ISRN SICS-R--95/09--SE

Best Probability of Activation and Performance Comparisons for several Designs of Sparse Distributed Memory

by

Jan Kristoferson

SICS Research Report
R95:09
ISSN 0283-3638

Report R95:09

ISRN : SICS-R--95/09-SE
ISSN : 0283-3638

Best Probability of Activation and Performance Comparisons for Several Designs of Sparse Distributed Memory

Jan Kristoferson

RWCP[1] Neuro SICS[2] Laboratory
Box 1263, S–164 28 KISTA, SWEDEN
Phone +46 8 752 15 00 Fax +46 8 751 72 30
E-mail: janke@sics.se

November 30, 1995

Abstract

The optimal probability of activation and the corresponding performance is studied for three designs of Sparse Distributed Memory, namely, Kanerva's original design, Jaeckel's selected-coordinates design and Karlsson's modification of Jaeckel's design. We will assume that the hard locations (in Karlsson's case, the masks), the storage addresses and the stored data are randomly chosen, and we will consider different levels of random noise in the reading address.

Keywords: Sparse Distributed Memory, Probability of Activation, Performance

Contents

---

[1] Real World Computing Partnership
[2] Swedish Institute of Computer Science

A *hard location* $\mathbf{h}$ is a location existing in the memory; let M be the total number of hard locations.

The *content* of a hard location is a U-dimensional vector of integers. The coordinates of the content vectors are called *positions*.

In Kanerva's design an address $\mathbf{X}$ *activates* a hard location $\mathbf{h}$ if the Hamming distance $d(\mathbf{X}, \mathbf{h})$ is at most equal to the *activation radius* R.

In the other two designs an address $\mathbf{X}$ *activates* a hard location $\mathbf{h}$ if all bits in $\mathbf{X}$ lying in $\mathbf{h}$'s mask match the corresponding bits in $\mathbf{h}$.

A *storage of a datum at the address* $\mathbf{X}$ means adding (as a vector) the datum to the contents of all the hard locations activated by $\mathbf{X}$. In the sequel we assume that $T + 1$ storages have been made at addresses $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_T$, the contents of all hard locations initially being zero.

A *reading at the address* $\mathbf{Y}$ means calculating by some method, from the contents of the hard locations activated by $\mathbf{Y}$, a datum to be read at $\mathbf{Y}$. In this paper we will just consider the standard-type reading: for each position, sum the (activated) contents and choose the reading 1 if the sum is ≥ 0, and -1 if the sum is < 0.

We now assume that we want to retrieve the datum stored at $\mathbf{X}_0$ and that the reading address $\mathbf{Y}$ is "near" $\mathbf{X}_0$, i.e., that the Hamming distance $d(\mathbf{Y}, \mathbf{X}_0)$ is small in some sense to be made precise below.

We will consider only the "maximally random" case which, for Kanerva's and Jaeckel's designs, means that we assume:

1. The M hard locations are randomly (uniformly and independently) distributed among all locations.

2. The storage addresses $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_T$ are randomly distributed among all addresses.

3. The stored data are randomly distributed among all possible data.

4. The reading address $\mathbf{Y}$ is chosen so that, for every bit in $\mathbf{Y}$ independently, the probability of this bit being different from the corresponding bit in $\mathbf{X}_0$, is equal to a small fixed number $\varepsilon \geq 0$, $\varepsilon = 0$ being the case of reading at the exact address ($\mathbf{Y} = \mathbf{X}_0$ with probability 1). Observe that this assumption implies that the stochastic variable $d(\mathbf{Y}, \mathbf{X}_0)$ has the binomial distribution $\mathrm{Bin}(N, \varepsilon)$.

To calculate the minimal value of $\mathrm{Pr}_{\mathrm{err}}(p)$ we use normal approximation of Z and get

$$\mathrm{Pr}_{\mathrm{err}}(p) \;=\; \mathrm{Pr}(Z < 0) \;\approx\; \Phi(-\rho) \qquad \text{.....................} \quad (2)$$

Here Φ is the standard normal distribution function, and $\rho = \mu/\sigma$ is the *signal-to-noise ratio*, where μ is the expected value and σ the standard deviation of Z.

A discussion of the correctness of the normal approximation is given in section 5.

4. Determination of the signal-to-noise ratio

Consider, for Kanerva's and Jaeckel's designs, the events "the hard location $\mathbf{h}$ is activated by $\mathbf{Y}$ and $\mathbf{X}_t$" and, for Karlsson's design, the events "$\mathbf{Y}$ and $\mathbf{X}_t$ coincide within the mask $\mathbf{m}$". For calculation of the distribution of L_t for a given $t \geq 1$, it would be fine if these events were independent for the different locations $\mathbf{h}$ (masks $\mathbf{m}$). In general this is not the case, however, since an activation of a hard location by $\mathbf{Y}$ and $\mathbf{X}_t$ indicates a relatively small value of $d(\mathbf{Y}, \mathbf{X}_t)$. But conditioned on a given distance $d(\mathbf{Y}, \mathbf{X}_t)$, the events are indeed independent (with probability depending on $d(\mathbf{Y}, \mathbf{X}_t)$). We observe that the distribution of $d(\mathbf{Y}, \mathbf{X}_t)$ is concentrated around $N/2$, the more so the larger N is. So by making the simplifying approximation that all these distances are equal to $N/2$, we get the desired independence. It is then pretty obvious that for a randomly chosen location $\mathbf{h}$, it is still true that the two events "$\mathbf{h}$ is activated by $\mathbf{Y}$" and "$\mathbf{h}$ is activated by $\mathbf{X}_t$" are independent, each having probability p.

According to Keeler [4] and Chou [5] this approximation will give a slight underestimate of the variance of the noise, hence a slight underestimate of the error probability. For not too small values of N the approximation will be good enough, though. These statements were verified by running a C-program for simulations of the SDM.

The calculations will now be easiest for Jaeckel's design, so we begin with that case. We find that

$$\mathrm{Pr}(\text{a random hard location } \mathbf{h} \text{ is activated by both } \mathbf{X}_t \text{ and } \mathbf{Y}) =$$

$$= \mathrm{Pr}(\mathbf{h} \text{ is activated by } \mathbf{X}_t) \cdot \mathrm{Pr}(\mathbf{h} \text{ is activated by } \mathbf{Y} \mid \mathbf{h} \text{ is activated by } \mathbf{X}_t),$$

which is $p{\cdot}p$ for $t = 1, 2, \ldots, T$, and $p{\cdot}(1-\varepsilon)^K$ for $t = 0$. Since $p = 2^{-K}$ we get $(1-\varepsilon)^K = p^\gamma$, where $\gamma = -\log_2(1-\varepsilon)$. It follows that

$t = 1, 2, \ldots, T$ and probability $(1-\varepsilon)^K = p^\gamma$ for $t = 0$, exactly one common hard location. Thus we see that

$$L_t \text{ has the distribution } \mathrm{Bin}(pM, p), \qquad \text{for } t = 1, 2, \ldots, T$$

and

$$L_0 \text{ has the distribution } \mathrm{Bin}(pM, p^\gamma).$$

So we get

$$\mu = E(Z) = Mp^{1+\gamma} \qquad \text{.................. (3Kar)}$$

and, using $\mathrm{Var}(L_0) = pMp^\gamma(1-p^\gamma)$ and $\mathrm{Var}(L_tB_t) = \mathrm{Var}(L_t) + [E(L_t)]^2 =$ $= pMp(1-p) + (Mp^2)^2$,

$$\sigma^2 = \mathrm{Var}(Z) = M\cdot[p^{1+\gamma} - p^{1+2\gamma} + Tp^2 - Tp^3 + TMp^4] \qquad \text{............ (4Kar)}$$

and thus

$$\rho^2 = h(p) = \frac{Mp^{1+\gamma}}{1 - p^\gamma + Tp^{1-\gamma} - Tp^{2-\gamma} + TMp^{3-\gamma}} \qquad \text{............ (5Kar)}$$

The maximal value of $h(p)$ is gotten for the p-value between 0 and 1 satisfying

$$p^{3-\gamma} - ap^{2-\gamma} - bp^{1-\gamma} + cp^\gamma - d = 0 \qquad \text{............ (6Kar)}$$

where $a = \dfrac{1-2\gamma}{2(1-\gamma)M}$, $b = \dfrac{\gamma}{(1-\gamma)M}$, $c = \dfrac{1}{2(1-\gamma)TM}$, $d = \dfrac{1+\gamma}{2(1-\gamma)TM}$.

Finally we look at Kanerva's original design. We still have that

$$L_t \text{ has the distribution } \mathrm{Bin}(M, p^2), \qquad \text{for } t = 1, 2, \ldots, T.$$

Putting q = Pr(a random hard location $\mathbf{h}$ is activated by both $\mathbf{X}_0$ and $\mathbf{Y}$)

$$L_0 \text{ has the distribution } \mathrm{Bin}(M, q).$$

So we get

$$\mu = E(Z) = Mq \qquad \text{.................. (3Kan)}$$

and

$$\sigma^2 = \mathrm{Var}(Z) = M\cdot[q - q^2 + Tp^2 + T(M-1)p^4] \qquad \text{........... (4Kan)}$$

This gives us the following expression for ρ^2:

$$q \approx 2^{-N} \cdot \sum_{k=0}^{R} \Phi\left(\frac{R-k+0.5-(N-2k)\varepsilon}{\sqrt{N\varepsilon(1-\varepsilon)}}\right) \cdot \binom{N}{k} \qquad \text{.......... (8Kan)}$$

This formula is used in the numerical calculations for $\varepsilon > 0$.

5. Discussion of the normal approximation of Z

We use the following realistic constraints: M is large, normally $M \geq 10^6$, T is of order $0.01M$, ε takes values between 0 and 0.15, and p will not be near 1. (It is obvious that p-values near 1 would give a very poor performance.)

Looking at the distributions of L_0, L_tB_t, and $\sum_{t=1}^{T} L_tB_t$ we find that approximate normal distribution of Z could be doubted only when p is very small, of order $1/M$ or less. Then any address activates only very few hard locations, if any, and it could be argued that for certain reasons such values of p are not realistic. However, let us estimate the error probability for some cases.

For Jaeckel's design, we look at the case $p = 1/M$ and ε $(= \gamma) = 0$. Then L_0 and $S = \sum_{t=1}^{T} L_t$ are well approximated by Poisson distributions with parameters $\lambda = 1$ and $\lambda = T/M$ respectively. It is easily seen that with respect to the values of L_0 and S the dominating event for which the reading is in error will be $B_0 = -1$ and $L_0 = S = 0$. This gives, for $T/M = 0.01$, $\Pr_{\text{err}} \approx 0.5 \cdot e^{-1} \cdot e^{-0.01} = 0.18$, a very poor performance. Unjustified use of equation (2), instead, would give $\Pr_{\text{err}} \approx \Phi(-1) \approx 0.16$, with unexpectedly good agreement. For values of $p \ll 1/M$, (2) also gives the obvious result $\Pr_{\text{err}} \approx 0.5$. For $p = 1/M$ and $\varepsilon > 0$, similar results are obtained. Although nothing is strictly proven for intermediate values of p it seems correct, in Jaeckel's case, to use the assumption of normal distribution, i.e., to use (6J), when seeking the optimal value of $\Pr_{\text{err}}(p)$.

Looking at Karlsson's design for $\varepsilon = 0$ we find that *if* we assume normal distribution of Z, (6Kar) would give $p = 1/(2M)$ so the optimal p would be the least possible value $p = 1/M$. But for this p, normal approximation would be very bad, and the following calculations will indeed show that this is definitely not the optimal value of p. For $p = 1/M$ we see that L_0 is

Let us first look at Kanerva's model. If the hard locations are drawn independently according to a uniform distribution it is possible that they, due to bad luck, are too clustered in Hamming space. In this case the variance of the L_t's will increase (still assuming that the storage addresses are randomly distributed), and performance will probably get worse. However, it is unlikely that a significant degradation of performance due to this type of clustering will occur, if the memory is not very small.

The same reasoning goes for Jaeckel's model (for Jaeckel's and Karlsson's designs, let us agree to measure the "nearness" of two hard locations by the number of matching bits in common mask coordinates). Observe that the problem should be somewhat greater for Karlsson's design, since then only the r masks are freely chosen, and r is small compared with M. One way to try to avoid the problem would be the following: generate the masks by successively choosing random coordinates, with the restriction that a coordinate must not be chosen if there are other coordinates that have been chosen less often; a new mask is generated each time that K coordinates have been chosen. (The risk of clustering will then diminish somewhat but not disappear). Another idea would be to use some deterministic algorithm to distribute the masks as evenly as possible in the space of masks. This is a difficult task, though. The theory of constant-weight codes deals with the problem, but gives results only for some special (and small) values of the parameters. (We remark that when we use another method for choosing masks than the ordinary random one, the optimal number of masks might also change.)

By using a stochastic algorithm programmed in C, some examples of mask distributions with high minimum distance between masks were found, with parameters fitting some simulations run by Karlsson, where the masks were chosen by the above-mentioned restricted random method. The result was a slight improvement of performance.

It is obvious that performance will deteriorate if the storage addresses are too clustered, e.g., if many addresses coincide on a large subset of the coordinates. An indication of how this problem could be handled is given in Sjödin [4].

Another problem could arise if the data are too clustered, depending on our use of the standard-type reading with threshold 0 for Z. How to find better thresholds (for all data positions) is discussed in Sjödin [4].

Finally, if the disturbance of the reading address is unevenly distributed over the coordinates, we could think of some way of lowering the importance of the coordinates that are disturbed the most.

8. Summary and conclusions

We have studied the performance of three different designs of Sparse Distributed Memory, namely, Kanerva's original one, Jaeckel's selected-coordinates design, and a modification of Jaeckel's design invented by Karlsson. We have made assumptions concerning randomness of hard locations, data and addresses, e. g., we have studied the effect of a random disturbance of the reading address. We have also given a short discussion of the effects of such assumptions.

We have derived formulas for the error probability when reading from the memory, and for calculating the optimal probability of activation. By numerical calculations we see that the performance is rather sensitive to the choice of probability of activation, and that the optimal probability varies with the disturbance of the reading address. For the optimal choice we find that for moderately disturbed reading addresses Jaeckel performs clearly better than Kanerva, and that Karlsson performs still slightly better. For very small (or none) disturbances of the reading address Kanerva performs as well as Jaeckel, though, and Karlsson performs clearly better. Scaling up, with more hard locations but the same load (number of stored patterns divided by the number of hard locations), we find that the performance for very small disturbances of the reading address gets better for Kanerva and Jaeckel, but that for larger disturbances it is worse for the bigger memory, for all three designs.

References

[1] Kanerva, Pentti — Sparse Distributed Memory and Related Models. In *Associative Neural Memories. Theory and Implementation*. Edited by Mohamad H. Hassoun. Oxford University Press. 1993.

[2] Jaeckel, Louis A. — An Alternative Design for a Sparse Distributed Memory. *RIACS Technical Report 89.28*. NASA Ames Research Center, 1989.

[3] Karlsson, Roland — A Fast Activation Mechanism for the Kanerva SDM Memory. In the *Proceedings of the '95 RWC Symposium*, pages 69–70, Tokyo, 1995.

[4] Sjödin, Gunnar — Improving the Capacity and Capabilities of SDM. 1995. To be published.

[5] Keeler, James D. — Comparison Between Kanerva's SDM and Hopfield-type Neural Networks. *Cognitive Science* 12: 299–329, 1988.

[6] Chou, Philip A. — The Capacity of the Kanerva Associative Memory. *IEEE Transactions on Information Theory,* 35 (2), 1989.

ε = 0.09			
Jaeckel	11 (optimal)	0.000488	0.032
	9	0.001953	0.13
	13	0.000122	0.049
Karlsson	11 (optimal)	0.000488	0.031
	9	0.001953	0.13
	13	0.000122	0.047
Kanerva	447 (optimal)	0.000445	0.082
	453	0.001626	0.17
	441	0.000106	0.11
ε = 0.12			
Jaeckel	11 (optimal)	0.000488	0.099
	9	0.001953	0.20
	13	0.000122	0.14
Karlsson	11 (optimal)	0.000488	0.099
	9	0.001953	0.20
	13	0.000122	0.14
Kanerva	447 (optimal)	0.000445	0.16
	453	0.001626	0.24
	441	0.000106	0.20
ε = 0.15			
Jaeckel	11 (optimal)	0.000488	0.19
	9	0.001953	0.27
	13	0.000122	0.24
Karlsson	11 (optimal)	0.000488	0.19
	9	0.001953	0.27
	13	0.000122	0.24
Kanerva	448 (optimal)	0.000558	0.24
	454	0.001992	0.31
	442	0.000136	0.27

Table 3. Error probabilities for optimal and nearly optimal values of the probability of actication.
Parameter values: $N = 1{,}000$ $M = 67{,}108{,}864$ $T = 6{,}710{,}886$

	K (R)	p	Pr(error)
$\varepsilon = \mathbf{0}$			
Jaeckel	17 (optimal)	0.000008	0.00089
	23 (cf Kar)	0.000000	0.018
Karlsson	23 (~optimal(?))	0.000000	0.00078
	17 (cf J)	0.000008	0.00080
Kanerva	432 (optimal)	0.000010	0.00088
$\varepsilon = \mathbf{0.03}$			
Jaeckel	15 (optimal)	0.000031	0.026
Karlsson	15 (optimal)	0.000031	0.026
Kanerva	436 (optimal)	0.000029	0.080
$\varepsilon = \mathbf{0.06}$			
Jaeckel	15 (optimal)	0.000031	0.11
Karlsson	15 (optimal)	0.000031	0.11
Kanerva	437 (optimal)	0.000038	0.19
$\varepsilon = \mathbf{0.15}$			
Jaeckel	14 (optimal)	0.000061	0.39
Karlsson	14 (optimal)	0.000061	0.39
Kanerva	439 (optimal)	0.000064	0.40

Table 4. Capacity T/M (in %) giving error probability 0.01
Parameter values: $N = 1{,}000$, $M = 1{,}000{,}000$

	Jaeckel	Karlsson	Kanerva
$\varepsilon = \mathbf{0}$	17.4	18.5	17.4
$\varepsilon = \mathbf{0.05}$	4.88	5.00	2.58
$\varepsilon = \mathbf{0.1}$	1.42	1.43	0.80
$\varepsilon = \mathbf{0.2}$	0.103	0.104	0.083

Table 5. Capacity T/M (in %) giving error probability 0.001
Parameter values: $N = 1{,}000$, $M = 1{,}000{,}000$

	Jaeckel	Karlsson	Kanerva
$\varepsilon = \mathbf{0}$	9.51	10.5	9.61
$\varepsilon = \mathbf{0.05}$	2.68	2.79	1.42
$\varepsilon = \mathbf{0.1}$	0.78	0.80	0.43
$\varepsilon = \mathbf{0.2}$	0.056	0.057	0.044

Report R95:09

ISRN : SICS-R--95/09-SE
ISSN : 0283-3638

Best Probability of Activation and Performance Comparisons for Several Designs of Sparse Distributed Memory

Jan Kristoferson

RWCP[1] Neuro SICS[2] Laboratory
Box 1263, S–164 28 KISTA, SWEDEN
Phone +46 8 752 15 00 Fax +46 8 751 72 30
E-mail: janke@sics.se

November 30, 1995

Abstract

The optimal probability of activation and the corresponding performance is studied for three designs of Sparse Distributed Memory, namely, Kanerva's original design, Jaeckel's selected-coordinates design and Karlsson's modification of Jaeckel's design. We will assume that the hard locations (in Karlsson's case, the masks), the storage addresses and the stored data are randomly chosen, and we will consider different levels of random noise in the reading address.

Keywords: Sparse Distributed Memory, Probability of Activation, Performance

Contents

[1] Real World Computing Partnership
[2] Swedish Institute of Computer Science

1. Introduction

We will look at Kanerva's original design of Sparse Distributed Memory, Jaeckel's design, and a modification of Jaeckel's design invented by Roland Karlsson at SICS, having great implementation advantages for simulations on a sequential computer. For a general discussion of these different designs see Kanerva [1], Jaeckel [2], and Karlsson [3].

The performance of the different designs will be studied assuming that

1. the hard locations in the memory, the data and the addresses for storing data are randomly chosen,
2. the reading procedure is the standard one (used in [1] and [2]; better performance can be achieved with other reading procedures, being investigated in Sjödin [4]),
3. the reading address may be disturbed (i.e., it may contain noise).

In particular, we will consider the question of determining the optimal probability p of activation. For his original design of the memory, Kanerva [1] derived a formula for the optimal p, when reading at the exact address. We will see how the optimal p changes when looking at the other designs, and when the reading address is disturbed. We will also calculate performance for optimal p. Jaeckel's claim of his design performing better than Kanerva's original one for moderately disturbed reading addresses will be confirmed, and we will see that Karlsson's design performs still slightly better.

2. General definitions and assumptions

We will use the following definitions:

An *address* $\mathbf{X}$ is a member of the set $\{0, 1\}^N$, i.e., a binary string of length N.

A *mask* $\mathbf{m}$ is a subset of K coordinates (out of N coordinates) in the address strings.

A *datum* is a member of $\{-1, 1\}^U$, i.e., a binary string of length U (the bits now being 1 or −1, also called "bipolar" representation).

A *location* is a potential place where to store data. In Kanerva's original design a location is given by an address, so there are totally 2^N locations. In the other two designs a location is given by a mask and K bits (0 or 1), one bit for each coordinate in the mask, so then there are totally $\binom{N}{K}\cdot 2^K$ locations.

A *hard location* $\mathbf{h}$ is a location existing in the memory; let M be the total number of hard locations.

The *content* of a hard location is a U-dimensional vector of integers. The coordinates of the content vectors are called *positions*.

In Kanerva's design an address $\mathbf{X}$ *activates* a hard location $\mathbf{h}$ if the Hamming distance $d(\mathbf{X}, \mathbf{h})$ is at most equal to the *activation radius* R.

In the other two designs an address $\mathbf{X}$ *activates* a hard location $\mathbf{h}$ if all bits in $\mathbf{X}$ lying in $\mathbf{h}$'s mask match the corresponding bits in $\mathbf{h}$.

A *storage of a datum at the address* $\mathbf{X}$ means adding (as a vector) the datum to the contents of all the hard locations activated by $\mathbf{X}$. In the sequel we assume that $T + 1$ storages have been made at addresses $\mathbf{X}_0$, $\mathbf{X}_1, \ldots,$ $\mathbf{X}_T$, the contents of all hard locations initially being zero.

A *reading at the address* $\mathbf{Y}$ means calculating by some method, from the contents of the hard locations activated by $\mathbf{Y}$, a datum to be read at $\mathbf{Y}$. In this paper we will just consider the standard-type reading: for each position, sum the (activated) contents and choose the reading 1 if the sum is ≥ 0, and -1 if the sum is < 0.

We now assume that we want to retrieve the datum stored at $\mathbf{X}_0$ and that the reading address $\mathbf{Y}$ is "near" $\mathbf{X}_0$, i.e., that the Hamming distance $d(\mathbf{Y}, \mathbf{X}_0)$ is small in some sense to be made precise below.

We will consider only the "maximally random" case which, for Kanerva's and Jaeckel's designs, means that we assume:

1. The M hard locations are randomly (uniformly and independently) distributed among all locations.
2. The storage addresses $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_T$ are randomly distributed among all addresses.
3. The stored data are randomly distributed among all possible data.
4. The reading address $\mathbf{Y}$ is chosen so that, for every bit in $\mathbf{Y}$ independently, the probability of this bit being different from the corresponding bit in $\mathbf{X}_0$, is equal to a small fixed number $\varepsilon \geq 0$, $\varepsilon = 0$ being the case of reading at the exact address ($\mathbf{Y} = \mathbf{X}_0$ with probability 1). Observe that this assumption implies that the stochastic variable $d(\mathbf{Y}, \mathbf{X}_0)$ has the binomial distribution $\mathrm{Bin}(N, \varepsilon)$.

Karlsson's design differs from Jaeckel's thus: for each used mask ("hard-location mask") all the 2^K possible bit combinations are used for hard locations. So $M = r \cdot 2^K$ if we have totally r hard-location masks. For Karlsson's design the first assumption (1) should be replaced by

1. The r hard-location masks are randomly (uniformly and independently) distributed among all possible masks.

Define the *probability of activation* p as the probability for a randomly chosen hard location to be activated by a randomly chosen address (e.g., by $\mathbf{X}_t$).

Observe that $p = 2^{-N} \cdot \sum_{k=0}^{R} \binom{N}{k}$ for Kanerva's design and $p = 2^{-K}$ for the other two designs.

3. The error probability and the signal-to-noise ratio

The material in this section is well-known from [1] and [2]. Define the stochastic variables

L_t = the number of hard locations activated by both $\mathbf{Y}$ and $\mathbf{X}_t$, $t = 0, 1, \ldots, T$.

Fix any position in the datum vectors and define the stochastic variables

B_t = the bit (1 or −1) stored in the considered position at the address $\mathbf{X}_t$, $t = 0, 1, \ldots, T$.

We see that, for the reading at address $\mathbf{Y}$, the sum Z of the (activated) contents in the considered fixed position will be

$$Z = \sum_{t=0}^{T} L_t B_t = L_0 + \sum_{t=1}^{T} L_t B_t \qquad \text{..................... (1)}$$

where the assumption $B_0 = 1$ is made in the last member. The probability of getting the wrong reading at the given position then equals the probability $\Pr(Z < 0)$. (Observe the asymmetry here: if we assume $B_0 = -1$ the probability of error equals $\Pr(Z \geq 0)$, where the inequality is nonstrict. This will be compensated for by not using "half-correction" in the normal approximation of Z below.)

For given values of N, M, T, and ε this probability depends only on p (i.e., on R in Kanerva's case and on K in the other two), and it will be denoted $\Pr_{\text{err}}(p)$. Minimizing $\Pr_{\text{err}}(p)$ gives the *best probability p of activation*.

To calculate the minimal value of $\Pr_{\text{err}}(p)$ we use normal approximation of Z and get

$$\Pr_{\text{err}}(p) \;=\; \Pr(Z < 0) \;\approx\; \Phi(-\rho) \qquad \text{..................... (2)}$$

Here Φ is the standard normal distribution function, and $\rho = \mu / \sigma$ is the *signal-to-noise ratio*, where μ is the expected value and σ the standard deviation of Z.

A discussion of the correctness of the normal approximation is given in section 5.

4. Determination of the signal-to-noise ratio

Consider, for Kanerva's and Jaeckel's designs, the events "the hard location $\mathbf{h}$ is activated by $\mathbf{Y}$ and $\mathbf{X}_t$" and, for Karlsson's design, the events "$\mathbf{Y}$ and $\mathbf{X}_t$ coincide within the mask $\mathbf{m}$". For calculation of the distribution of L_t for a given $t \geq 1$, it would be fine if these events were independent for the different locations $\mathbf{h}$ (masks $\mathbf{m}$). In general this is not the case, however, since an activation of a hard location by $\mathbf{Y}$ and $\mathbf{X}_t$ indicates a relatively small value of $d(\mathbf{Y}, \mathbf{X}_t)$. But conditioned on a given distance $d(\mathbf{Y}, \mathbf{X}_t)$, the events are indeed independent (with probability depending on $d(\mathbf{Y}, \mathbf{X}_t)$). We observe that the distribution of $d(\mathbf{Y}, \mathbf{X}_t)$ is concentrated around $N/2$, the more so the larger N is. So by making the simplifying approximation that all these distances are equal to $N/2$, we get the desired independence. It is then pretty obvious that for a randomly chosen location $\mathbf{h}$, it is still true that the two events "$\mathbf{h}$ is activated by $\mathbf{Y}$" and "$\mathbf{h}$ is activated by $\mathbf{X}_t$" are independent, each having probability p.

According to Keeler [4] and Chou [5] this approximation will give a slight underestimate of the variance of the noise, hence a slight underestimate of the error probability. For not too small values of N the approximation will be good enough, though. These statements were verified by running a C-program for simulations of the SDM.

The calculations will now be easiest for Jaeckel's design, so we begin with that case. We find that

$$\Pr(\text{a random hard location } \mathbf{h} \text{ is activated by both } \mathbf{X}_t \text{ and } \mathbf{Y}) =$$

$$= \Pr(\mathbf{h} \text{ is activated by } \mathbf{X}_t) \cdot \Pr(\mathbf{h} \text{ is activated by } \mathbf{Y} \mid \mathbf{h} \text{ is activated by } \mathbf{X}_t),$$

which is $p \cdot p$ for $t = 1, 2, \ldots, T$, and $p \cdot (1 - \varepsilon)^K$ for $t = 0$. Since $p = 2^{-K}$ we get $(1 - \varepsilon)^K = p^\gamma$, where $\gamma = -\log_2(1 - \varepsilon)$. It follows that

L_t has the distribution $\mathrm{Bin}(M, p^2)$, for $t = 1, 2, \ldots, T$,

and

L_0 has the distribution $\mathrm{Bin}(M, p^{1+\gamma})$, where $\gamma = -\log_2(1-\varepsilon)$.

The expected value $\mu = E(Z)$ and the standard deviation $\sigma = \sqrt{\mathrm{Var}(Z)}$ are to be determined. L_t and B_t are independent. For $t \geq 1$, $E(B_t) = 0$ and consequently $E(L_tB_t) = E(L_t)\cdot E(B_t) = 0$. Thus (1) gives $E(Z) = E(L_0) + 0$, so

$$\mu = E(Z) = Mp^{1+\gamma} \qquad \text{(3J)}$$

Furthermore, $\mathrm{Var}(L_0) = Mp^{1+\gamma}(1-p^{1+\gamma})$, and for $t \geq 1$ we have $\mathrm{Var}(L_tB_t) = E(L_t^2B_t^2) - [E(L_tB_t)]^2 = E(L_t^2) - 0 = \mathrm{Var}(L_t) + [E(L_t)]^2 = Mp^2(1-p^2) + (Mp^2)^2 = Mp^2 + M(M-1)p^4$. Since all $T + 1$ terms in the sum for Z in (1) are independent, we get

$$\sigma^2 = \mathrm{Var}(Z) = M\cdot[p^{1+\gamma} - p^{2+2\gamma} + Tp^2 + T(M-1)p^4] \qquad \text{(4J)}$$

So we get the following expression for ρ^2:

$$\rho^2 = g(p) = \frac{Mp^{1+\gamma}}{1 - p^{1+\gamma} + Tp^{1-\gamma} + T(M-1)p^{3-\gamma}} \qquad \text{(5J)}$$

By differentiating g, etc., it is found that the maximal value of $g(p)$ is gotten for the single p-value between 0 and 1 satisfying the equation

$$p^{3-\gamma} - \alpha p^{1-\gamma} - \beta = 0 \qquad \text{(6J)}$$

where $\alpha = \dfrac{\gamma}{(1-\gamma)(M-1)}$ and $\beta = \dfrac{1+\gamma}{2(1-\gamma)T(M-1)}$. Observe that this value of p normally is not exactly the desired optimal value, since this must be $p = 2^{-K}$ for some integer K. We find that for reading at the exact address, i.e., $\varepsilon = \gamma = 0$, we get the same formula $p = \sqrt[3]{\dfrac{1}{2MT}}$ for the optimal probability of activation as Kanerva got for his original design (approximating $M - 1$ by M).

We now look at Karlsson's design. It is clear that every address $\mathbf{X}$ activates exactly one hard location for each hard-location mask, consequently $r = pM$ hard locations totally. So, for any hard-location mask, $\mathbf{Y}$ and $\mathbf{X}_t$ activate either no common hard location or, with probability p for

$t = 1, 2, \ldots, T$ and probability $(1 - \varepsilon)^K = p^\gamma$ for $t = 0$, exactly one common hard location. Thus we see that

$$L_t \text{ has the distribution } \mathrm{Bin}(pM, p), \qquad \text{for } t = 1, 2, \ldots, T$$

and

$$L_0 \text{ has the distribution } \mathrm{Bin}(pM, p^\gamma) .$$

So we get

$$\mu = E(Z) = Mp^{1+\gamma} \qquad \text{.................. (3Kar)}$$

and, using $\mathrm{Var}(L_0) = pMp^\gamma(1 - p^\gamma)$ and $\mathrm{Var}(L_t B_t) = \mathrm{Var}(L_t) + [E(L_t)]^2 =$ $= pMp(1-p) + (Mp^2)^2$,

$$\sigma^2 = \mathrm{Var}(Z) = M \cdot [p^{1+\gamma} - p^{1+2\gamma} + Tp^2 - Tp^3 + TMp^4] \qquad \text{............ (4Kar)}$$

and thus

$$\rho^2 = h(p) = \frac{Mp^{1+\gamma}}{1 - p^\gamma + Tp^{1-\gamma} - Tp^{2-\gamma} + TMp^{3-\gamma}} \qquad \text{............ (5Kar)}$$

The maximal value of $h(p)$ is gotten for the p-value between 0 and 1 satisfying

$$p^{3-\gamma} - ap^{2-\gamma} - bp^{1-\gamma} + cp^\gamma - d = 0 \qquad \text{............ (6Kar)}$$

where $a = \dfrac{1 - 2\gamma}{2(1-\gamma)M}$, $b = \dfrac{\gamma}{(1-\gamma)M}$, $c = \dfrac{1}{2(1-\gamma)TM}$, $d = \dfrac{1+\gamma}{2(1-\gamma)TM}$.

Finally we look at Kanerva's original design. We still have that

$$L_t \text{ has the distribution } \mathrm{Bin}(M, p^2), \qquad \text{for } t = 1, 2, \ldots, T.$$

Putting q = Pr(a random hard location **h** is activated by both $\mathbf{X}_0$ and $\mathbf{Y}$)

$$L_0 \text{ has the distribution } \mathrm{Bin}(M, q).$$

So we get

$$\mu = E(Z) = Mq \qquad \text{.................. (3Kan)}$$

and

$$\sigma^2 = \mathrm{Var}(Z) = M \cdot [q - q^2 + Tp^2 + T(M-1)p^4] \qquad \text{........... (4Kan)}$$

This gives us the following expression for ρ^2:

$$\rho^2 = k(p) = \frac{Mq^2}{q - q^2 + Tp^2 + T(M-1)p^4} \qquad \text{.................. (5Kan)}$$

If $\varepsilon = 0$ then $q = p$, and we get the same formula for ρ^2 as in Jaeckel's case. But, unlike in the other cases, we have no simple way to write (approximate) q in terms of p (and the other parameters), if $\varepsilon > 0$. So we will not optimize p (i.e., optimize R) by differentiation, etc., as before. Let us see how q can be approximated. Without loss of generality we can assume that the hard location $\mathbf{h}$ is the zero vector. Then $d(\mathbf{X}, \mathbf{h}) = w(\mathbf{X})$, where the *weight* $w(\mathbf{X})$ is the number of ones in $\mathbf{X}$. Let A and B be the events $w(\mathbf{X}_0) \leq R$ and $w(\mathbf{Y}) \leq R$, i.e., that $\mathbf{h}$ is activated by $\mathbf{X}_0$ and that $\mathbf{h}$ is activated by $\mathbf{Y}$, respectively. Let A_k be the event $w(\mathbf{X}_0) = k$. Then

$$q = \Pr(A \cap B) = \Pr(B \cap (A_0 \cup A_1 \cup \ldots \cup A_R)) = \sum_{k=0}^{R} \Pr(B \cap A_k), \text{ so}$$

$$q = \sum_{k=0}^{R} \Pr(B \mid A_k) \cdot \Pr(A_k) \qquad \text{.................... (7Kan)}$$

Here $\Pr(A_k) = 2^{-N} \cdot \binom{N}{k}$. To calculate $\Pr(B \mid A_k)$, suppose $w(\mathbf{X}_0) = k$. We need to consider only cases where $k \leq R < N/2$, and R rather near $N/2$. Define the stochastic variable C_0 to be the number of zeros in $\mathbf{X}_0$ being flipped in $\mathbf{Y}$; similarly, C_1 is the number of ones flipped. Then $\Pr(B \mid A_k) = \Pr(C_0 - C_1 \leq R - k)$. C_0 and C_1 have distributions $\mathrm{Bin}(N-k, \varepsilon)$ and $\mathrm{Bin}(k, \varepsilon)$, respectively. C_0 is well approximated by the normal distribution

$$\mathrm{Norm}(\ (N-k)\varepsilon, \sqrt{(N-k)\varepsilon(1-\varepsilon)}\ )$$

if the variance $(N-k)\varepsilon(1-\varepsilon)$ is not too small, say, if it is > 10. Since $N - k$ is $> N/2$ the approximation is good if $\varepsilon(1-\varepsilon) > 20/N$. Similarly, C_1 is well approximated by

$$\mathrm{Norm}(\ k\varepsilon, \sqrt{k\varepsilon(1-\varepsilon)}\ )$$

if $\varepsilon(1-\varepsilon) > 20/N$ and k is near R. So $C_0 - C_1$ is well approximated by

$$\mathrm{Norm}(\ (N-2k)\varepsilon, \sqrt{N\varepsilon(1-\varepsilon)}\ )$$

if $\varepsilon(1-\varepsilon) > 20/N$, and k is near R. In fact, if $\varepsilon(1-\varepsilon) > 20/N$ this approximation can be used in (7Kan) for all k, since $\Pr(A_k)$ is very small when k is not near R.

So, for $\varepsilon(1-\varepsilon) > 20/N$ we have (using "half-correction" in the normal approximation)

$$q \approx 2^{-N} \cdot \sum_{k=0}^{R} \Phi\left(\frac{R-k+0.5-(N-2k)\varepsilon}{\sqrt{N\varepsilon(1-\varepsilon)}}\right) \cdot \binom{N}{k} \qquad \text{.......... (8Kan)}$$

This formula is used in the numerical calculations for $\varepsilon > 0$.

5. Discussion of the normal approximation of Z

We use the following realistic constraints: M is large, normally $M \geq 10^6$, T is of order $0.01M$, ε takes values between 0 and 0.15, and p will not be near 1. (It is obvious that p-values near 1 would give a very poor performance.)

Looking at the distributions of L_0, L_tB_t, and $\sum_{t=1}^{T} L_tB_t$ we find that approximate normal distribution of Z could be doubted only when p is very small, of order $1/M$ or less. Then any address activates only very few hard locations, if any, and it could be argued that for certain reasons such values of p are not realistic. However, let us estimate the error probability for some cases.

For Jaeckel's design, we look at the case $p = 1/M$ and ε $(= \gamma) = 0$. Then L_0 and $S = \sum_{t=1}^{T} L_t$ are well approximated by Poisson distributions with parameters $\lambda = 1$ and $\lambda = T/M$ respectively. It is easily seen that with respect to the values of L_0 and S the dominating event for which the reading is in error will be $B_0 = -1$ and $L_0 = S = 0$. This gives, for $T/M = 0.01$, $\mathrm{Pr}_{\mathrm{err}} \approx 0.5 \cdot e^{-1} \cdot e^{-0.01} = 0.18$, a very poor performance. Unjustified use of equation (2), instead, would give $\mathrm{Pr}_{\mathrm{err}} \approx \Phi(-1) \approx 0.16$, with unexpectedly good agreement. For values of $p << 1/M$, (2) also gives the obvious result $\mathrm{Pr}_{\mathrm{err}} \approx 0.5$. For $p = 1/M$ and $\varepsilon > 0$, similar results are obtained. Although nothing is strictly proven for intermediate values of p it seems correct, in Jaeckel's case, to use the assumption of normal distribution, i.e., to use (6J), when seeking the optimal value of $\mathrm{Pr}_{\mathrm{err}}(p)$.

Looking at Karlsson's design for $\varepsilon = 0$ we find that *if* we assume normal distribution of Z, (6Kar) would give $p = 1/(2M)$ so the optimal p would be the least possible value $p = 1/M$. But for this p, normal approximation would be very bad, and the following calculations will indeed show that this is definitely not the optimal value of p. For $p = 1/M$ we see that L_0 is constant = 1, and that $S = \sum_{t=1}^{T} L_t$ is again approximated by the Poisson distribution with parameter $\lambda = T/M$. The dominating event giving erroneous reading is $B_0 = -1$, $S = 1$, and the one and only activated noise bit B_t being = 1. For $T/M = 0.01$ this gives $\Pr_{err} \approx 0.5 \cdot 0.01 e^{-0.01} \cdot 0.5 = 0.0025$. Unjustified use of (6Kar) and (2), instead, would give $\Pr_{err} \approx \Phi(-10) \approx 10^{-22}$. The analogous analysis goes for $p = r/M$, $\varepsilon = 0$, and a few very small values of r (say $r \le 5$), but then there is a gap with great combinatorial complexity before we get to the p-values for which normal approximation can be safely used. If $r = 2$, then L_0 is constant = 2, and the Poisson distribution for S has the parameter value $\lambda = 4T/M$. The dominating event giving erroneous reading is $B_0 = -1$, $S = 2$, and the two activated noise bits being equal to 1. $T/M = 0.01$ then gives $\Pr_{err} \approx 0.5 \cdot (0.04)^2 \cdot 0.5 e^{-0.04} \cdot (0.5)^2 = 0.0001$. If $r = 3$ we get $\Pr_{err} \approx 0.000007$. We now observe that for $10/(TM) \le p^2 << 1/M$ we can use the normal approximation, but still get very near the maximum of $h(p)$ gotten for $p = 1/M$. Despite the lack of proof we find it reasonable to assume that such a value of p, for instance $p \approx \sqrt{10/(TM)}$, gives nearly the optimal performance, and we will use this value in the numerical calculations.

For not too small values of ε, say $\varepsilon \ge 0.01$, the situation is about the same as for Jaeckel's design. Numerical tests have shown that if we use (6Kar) for $\varepsilon > 0.00001$, and use $p = \sqrt{10/(TM)}$ for $\varepsilon \le 0.00001$, the results look reasonable. The figures are somewhat uncertain for very small positive values of ε, though.

Finally, for Kanerva's design the situation is about the same as for Jaeckel's, so we will use normal approximation, i.e., (5Kan) and (2), throughout.

6. Discussion of the randomness assumptions for hard locations, storage addresses etc.

All calculations so far were based upon the assumptions 1–4 of "maximal randomness" in section 2. The relevance of these assumptions depends on the application, of course. For instance, Karlsson (cf. [3]) has studied some examples where some of the problems discussed below appeared very clearly. Now, imagine any situation with given hard locations, storage addresses, data, and distribution of noise in the reading addresses.

Let us first look at Kanerva's model. If the hard locations are drawn independently according to a uniform distribution it is possible that they, due to bad luck, are too clustered in Hamming space. In this case the variance of the L_t's will increase (still assuming that the storage addresses are randomly distributed), and performance will probably get worse. However, it is unlikely that a significant degradation of performance due to this type of clustering will occur, if the memory is not very small.

The same reasoning goes for Jaeckel's model (for Jaeckel's and Karlsson's designs, let us agree to measure the "nearness" of two hard locations by the number of matching bits in common mask coordinates). Observe that the problem should be somewhat greater for Karlsson's design, since then only the r masks are freely chosen, and r is small compared with M. One way to try to avoid the problem would be the following: generate the masks by successively choosing random coordinates, with the restriction that a coordinate must not be chosen if there are other coordinates that have been chosen less often; a new mask is generated each time that K coordinates have been chosen. (The risk of clustering will then diminish somewhat but not disappear). Another idea would be to use some deterministic algorithm to distribute the masks as evenly as possible in the space of masks. This is a difficult task, though. The theory of constant-weight codes deals with the problem, but gives results only for some special (and small) values of the parameters. (We remark that when we use another method for choosing masks than the ordinary random one, the optimal number of masks might also change.)

By using a stochastic algorithm programmed in C, some examples of mask distributions with high minimum distance between masks were found, with parameters fitting some simulations run by Karlsson, where the masks were chosen by the above-mentioned restricted random method. The result was a slight improvement of performance.

It is obvious that performance will deteriorate if the storage addresses are too clustered, e.g., if many addresses coincide on a large subset of the coordinates. An indication of how this problem could be handled is given in Sjödin [4].

Another problem could arise if the data are too clustered, depending on our use of the standard-type reading with threshold 0 for Z. How to find better thresholds (for all data positions) is discussed in Sjödin [4].

Finally, if the disturbance of the reading address is unevenly distributed over the coordinates, we could think of some way of lowering the importance of the coordinates that are disturbed the most.

7. Numerical calculations

A program in C was written and run on a SUN Sparc 5 workstation. We will show some numerical results in tables 1–5, pages 14–17.

Tables 1–3 show values of $\mathrm{Pr}_{\mathrm{err}}(p)$ for several choices of M, T/M, and ε. The value of N, if reasonably large, has a rather small influence, so we just look at $N = 1{,}000$. In tables 4 and 5 we give the values of the *capacity* T/M (using optimal values of p) for two given values of $\mathrm{Pr}_{\mathrm{err}}(p)$, namely 0.01 and 0.001, and for some different values of ε. We use the value M = 1,000,000 in accordance with Jaeckel's examples in [2]. (This is somewhat improper for Karlsson's design, 1,000,000 not containing high enough powers of 2, but for a comparative study it is quite all right.) Below we give a few comments on the results:

Table 1. Error probabilities:
We have chosen $M = 1{,}048{,}576 = 2^{20}$ and $T/M \approx 2.86\%$. We see that the performance is rather sensitive to the choice of K (R). For the optimal choice we see that for intermediate values of ε Jaeckel performs clearly better than Kanerva, and that (for $\varepsilon \geq 0.03$ at least) Karlsson performs slightly better than Jaeckel.

Table 2. Error probabilities:
Here $T/M = 10\%$. We see that for very small values of ε Karlsson performs clearly better than Jaeckel, and Kanerva performs as well as Jaeckel (slightly better).

Table 3. Error probabilities:
Here we have $M = 2^{26}$ (64 times larger than in tables 1 and 2), but the value of T/M is the same (10%) as in table 2. Comparing with table 2, we see that for Jaeckel and Kanerva this gives better performance for $\varepsilon = 0$, but for $\varepsilon \geq 0.03$ the performance is worse for all three designs.

Tables 4–5. Capacity:
The relations between the figures agree with the results in tables 1–3. The same goes for the influence of increasing M: e.g., M = 100,000,000 gives 18.2 instead of 17.4, and 3.62 instead of 5.00. The figures in the second and third rows give about the same relation between the capacity of Jaeckel's and Kanerva's design for intermediate values of ε as Jaeckel claimed in [2] (where he used nonoptimal probabilities of activation).

8. Summary and conclusions

We have studied the performance of three different designs of Sparse Distributed Memory, namely, Kanerva's original one, Jaeckel's selected-coordinates design, and a modification of Jaeckel's design invented by Karlsson. We have made assumptions concerning randomness of hard locations, data and addresses, e. g., we have studied the effect of a random disturbance of the reading address. We have also given a short discussion of the effects of such assumptions.

We have derived formulas for the error probability when reading from the memory, and for calculating the optimal probability of activation. By numerical calculations we see that the performance is rather sensitive to the choice of probability of activation, and that the optimal probability varies with the disturbance of the reading address. For the optimal choice we find that for moderately disturbed reading addresses Jaeckel performs clearly better than Kanerva, and that Karlsson performs still slightly better. For very small (or none) disturbances of the reading address Kanerva performs as well as Jaeckel, though, and Karlsson performs clearly better. Scaling up, with more hard locations but the same load (number of stored patterns divided by the number of hard locations), we find that the performance for very small disturbances of the reading address gets better for Kanerva and Jaeckel, but that for larger disturbances it is worse for the bigger memory, for all three designs.

References

[1] Kanerva, Pentti — Sparse Distributed Memory and Related Models. In *Associative Neural Memories. Theory and Implementation*. Edited by Mohamad H. Hassoun. Oxford University Press. 1993.

[2] Jaeckel, Louis A. — An Alternative Design for a Sparse Distributed Memory. *RIACS Technical Report 89.28*. NASA Ames Research Center, 1989.

[3] Karlsson, Roland — A Fast Activation Mechanism for the Kanerva SDM Memory. In the *Proceedings of the '95 RWC Symposium,* pages 69–70, Tokyo, 1995.

[4] Sjödin, Gunnar — Improving the Capacity and Capabilities of SDM. 1995. To be published.

[5] Keeler, James D. — Comparison Between Kanerva's SDM and Hopfield-type Neural Networks. *Cognitive Science* 12: 299–329, 1988.

[6] Chou, Philip A. — The Capacity of the Kanerva Associative Memory. *IEEE Transactions on Information Theory,* 35 (2), 1989.

Table 1. Error probabilities for optimal and nearly optimal values of the probability of actication.
Parameter values: $N = 1{,}000$ $M = 1{,}048{,}576$ $T = 30{,}000$

	K (R)	p	Pr(error)
$\varepsilon = \mathbf{0}$			
Jaeckel	12 (optimal)	0.000244	0.000000
	10	0.000977	0.000017
	14	0.000061	0.000001
	16 (cf Kar)	0.000015	0.00046
Karlsson	16 (~optimal(?))	0.000015	0.000000
	14	0.000061	0.000000
	12 (cf J)	0.000244	0.000000
Kanerva	445 (optimal)	0.000280	0.000000
	451	0.001072	0.000038
	439	0.000064	0.000001
$\varepsilon = \mathbf{0.03}$			
Jaeckel	12 (optimal)	0.000244	0.000069
	10	0.000977	0.0011
	14	0.000061	0.00047
Karlsson	12 (optimal)	0.000244	0.000043
	10	0.000977	0.0010
	14	0.000061	0.00014
Kanerva	445 (optimal)	0.000280	0.0022
	451	0.001072	0.014
	439	0.000064	0.0060
$\varepsilon = \mathbf{0.06}$			
Jaeckel	12 (optimal)	0.000244	0.0040
	10	0.000977	0.013
	14	0.000061	0.013
Karlsson	12 (optimal)	0.000244	0.0036
	10	0.000977	0.012
	14	0.000061	0.0099
Kanerva	446 (optimal)	0.000354	0.025
	452	0.001323	0.077
	440	0.000083	0.041

ε = 0.09			
Jaeckel	11 (optimal)	0.000488	0.032
	9	0.001953	0.13
	13	0.000122	0.049
Karlsson	11 (optimal)	0.000488	0.031
	9	0.001953	0.13
	13	0.000122	0.047
Kanerva	447 (optimal)	0.000445	0.082
	453	0.001626	0.17
	441	0.000106	0.11
ε = 0.12			
Jaeckel	11 (optimal)	0.000488	0.099
	9	0.001953	0.20
	13	0.000122	0.14
Karlsson	11 (optimal)	0.000488	0.099
	9	0.001953	0.20
	13	0.000122	0.14
Kanerva	447 (optimal)	0.000445	0.16
	453	0.001626	0.24
	441	0.000106	0.20
ε = 0.15			
Jaeckel	11 (optimal)	0.000488	0.19
	9	0.001953	0.27
	13	0.000122	0.24
Karlsson	11 (optimal)	0.000488	0.19
	9	0.001953	0.27
	13	0.000122	0.24
Kanerva	448 (optimal)	0.000558	0.24
	454	0.001992	0.31
	442	0.000136	0.27

Table 2. Error probabilities for optimal and nearly optimal values of the probability of actication. Parameter values: $N = 1{,}000$ $M = 1{,}048{,}576$ $T = 104{,}858$

	K (R)	p	Pr(error)
$\varepsilon = 0$			
Jaeckel	13 (optimal)	0.000122	0.0012
	17 (cf Kar)	0.000008	0.018
Karlsson	17 (~optimal(?))	0.000008	0.00078
	13 (cf J)	0.000122	0.00085
Kanerva	443 (optimal)	0.000174	0.0012
$\varepsilon = 0.00001$			
Jaeckel	13 (optimal)	0.000122	0.0013
Karlsson	17 (~optimal(?))	0.000008	0.00079
	13 (cf J)	0.000122	0.00085
Kanerva	443 (optimal)	0.000174	0.0012
$\varepsilon = 0.00002$			
Jaeckel	13 (optimal)	0.000122	0.0013
Karlsson	16 (optimal)	0.000015	0.00079
Kanerva	443 (optimal)	0.000174	0.0012
$\varepsilon = 0.03$			
Jaeckel	12 (optimal)	0.000244	0.018
Karlsson	12 (optimal)	0.000244	0.017
Kanerva	445 (optimal)	0.000280	0.060
$\varepsilon = 0.06$			
Jaeckel	12 (optimal)	0.000244	0.074
Karlsson	12 (optimal)	0.000244	0.073
Kanerva	446 (optimal)	0.000354	0.14
$\varepsilon = 0.15$			
Jaeckel	11 (optimal)	0.000488	0.32
Karlsson	11 (optimal)	0.000488	0.32
Kanerva	448 (optimal)	0.000558	0.35

Table 3. Error probabilities for optimal and nearly optimal values of the probability of actication.
Parameter values: N = 1,000 M = 67,108,864 T = 6,710,886

	K (R)	p	Pr(error)
$\varepsilon = \mathbf{0}$			
Jaeckel	17 (optimal)	0.000008	0.00089
	23 (cf Kar)	0.000000	0.018
Karlsson	23 (~optimal(?))	0.000000	0.00078
	17 (cf J)	0.000008	0.00080
Kanerva	432 (optimal)	0.000010	0.00088
$\varepsilon = \mathbf{0.03}$			
Jaeckel	15 (optimal)	0.000031	0.026
Karlsson	15 (optimal)	0.000031	0.026
Kanerva	436 (optimal)	0.000029	0.080
$\varepsilon = \mathbf{0.06}$			
Jaeckel	15 (optimal)	0.000031	0.11
Karlsson	15 (optimal)	0.000031	0.11
Kanerva	437 (optimal)	0.000038	0.19
$\varepsilon = \mathbf{0.15}$			
Jaeckel	14 (optimal)	0.000061	0.39
Karlsson	14 (optimal)	0.000061	0.39
Kanerva	439 (optimal)	0.000064	0.40

Table 4. Capacity T/M (in %) giving error probability 0.01
Parameter values: N = 1,000, M = 1,000,000

	Jaeckel	Karlsson	Kanerva
$\varepsilon = \mathbf{0}$	17.4	18.5	17.4
$\varepsilon = \mathbf{0.05}$	4.88	5.00	2.58
$\varepsilon = \mathbf{0.1}$	1.42	1.43	0.80
$\varepsilon = \mathbf{0.2}$	0.103	0.104	0.083

Table 5. Capacity T/M (in %) giving error probability 0.001
Parameter values: N = 1,000, M = 1,000,000

	Jaeckel	Karlsson	Kanerva
$\varepsilon = \mathbf{0}$	9.51	10.5	9.61
$\varepsilon = \mathbf{0.05}$	2.68	2.79	1.42
$\varepsilon = \mathbf{0.1}$	0.78	0.80	0.43
$\varepsilon = \mathbf{0.2}$	0.056	0.057	0.044