corollary 21: $\eta_{SA}$ monotonically decreases with $T$.

VI. Numerical Evaluations

In this section, we demonstrate numerical evaluations of the correct transition probabilities of the PK, ZCVJ, SA and EHE algorithms. It will be shown that $\eta_{EHE} > \eta_{PK} = \eta_{ZCVJ} > \eta_{SA}$.

The conditions in all numerical evaluations are the same as those used in the simulations of [16]: 1) the original image $x^0$ is the $128 \times 128$ Lena image; and so $n = 16384$ and $L = 256$; 2) the blurring matrix $A$ represents a convolution of uniform blur over a window of size $(2c - 1) \times (2c - 1)$, i.e., $f_A(i, j) = 1/(2c-1)^2$ for $|i| < c$ and $|j| < c$ and $f_A(i, j) = 0$ elsewhere. The larger is $c$, the severer is the blurring; 3) the difference matrix $D$ denotes a convolution

$$f_D(i, j) = \frac{1}{\sigma_{x}} \begin{bmatrix} 0.7 & 1 & 0.7 \\ 1 & -6.8 & 1 \\ 0.7 & 1 & 0.7 \end{bmatrix}$$

4) In order to evaluate the correct transition probabilities for the image data used in the simulation of [16], we consider $\Delta x_i^c(0) = x^0 - x(0)$ where $x^0$ is the Lena image and $x(0) = y$ is the blurred noisy image with $c = 3$ and $SNR = 25$ dB. This is close to the initial images used in the simulations of [16]. As estimated, the distribution of $\Delta x_i^c(k) = l$ is symmetric to zero and $C(k) \neq 0$. In this case, $u_i^c = u_i = 1/2$. All other parameters of statistics in the evaluation of the correct transition probability are estimated from the practical images in corresponding conditions of $A$, $D$, noise, and $\lambda$; 5) The signal-to-noise ratio (SNR) of the blurred image is defined as $SNR = 10 \log_{10} (\sigma^2 / \sigma_{x}^2)$ where $\sigma^2$ is the pixel variance estimated from the noise-free blurred image $Ax^0$, and $\sigma_{x}^2$ is the pixel variance of the noise added to the blurred image. Given the SNR, we can know $\sigma^2$.

A. Correct Transition Probabilities Versus Severity of Blurring

Fig. 1 shows the correct transition probability versus the severity of blurring with $SNR = 25$ dB and $\lambda = 0.1$. As shown in Fig. 1, for all severity of blurring, $\eta_{EHE} > \eta_{PK} = \eta_{ZCVJ} > \eta_{SA}$ and $\eta_{EHE}$ is much higher than $\eta_{PK}$, $\eta_{ZCVJ}$, and $\eta_{SA}$. Fig. 1 also shows that all the correct transition probabilities are monotonically decreasing functions of $c$. This means that the more severely an image is blurred, the more difficult it is for these algorithms to restore. Corollaries 2, 9, and 10 and Propositions 1 and 2 are verified in Fig. 1.

B. Correct Transition Probabilities Versus SNR

Fig. 2 shows the correct transition probability versus SNR with $c = 3$ and $\lambda = 0.1$. For all SNR, $\eta_{EHE} > \eta_{PK} = \eta_{ZCVJ} > \eta_{SA}$, and $\eta_{EHE}$ is much higher than $\eta_{PK}$, $\eta_{ZCVJ}$, and $\eta_{SA}$ except for the extreme severity of noise. This confirms the simulation results in [16] that the correct transition rates of the EHE algorithms are much higher than those of the PK, ZCVJ, and SA algorithms. The less severe the noise is, the larger the difference is between $\eta_{EHE}$ and $\eta_{PK}$, $\eta_{ZCVJ}$, and $\eta_{SA}$. As shown in Fig. 2, all the correct transition probabilities are monotonically increasing functions of SNR. When the noise is severe, all these
For all severity of blurring, $\eta_{EH} > \eta_{PK} = \eta_{ZCVJ} > \eta_{SA}$, and $\eta_{EH}$ is much higher than $\eta_{PK}$, $\eta_{ZCVJ}$, and $\eta_{SA}$. All the correct transition probabilities monotonically decrease with the severity of blurring.

For all SNR, $\eta_{EH} > \eta_{PK} = \eta_{ZCVJ} > \eta_{SA}$, and $\eta_{EH}$ is much higher than $\eta_{PK}$, $\eta_{ZCVJ}$, and $\eta_{SA}$ except for the extremely large $\lambda$. The smaller $\lambda$ is, the larger the difference between them is. All the correct transition probabilities monotonically decrease with $\lambda$.

For all $\lambda$, $\eta_{EH} > \eta_{PK} = \eta_{ZCVJ} > \eta_{SA}$, and $\eta_{EH}$ is much higher than $\eta_{PK}$, $\eta_{ZCVJ}$, and $\eta_{SA}$ except for the extremely large $\lambda$. The smaller $\lambda$ is, the larger the difference between them is. All the correct transition probabilities monotonically decrease with $\lambda$.

For all $\lambda$, $\eta_{EH} > \eta_{PK} = \eta_{ZCVJ} > \eta_{SA}$, and $\eta_{EH}$ is much higher than $\eta_{PK}$, $\eta_{ZCVJ}$, and $\eta_{SA}$ except for the extremely large $\lambda$. The smaller $\lambda$ is, the larger the difference between them is. All the correct transition probabilities monotonically decrease with $\lambda$.

Correct transition probabilities are slightly smaller than 0.5 due to $\mathcal{C}(k) \neq 0$. Corollaries 4 and 12 and Propositions 1 and 2 are verified by Fig. 2.

**C. Correct Transition Probability Versus $\lambda$**

Fig. 3 shows the correct transition probability versus $\lambda$ with $c = 2$ and SNR = 25 dB. For all $\lambda$, $\eta_{EH} > \eta_{PK} = \eta_{ZCVJ} > \eta_{SA}$, and $\eta_{EH}$ is much higher than $\eta_{PK}$, $\eta_{ZCVJ}$, and $\eta_{SA}$ except for the extremely large $\lambda$. The smaller $\lambda$ is, the larger is the difference between them. When the regularization term is extremely large, all the correct transition probabilities are slightly lower than 0.5. All the correct transition probabilities are monotonically decreasing functions of $\lambda$. In contrast, as observed in simulation of [16], when the noise is strong, the correct transition rates of these algorithms are convex down functions of $\lambda$ instead of monotonic functions.

This difference between the correct transition probability and correct transition rate can be explained as follows. First, we note that the correct transition probability measures the performance at a particular time instance. In terms of signal detection theory, given statistics (including the currently restored image), any other term unrelated to the original signal to be detected is an interference to the signal. Hence, the additional regularization term decreases the correct detection probability of the signal. Second, the correct transition rate estimated from a simulation is an estimate of the averaged correct transition probability over all transitions in a long restoration processing. The correct transition rate and the correct transition probability have different statistical meanings. Consider the case of strong noise. If $\lambda = 0$, a neuron-state transition is wrong with a large probability equal to the complement of the correct transition probability, $1 - \eta$. The error produced by the wrong transition propagates and influences the later transitions of other neurons in the neighborhood. The finally restored image is rough and noisy. If $\lambda \neq 0$ is properly set up, because of the regularization term, the neural transition is biased to producing a pixel value closer to the mean of the currently restored image in the neighborhood of the pixel. This bias is toward producing a smoother image and suppresses the influence of the noise and the propagation of errors. Hence, although the nonzero regularization term decreases the correct transition probability at present time, it may increase the correct transition probabilities in later transitions. In other words, as a whole, the nonzero value $\lambda$ may increase the average correct transition probability over all neural transitions in a long run of image restoration. The correct transition rates estimated in the simulation of [16] are estimates of the averaged correct transition probabilities, which reflect on this dynamic effect of the regularization term in a long run of image restoration.
Hence, it is understandable that the correct transition probability monotonically decreases with $\lambda$ while the correct transition rate is convex down with respect to $T$. Corollaries 5 and 13 and Propositions 1 and 2 are verified by Fig. 3.

**D. Correct Transition Probabilities of SA Algorithm Versus $T$**

Fig. 4 shows the correct transition probability of the SA algorithm versus temperature $T$ with $c = 2$, SNR = 40 dB, and $\lambda = 0.1$. As expected, $\eta_{SA}$ monotonically decreases with $T$. When $T$ is large enough, $\eta_{SA} < 0.5$. By comparing with Fig. 1, we can see that $\eta_{SA} = \eta_{PK} = \eta_{ZCVJ}$ when $T = 0$. Corollaries 19, 20, and 21 are verified by Fig. 4.

**VII. CONCLUSIONS**

While the popularly used SNR of finally recovered images can only measure the performance of an iterative algorithm as a whole in a long recovering processing of an image, the correct transition probability as a new performance measure can reveal insight of performance of an iterative algorithm because it measures the performance of the algorithm at every iteration. The higher the correct transition probability is, the faster an algorithm converges to an image. If the correct transition probability is one for all iterations, the algorithm can recover an error-free image. If the correct transition probability is lower than 0.5 in all iterations, the algorithm has no capability to improve an blurred image.

The analysis of the PK, ZCVJ, EHE, and SA algorithms in terms of the correct transition probability shows that the EHE algorithm has much better performance at each transition in all conditions of severity of blurring, severity of noise, and degree of regularization. This confirms the results in many previous simulations that all versions of EHE algorithms can converge to higher-SNR images with much fewer iterations and better visual quality, and have much higher correct transition rates than other algorithms. The analysis also shows that the PK and ZCVJ algorithms have higher correct transition probability than the SA algorithm in all conditions, and are instance of the SA algorithm when the temperature is zero. Intuitively confirming the phenomena observed in previous simulations, it can be seen through the analysis of the correct transition probability that the more severely an image is blurred, the more hardly it is restored; the stronger the noise is, the more hardly the image is restored. In return, the correct transition probability is shown to be a rational performance measure.

Previous simulations demonstrated that the performance of these four algorithms in terms of the SNR and correct transition rate is convex down with respect to the degree of the regularization if the noise is strong. However, the correct transition probability studied in this paper is a monotonically decreasing function of the degree of regularization. This is due to the fact that the SNR and the correct transition rate account for the performance of the whole recovering processing while the correct transition probability measures only the performance at a particular time instance. The effect of the regularization on the finally recovered image is a dynamic process carried through the long processing of the image recovering iterations.

The correct transition probability is applicable for performance evaluation to other iterative, neural or nonneural, algorithms for image restoration and reconstruction, and for other tasks of signal and image processing with the similar problem formulation of this paper, e.g., CDMA multiuser detection.

It is an interesting problem to seek better energy functions by means of the analysis presented in this paper.

**APPENDIX**

**Proof of Property 4:**

If $\Pr[|\delta x_i(k)| = 0] = 1 = \alpha_i \neq 0$ and $\Pr[|\delta x_k| = 1] = 1 = \beta_k \neq 0$ are constants independent of $k$, the transitions are independent of the original image. Hence, the algorithm has no restoration capability. In this case, $\eta_k$ in (14) becomes

\begin{equation}
\eta_k = \frac{1}{L+1} \sum_{i=1}^{L+1} p_{i}^{(k)} \left( \sum_{i=1}^{L+1} u_i^{(k)} - 1 \right) = \frac{1}{L+1} \sum_{i=1}^{L+1} p_{i}^{(k)} \left( \sum_{i=1}^{L+1} u_i^{(k)} \right).
\end{equation}

If further the distribution of $\delta x_k = I$ is symmetric to zero,

\begin{equation}
\frac{\sum_{i=1}^{L+1} p_{i}^{(k)}}{2} \left( \sum_{i=1}^{L+1} u_i^{(k)} \right) \leq \frac{1}{2}
\end{equation}

and

\begin{equation}
\frac{\sum_{i=1}^{L+1} p_{i}^{(k)}}{2} \left( \sum_{i=1}^{L+1} u_i^{(k)} \right) \leq \frac{1}{2}
\end{equation}

where the equalities hold if and only if $\Pr[C(k) = \emptyset] = 1$. Hence, $\eta_k \leq 1/2 u_0^{+}(k) + 1/2 u_0^{+}(k) = 1$ due to $u_0^{+}(k) + u_0^{+}(k) = 1$. (Q.E.D.)
Proof of Theorem 1: According to the PK (ZCVJ) algorithm and by means of Property 6, we obtain
\[ F_i^-(l, k) = \Pr(\Delta x_i(k) = -1|\Delta x_i^2(k) = l) \]
\[ = \Pr(X_{ji} < -t) \]
\[ = \Phi \left( \frac{-m_{ij} + t}{\sigma_{q_i}} \right) . \]  
(A4)

Similarly,
\[ F_i^+(l, k) = \Pr(\Delta x_i(k) = 1|\Delta x_i^2(k) = l) \]
\[ = \Pr(X_{ji} > t) \]
\[ = \Phi \left( \frac{m_{ij} - t}{\sigma_{q_i}} \right) . \]  
(A5)

Due to the definition of the correct transition probability, (22) and (23) result in (24). (Q.E.D.)

Proof of Corollary 1: In terms of Property 6 and (8),
\[ F_i^-(l, k) = \Phi \left( \frac{|w_{ij}| (l + \frac{1}{2}) - m_{\Delta x^o} \sum_{j \in \Psi_{uw}, j \neq i} w_{ij} + \lambda m_{gx_i}}{\sqrt{\sigma_{\Delta x^o}^2 \sum_{j \in \Psi_{uw}, j \neq i} w_{ij}^2 + \lambda^2 \sigma_{gx_i}^2 + |w_{ii}|^2}} \right) . \]  
(A6)

and
\[ F_i^+(l, k) = \Phi \left( \frac{|w_{ij}| (l - \frac{1}{2}) - m_{\Delta x^o} \sum_{j \in \Psi_{uw}, j \neq i} w_{ij} + \lambda m_{gx_i}}{\sqrt{\sigma_{\Delta x^o}^2 \sum_{j \in \Psi_{uw}, j \neq i} w_{ij}^2 + \lambda^2 \sigma_{gx_i}^2 + |w_{ii}|^2}} \right) . \]  
(A7)

Since \( \Phi(x) \) is a monotonically increasing function of \( x \), (A6) and (A7) imply that \( F_i^-(l, k) \) and \( F_i^+(l, k) \) monotonically decrease and increase with \( l \), respectively. (Q.E.D.)

Proof of Corollary 2: Since \( m_{\Delta x^o} = 0 \) and \( \lambda = 0 \), from (A6) and (A7), we obtain
\[ F_i^-(l, k) = \Phi \left( \frac{l + \frac{1}{2}}{\sum_{j \in \Psi_{uw}, j \neq i} |w_{ij}|^2 / \sum_{j \in \Psi_{uw}, j \neq i} w_{ij}^2 + \sigma^2 / |w_{ii}|} \right) \]  
(A8)

and
\[ F_i^+(l, k) = \Phi \left( \frac{l - \frac{1}{2}}{\sum_{j \in \Psi_{uw}, j \neq i} w_{ij}^2 / \sum_{j \in \Psi_{uw}, j \neq i} w_{ij}^2 + \sigma^2 / |w_{ii}|} \right) . \]  
(A9)

Clearly, \( F_i^-(l, k) \) for \( l < 0 \) monotonically increases with \( |w_{ij}| \) and \( F_i^+(l, k) \) for \( l > 0 \) monotonically decreases with \( |w_{ij}| \). Hence, \( \eta_{PK} \) and \( \eta_{ZCVJ} \) monotonically increase with \( |w_{ij}| \). It is also obvious that \( F_i^-(l, k) \) for \( l < 0 \) and \( F_i^+(l, k) \) for \( l > 0 \) monotonically decrease with \( \sum_{j \in \Psi_{uw}, j \neq i} w_{ij}^2 \); \( F_i^-(l, k) \) for \( l < 0 \) and \( F_i^+(l, k) \) for \( l > 0 \) monotonically increase with \( \sum_{j \in \Psi_{uw}, j \neq i} |w_{ij}|^2 \). Hence, \( \eta_{PK} \) and \( \eta_{ZCVJ} \) monotonically decrease with \( \sum_{j \in \Psi_{uw}, j \neq i} |w_{ij}|^2 \).

Proof of Corollary 3: Since \( \sum_{j \in \Psi_{uw}, j \neq i} w_{ij} \leq 0 \), due to (A6) and (A7), \( F_i^-(l, k) \) and \( F_i^+(l, k) \) are monotonically decreasing and increasing functions of \( m_{\Delta x^o} \), respectively.

Proof of Corollary 4: In terms of (A6) and (A7), as \( \sigma \rightarrow \infty \), \( F_i^-(l, k) = 1/2 \) and \( F_i^+(l, k) = 1/2 \). By means of Property 4, all results follow.

Proof of Corollary 5: As \( \lambda \rightarrow \infty \), in probabilistic sense, (16) yields
\[ X_{q_i} = \lambda \sum_{j \in \Psi_{gw}} g_{kj} x_j(k) \]  
(A10)

which is independent of \( l \). This implies that \( F_i^-(l, k) \) and \( F_i^+(l, k) \) are constants independent of \( l \). In terms of Property 4, we obtain 1) and 2). If \( x_i(k) \) is constant for all \( i \in \{1, \cdots, n\} \), \( X_{q_i} = 0 \) due to \( \sum_{j \in \Psi_{gw}} g_{ij} = 0 \). Hence, \( F_i^-(l, k) = 0 \) and \( F_i^+(l, k) = 0 \) by means of Property 5, \( \mathbf{x}(k) \) is a fixed point of the PK and ZCVJ algorithms with probability one. (Q.E.D.)

Proof of Corollary 6: Since \( \mathbf{A} = \mathbf{I}, \mathbf{W} = \mathbf{I} \) and so \( w_{ii} = -1 \) and \( w_{ij} = 0 \) for \( \forall i \neq j \). Meanwhile, \( \sigma = 0 \) and \( \lambda = 0 \). (A6) and (A7) yield
\[ F_i^-(l, k) = \frac{1}{2} (1 - \text{sgn}(l + \frac{1}{2})) \]  
(A11)

and
\[ F_i^+(l, k) = \frac{1}{2} (1 + \text{sgn}(l - \frac{1}{2})) , \]  
(A12)

respectively. In terms of (24), this implies \( \eta_{PK} = \eta_{ZCVJ} = 1 \).

Proof of Corollary 7: Since \( \alpha_{ij} = (A_{ij})_{ij} = 1/n \) for all \( i, j \in \{1, \cdots, n\} \), \( w_{ij} = -1/n \) and \( \Psi_{uw} = \{1, \cdots, n\} \). Because of the uniform distribution of \( \Delta x_i^2(k) \) on \\{-L + 1, \cdots, -1, 1, \cdots, L - 1\}, \( p^{(0)}(k) = 1/(2L - 2) \), \( m_{\Delta x^o} = 0 \) and \( \sigma_{\Delta x^o} = \sqrt{L(2L - 1)/6} \). Since \( \sigma = 0 \) and \( \lambda = 0 \), \( m_{q_i} = 1/n \) and \( \sigma_{q_i}^2 = \sigma_{\Delta x}^2 |\Psi_{uw}| - 1) / n^2 = \sigma_{q_i}^2(k) = L(2L - 1)(n - 1)/(6n^2) \). From (A6) and (A7),
\[ F_i^-(l, k) = \Phi \left( -\frac{(l + \frac{1}{2}) \sqrt{6}}{L(2L - 1)(n - 1)} \right) \]  
(A13)

and
\[ F_i^+(l, k) = \Phi \left( \frac{(l - \frac{1}{2}) \sqrt{6}}{L(2L - 1)(n - 1)} \right) . \]  
(A14)

Since \( |l| \leq L - 1 \) and \( L \ll n \), \( F_i^-(l, k) \approx 1/2 \) and \( F_i^+(l, k) \approx 1/2 \). These produce \( \sum_{l=-L}^{-1} F_i^-(l, k) \approx \)
1/4 and \( \sum_{l=L+1}^{L+n} F_{-\ell}(l, k) p_{\ell}^{(k)}(k) \approx 1/2 \). Similarly, we can obtain \( \sum_{l=L+1}^{L+n} F_{\ell}(l, k) p_{\ell}^{(k)}(k) \approx 1/4 \) and \( \sum_{l=L+1}^{L+n} F_{+\ell}(l, k) p_{\ell}^{(k)}(k) \approx 1/2 \). Hence, \( \eta_{\text{FC}} = \eta_{\text{CVJ}} \approx 1/2 \).

Proof of Theorem 2: Consider \( i \in L(k) \). In terms of the EHE algorithm, \( i \in J(k) \) and \( |h_i(k)| = \max_{j \in J(k)} |h_j(k)| \). Because of assumption that \( 0 < x_i(k) - L + 1 \) for \( \forall i \in \{1, \ldots, n\} \), \( |h_i(k)| = \max_{j \in \{1, \ldots, n\}} |h_j(k)| \). By means of Property 6

\[
Q_{\tau}^+(l, k) = \Pr(\Delta x_i(k) = 1 | \Delta x_{\ell}^{(k)}(l) = 1) = \Pr(h_i(k) > t, h_{j}(k) < h_{j}(k) < h_{k}(k)) \quad \text{for} \quad j \neq i, j \in \{1, \ldots, n\} \mid \Delta x_{\ell}^{(k)}(l) = 1 = \Pr(X_{d_i} > t, X_{d_i} < X_{d_i} < X_{d_i}) \quad \text{for} \quad j \neq i, j \in \{1, \ldots, n\} = \frac{1}{\sqrt{2\pi} \sigma_{d_i}} \int_{-\infty}^{t} \exp\left(-\frac{(x_i - m_{d_i})^2}{2 \sigma_{d_i}^2}\right) \prod_{j=1, j \neq i}^{n} \exp\left(-\frac{(x_j - m_{d_j})^2}{2 \sigma_{d_j}^2}\right) dx_i \times \prod_{j=1, j \neq i}^{n} \Phi\left(-\frac{x_i - m_{d_i}}{\sigma_{d_i}}\right) - \Phi\left(\frac{x_i - m_{d_i}}{\sigma_{d_i}}\right) \quad \text{(A15)}
\]

which produces (25). Similarly

\[
Q_{\tau}^+(l, k) = \Pr(\Delta x_i(k) = -1 | \Delta x_{\ell}^{(k)}(l) = 1) = \Pr(h_i(k) < -t, -h_{j}(k) < h_{j}(k) < h_{k}(k)) \quad \text{for} \quad j \neq i, j \in \{1, \ldots, n\} \mid \Delta x_{\ell}^{(k)}(l) = 1 = \Pr(X_{d_i} < t, X_{d_i} > X_{d_i} < X_{d_i}) \quad \text{for} \quad j \neq i, j \in \{1, \ldots, n\} = \frac{1}{\sqrt{2\pi} \sigma_{d_i}} \int_{t}^{\infty} \exp\left(-\frac{(x_i - m_{d_i})^2}{2 \sigma_{d_i}^2}\right) \prod_{j=1, j \neq i}^{n} \exp\left(-\frac{(x_j - m_{d_j})^2}{2 \sigma_{d_j}^2}\right) dx_i \times \prod_{j=1, j \neq i}^{n} \Phi\left(-\frac{x_i - m_{d_i}}{\sigma_{d_i}}\right) - \Phi\left(\frac{x_i - m_{d_i}}{\sigma_{d_i}}\right) \quad \text{(A16)}
\]

which yields (26). According to the EHE criterion, \( \Pr(L(k) = \{i\}) = 1 \) if \( |h_i(k)| = \max_{j \in J(k)} |h_j(k)| \) and \( \Pr(L(k) = \{i\}) = 0 \) if \( |h_i(k)| < \max_{j \in J(k)} |h_j(k)| \). Hence, \( \eta_{\text{EHE}} = \eta_{\text{EHE}} \) that is given by (27).

Proof of Corollary 8: From (A15), we obtain

\[
Q_{\tau}^+(l, k) = \frac{1}{\sqrt{2\pi} \sigma_{d_i}} \int_{-\infty}^{m_{d_i}} \exp\left(-\frac{(x - t)^2}{2 \sigma_{d_i}^2}\right) \times \prod_{j=1, j \neq i}^{n} \left[ \Phi\left(\frac{-x_i - m_{d_i} - m_{d_j} + t}{\sigma_{d_i}}\right) - \Phi\left(\frac{x_i + m_{d_i} - m_{d_j} - t}{\sigma_{d_i}}\right) \right] dx_i. \quad \text{(A17)}
\]

Take derivative of (A17) with respect to \( l \), we have

\[
dQ_{\tau}^+(l, k)/dl = \sum_{m=1}^{n} t_m
\]

where

\[
t_0 = \frac{1}{\sqrt{2\pi} \sigma_{d_i}} \exp\left(-\frac{(m_{d_i} - t)^2}{2 \sigma_{d_i}^2}\right)
\]

and

\[
t_m = \frac{1}{\sqrt{2\pi} \sigma_{d_i}} \int_{-\infty}^{m_{d_i}} \exp\left(-\frac{(x - t)^2}{2 \sigma_{d_i}^2}\right) \times \prod_{j=1, j \neq i}^{n} \left[ \Phi\left(\frac{-x_i - m_{d_i} - m_{d_j} + t}{\sigma_{d_i}}\right) - \Phi\left(\frac{x_i + m_{d_i} - m_{d_j} - t}{\sigma_{d_i}}\right) \right] \frac{1}{\sqrt{2\pi} \sigma_{d_i}} \left[ \exp\left(-\frac{(x - m_{d_i})^2}{2 \sigma_{d_i}^2}\right) \frac{w_{ii} + w_{mi}}{\sigma_{m_i}} + \exp\left(-\frac{(x_{d_i} - m_{d_i} - m_{d_i})^2}{2 \sigma_{d_i}^2}\right) \frac{w_{ii} - w_{mi}}{\sigma_{m_i}} \right] dx_i. \quad \text{(A19)}
\]

Since \( w_{ii} < 0 \) and \( |w_{ii}| > |w_{mi}| \) due to (10), \( t_0 < 0 \) and \( t_m < 0 \). Hence, \( dQ_{\tau}^+(l, k)/dl = \sum_{m=0}^{n} t_m < 0 \). Similarly, we can obtain \( dQ_{\tau}^+(l, k)/dl > 0 \). (Q.E.D.)

Proof of Corollary 12: From (25) and (26), as \( \sigma \rightarrow \infty \).

\[
Q_{\tau}^+(l, k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \exp\left(-\frac{x^2}{2}\right) \left[ \Phi(-x) - \Phi(x) \right]^{n-1} dx = \frac{1}{2n}, \quad \text{(A20)}
\]

and

\[
Q_{\tau}^+(l, k) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \exp\left(-\frac{x^2}{2}\right) \left[ \Phi(x) - \Phi(-x) \right]^{n-1} dx = \frac{1}{2n}. \quad \text{(A21)}
\]

In terms of Property 4, all results follow. (Q.E.D.)
Proof of Corollary 14: If $A = I$, $\sigma = 0$ and take $\lambda = 0$, then $t = 1/2$, $X_{q_k} = l$ and $X_{\lambda_k} = \Delta x_j(k)$. Hence,

$$Q^+_i(l, k) = \Pr(X_{\lambda_i} < -t, X_{q_i} < X_{ji} < X_{j_i} < X_{q_i};$$

for $j \neq i, j \in \{1, \ldots, n\} \}
\Pr(l < -\frac{1}{2} l < \Delta x_j^2(k) < -l$
\text{for } j \neq i, j \in \{1, \ldots, n\} 
= \Pr(l < 0, |\Delta x_j^2(k)| < |l| \text{ for all } j \neq i, j \in \{1, \ldots, n\} 
= \begin{cases} 
1, & \text{if } l < 0 \text{ and } |\Delta x_j^2(k)| < |l| \text{ for all } j \neq i, \\
0, & \text{otherwise}. 
\end{cases} 
(A22)

In the same way, we obtain

$$Q^+_i(l, k) = \begin{cases} 
1, & \text{if } l > 0 \text{ and } |\Delta x_j^2(k)| < |l| \text{ for all } j \neq i, \\
0, & \text{otherwise}. 
\end{cases} 
(A23)

This implies $\eta_{EHE} = 1$.

Proof of Corollary 15: Since $a_{ij} = (A)_{ij} = 1/n$ for all $i, j \in \{1, \ldots, n\}$ and $w_{ji} = -1/n$ and $\Psi_{w_i} = \{1, \ldots, n\}$. From (17),

$$X_{m_{ji}} = \frac{l}{n} + \frac{1}{n} \sum_{j=1}^{n} \Delta x_j^2(k) 
(A24)

which means that $X_{m_{ji}}$'s are identical. Hence, $Q^+_i(l, k) = E^+(l, l, 1)$ and $Q^+_i(l, k) = E^+(l, l, 1)$. Corollary 7 is applicable and all results follow.

Proof of Corollary 16: From (A15) and (A4),

$$Q^+_i(l, k) = \Pr(X_{\lambda_i} < -t, X_{q_i} < X_{ji} < X_{j_i} < X_{q_i};$$

for $j \neq i, j \in \{1, \ldots, n\} \}
\leq \Pr(X_{\lambda_i} < -t)
= E^+_i(l, k). 
(A25)

Similarly, from (A16) and (A5), we obtain $Q^+_i(l, k)E^+_i(l, k)$. (Q.E.D.)

Proof of Corollary 17: The image $\mathbf{x}(k)$ may contain many isolated error pixels. Consider any such an error pixel $x_j(k)$ and its neighborhood such that $\Delta x_j^2(k) \neq 0$ but $\Delta x_j^2(k) = 0$ for $\forall j \in \Psi_{w_i}$. Clearly, we have $h_j(k) = -\Delta x_j^2(k)$ for $j \in \Psi_{w_i}$.

a) For the PK and ZCVJ algorithms, since there exists at least one $j \in \Psi_{w_i}$ such that $w_{ji} = 2[w_{ji}]$ or equivalently $h_j(k) = -w_{ji} \Delta x_j^2(k) \geq |w_{ji}| > 1/2|w_{ji}| = t$, either 1) $E^{-}_i(0, k) > 0$ for $h_j(k) < -t$ or 2) $E^+_i(0, k) > 0$ for $h_j(k) > t$. This implies that at least one wrong transition occurs with a probability larger than zero. Hence, $\eta_{PK} = \eta_{ZCVJ} < 1$.

b) For the EHE algorithm, since $w_{ji} > |w_{ji}|$ for $\forall j \neq i, |h_j(k)| = -w_{ji} \Delta x_j^2(k) < |w_{ji} \Delta x_j^2(k)| = |h_j(k)|$.

This implies that in the $i$th pixel's neighborhood $\Psi_{w_i}$, it is impossible for the EHE algorithm to choose any $j$ other than $i$ to form $L(k)$, i.e., $Pr(L(k) = \{j\}) = 0$ for $j \neq i$. Only $i$ is possible to be chosen to form $L(k) = \{i\}$, i.e., $Pr(L(k) = \{i\}) > 0$. The same result is true in the neighborhoods of any other error pixels. Due to the fact $w_{ji} < 0$, $h_j(k) = |w_{ji}| \Delta x_j^2(k)$. So according to the EHE algorithm, once $i$ is chosen, $\Delta x_i^2(k) = -1$ if $\Delta x_i^2(k) = l < 0$ or $\Delta x_i^2(k) = 1$ if $\Delta x_i^2(k) = l > 0$. Hence, $Q^+_i(l, k) = 0$ for all $l \geq 0$ and $Q^+_i(l, k) = 0$ for all $l \leq 0$. This implies $\eta_{EHE} = 1$.

Proof of Theorem 3: For $\Delta x_i^2(k) = -1$, according to (10), $\Delta E_i(k) = h_i(k) + t \cdot p_i^-(k) = \exp[-\Delta E_i(k)/T] > \gamma$ is equivalent to $h_i(k) < -t - T \ln \gamma$, and hence

$$R_i^-(l, k) = \Pr(\Delta x_i^2(k) = -1 | \Delta x_i^2(k) = l)$$

$$= \Pr(i \in L(k), h_i(k) < -t - T \ln \gamma | \Delta x_i^2(k) = l)$$

$$= \Pr(X_{\lambda_i} < -t - T \ln \gamma)$$

$$= \frac{1}{\sigma_{\lambda_i}} \int_{0}^{\gamma} \Phi\left(\frac{-m_{\lambda_i} + t - T \ln \gamma}{\sigma_{\lambda_i}}\right) d\gamma. 
(A26)

For $\Delta x_i^2(k) = 1$, $\Delta E_i(k) = h_i(k) + t \cdot p_i^+(k) = \exp[-\Delta E_i(k)/T] > \gamma$ is equivalent to $h_i(k) > t + T \ln \gamma$, thus yielding

$$R_i^+(l, k) = \Pr(\Delta x_i^2(k) = 1 | \Delta x_i^2(k) = l)$$

$$= \Pr(i \in L(k), h_i(k) > t + T \ln \gamma | \Delta x_i^2(k) = l)$$

$$= \Pr(X_{\lambda_i} > t + T \ln \gamma)$$

$$= \int_{0}^{\gamma} \Phi\left(\frac{m_{\lambda_i} - t - T \ln \gamma}{\sigma_{\lambda_i}}\right) d\gamma. 
(A27)

Due to the definition of the correct transition probability, we obtain (30). (Q.E.D.)

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their careful review and constructive comments.

REFERENCES


Yi Sun (M’92) received the B.S. and M.S. degrees in electrical engineering from the Shanghai Jiao Tong University, Shanghai, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 1997. From 1985 to 1993, he was a Lecturer with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University. In the Summer of 1993, he was a Visiting Scientist with the Department of Mechanical Engineering, Concordia University, Montreal, P.Q., Canada. From March to September 1997, he was a Postdoctoral Research Fellow with the Radiology Department, University of Utah, Salt Lake City, where he studied MRI imaging. From October 1997 to August 1998, he was a Postdoctoral Research Associate working on wireless communications in the Department of Electrical and Systems Engineering, University of Connecticut, Storrs. Since September 1998, he has been an Assistant Professor with the Department of Electrical Engineering, City College, City University of New York. His research interests are in the areas of wireless communications (with focus on CDMA multiuser detection, slotted CDMA networks, channel equalization and sequence detection, and multicarrier CDMA systems), image and signal processing, medical imaging, and neural networks.