Finding a Needle in a Haystack: from Baldwin Effect to Quantum Computation

Akira Imada

Brest State Technical University, Belarus

Alora's proposition (2007)

Brilliance = Ability to find a needle in a haystack.

"Art is to create a rarely possible appreciation somewhere in a huge universe of our brain."

(Cont'd)

(Artificial) *INTELLIGENCE*
can search, classify and recognize normal objects
while
only *BRILLIANCE*
can search for a needle in a hay.

Could an endlessly typing monkey eventually create Shakespeare by chance?

Hopefully not!

Then what about *haiku*?
Is there really a chance it would happen at all?

うぐいすの
かさおとしたる
つばきかな

U-Gu-I-Su-No
Ka-Sa-O-To-Shi-Ta-Ru
Tsu-Ba-Ki-Ka-Na

How many possible different candidates

$$76^{17} \approx 10^{31}$$

It's still like a search for *a-needle-in-a-haystack* for a monkey to create a meaningful one.

Computational analogue of a-needle-in-a-haystack

A specific rare cases in a huge database;
A real necessary information from world-wide-web;
Oil spills in the ocean from satellite image;
A cause of failures in diagnosing a huge code;
A collision of a hash function;
etc......
⇓
Such searches are very important for us.

Associative memory by Spiking neurons

Fitness landscape

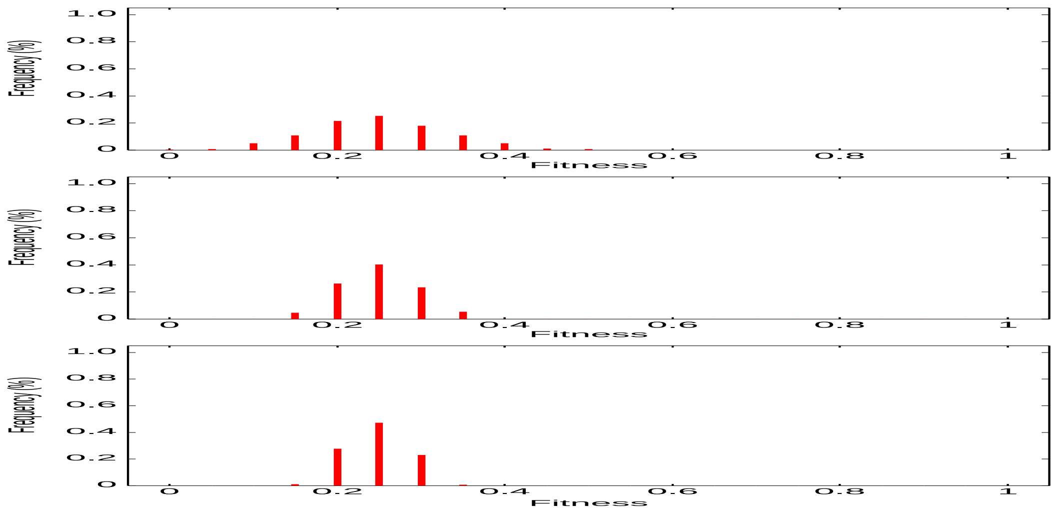
$\Downarrow$

Hopfield network ... many peaks in a rugged land

vs.

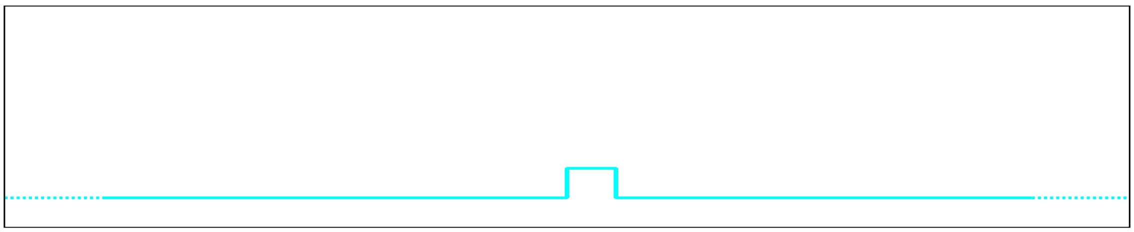
Spiking neurons ... like a flat island in a huge lake

A standard initial distribution of fitness values

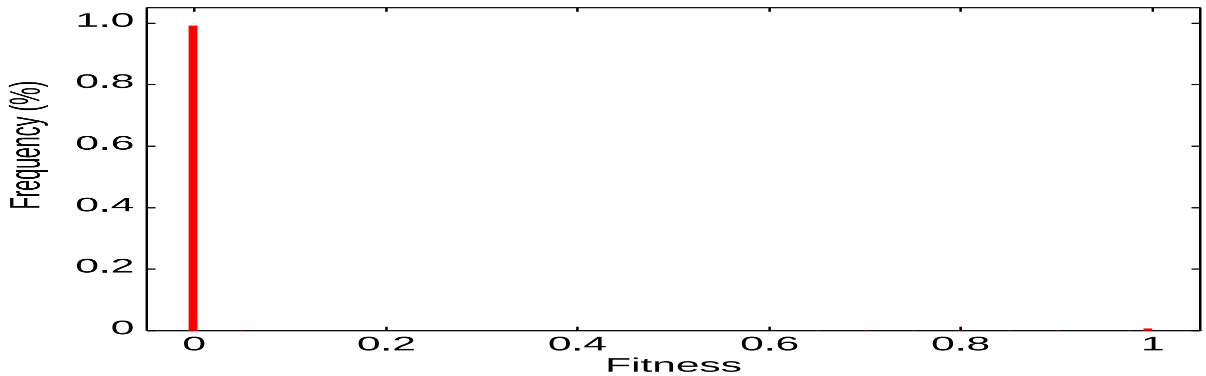


A-tiny-flat-island-in-a-huge-lake

A fictitious 1-D fitness-landscape
of Hebbian-learning by spiking neurons



Then, fitness distribution is ...



It's not possible to hillclimb if not a hill to climb!

Network intrusion data among normal transaction data

E.g. $\Rightarrow$ 4 attack types in KDD-cup-99 dataset

The winner's result

Probe	DoS	U2R	R2L
83.3%	97.1%	13.2%	8.4%

Detection rate by Sabhnani et al. (2003)

	Probe	DoS	U2R	R2L
Multi-layer Perceptron	88.7	97.2	13.2	5.6
Gaussian Classifier	90.2	82.4	22.8	9.6
K-mean Clustering	87.6	97.3	29.8	6.4
Nearest Cluster Algorithm	88.8	97.1	2.2	3.4
Radial Basis Function	93.2	73.0	6.1	5.9
Leader Algorithm	83.8	97.2	6.6	1.0
Hypersphere Algorithm	84.8	97.2	8.3	1.0
Fuzzy Art Map	77.2	97.0	6.1	3.7
C4.5 Decision Tree	80.8	97.0	1.8	4.6

Some Intrusions are immune to any detectors

Probably because they are like needles in a huge hay.

(Joshi et al. 2001)
"If a class is very rare, e.g., 0.5%, prediction of non-target will be 99.5%."

(Nevertheless, we have still many innocent optimistic reports of success.)

We now look at the other 4 cases more in detail!

I. Hinton & Nowlan's experiment

II. Reduced even-parity problem as a benchmark

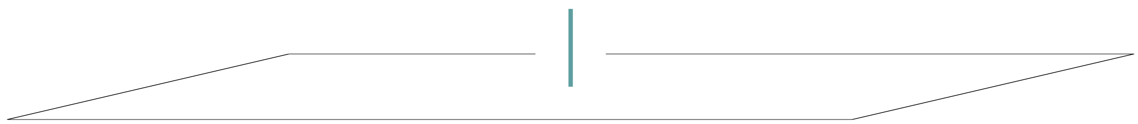
III. A robot navigation

IV. Quantum computation

$< I >$

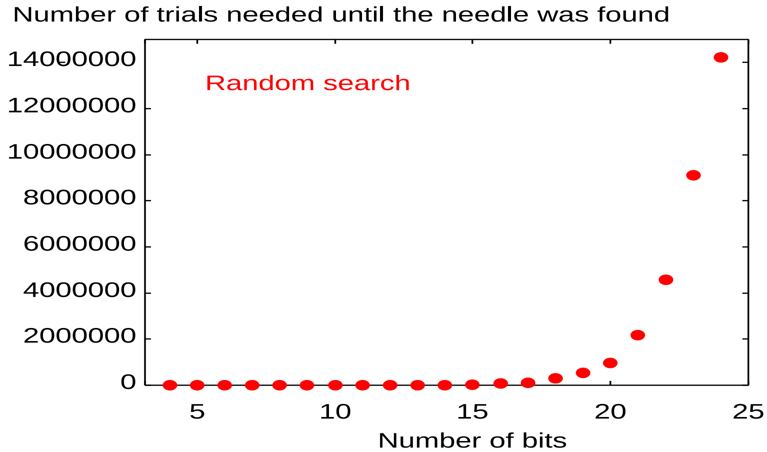
Hinton & Nowlan's Simplest Needle in a Hay.

Hinton & Nowlan's Needle (1987)



- A-needle ⇒ A configuration of 20-bit binary strings
- Haystack ⇒ $2^{20} - 1$ search points

Their choice of 20-bit was a good one!



Evolution under Baldwin effect – Lifetime learning of phenotype

A genotype:

(10901099011001001091)

$\Downarrow$

Its phenotype:

(10101001011001001001)
(10001011011001001011)

...

(Cont'd)

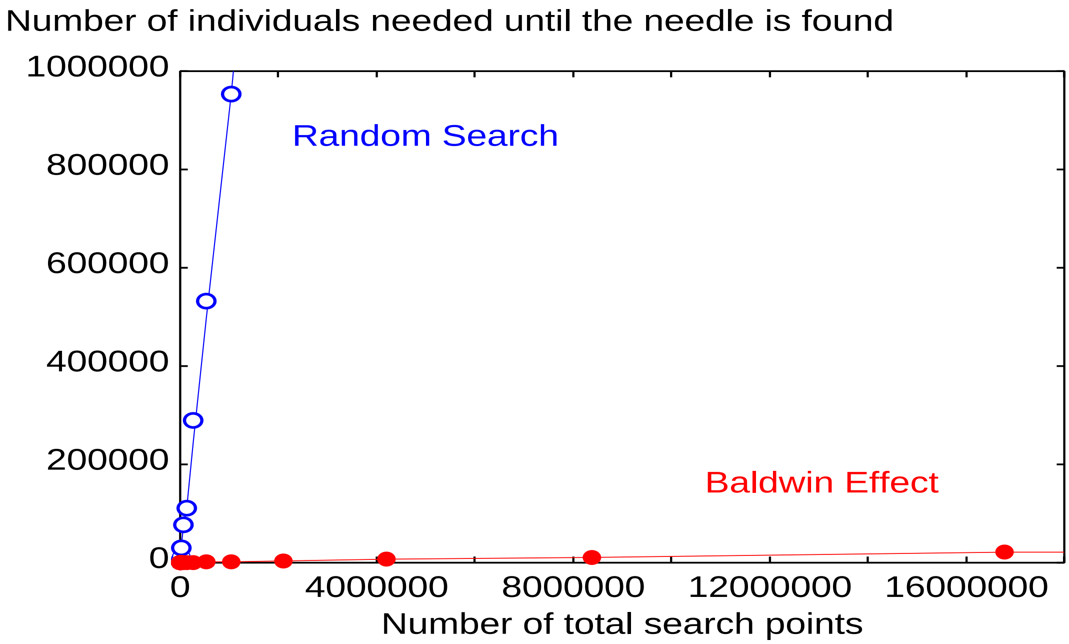
(Their assumption)
The closer the *genotype* to the needle,
the faster the learning of *phenotype.*
$\Downarrow$
This makes the needle-like-peak smoother.

Their GA implementation

1000 genotypes with 25% flexible genes,
each allowed 1000 lifetime learnings;

Roulette wheel selection;
One-point cross over;
Mutation;
......

Baldwin effect looks great!



(Cont'd)

Note, anyway, in both cases,
$O(N)$ steps are necessary!

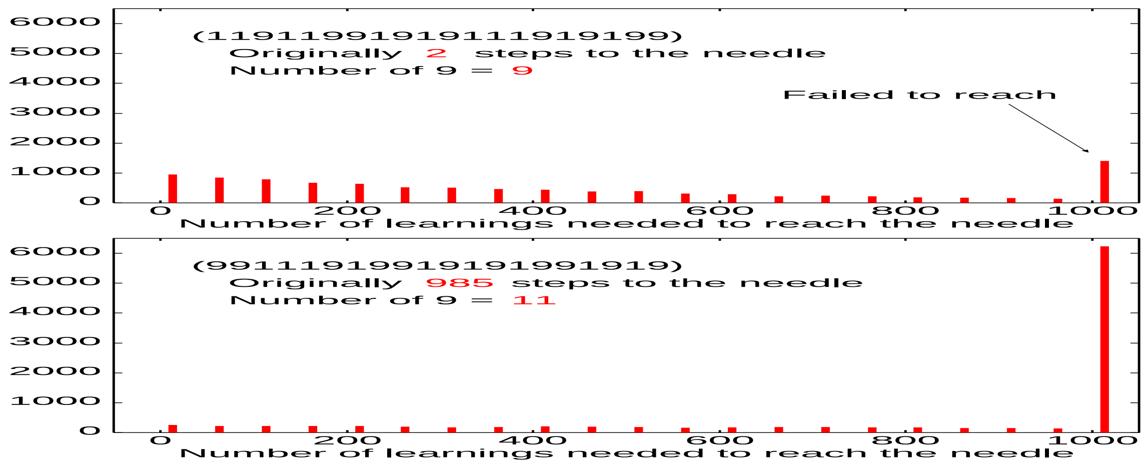
Are we happy with this Baldwin effect?

Why should we continue
when lifetime-learning already has found the needle?
$\Downarrow$
(Turney 1987)

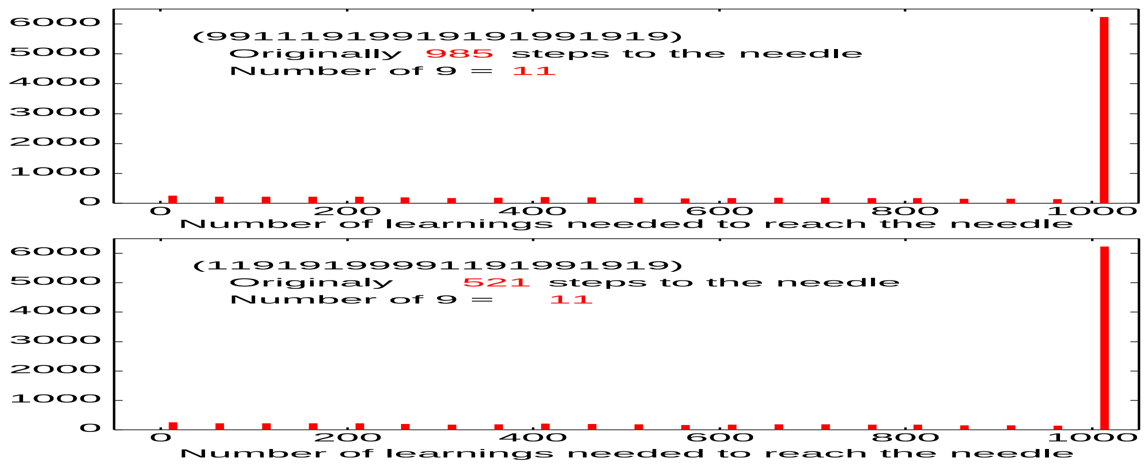
"Not from an engineering but a biological interest."

Is a high fitness gene a good gene?

(How many successes out of 1000 learnings of 1000 phenotypes)



Goodness depends on number of 9?



(Cont'd)

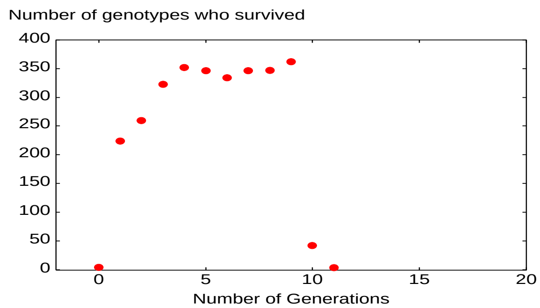
How many 9's are optimal?

or

Is the number of 9's decreasing?

An extinction!

In the 1st generation,
only 4 successful genotypes are found at the luckiest case.



Evolution was not successful.

What about Lamarckian Inheritance?

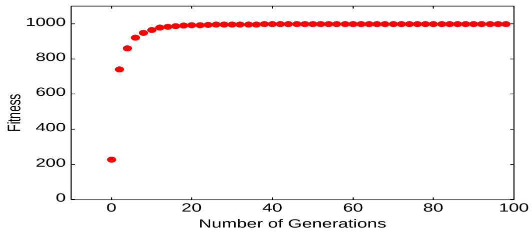
An inverse mapping from phenotype to genotype is necessary.

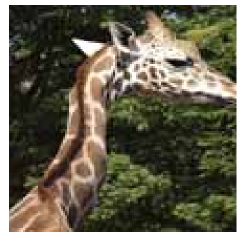
Turney (1996)

"We believe that computing this mapping is intractable,"

Why giraffe has a long neck?

Let's re-map a few of successful 9's in phenotype
to its genotype.



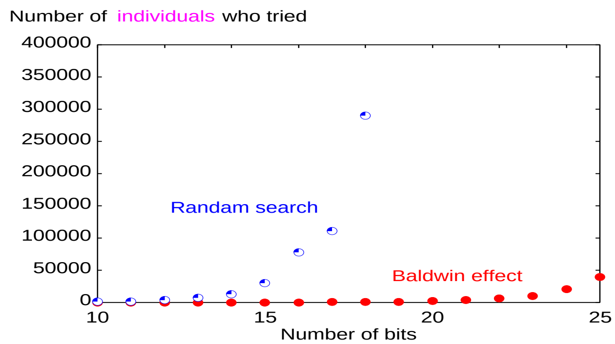


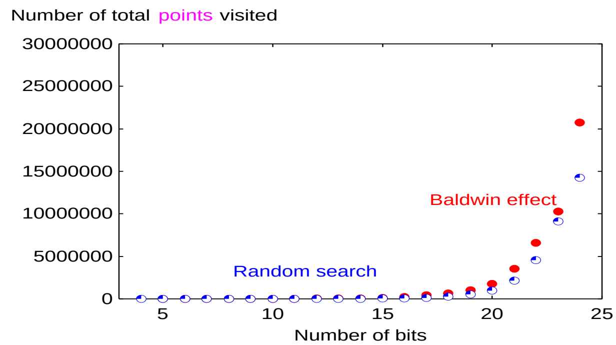
Eventually, all 9's disappeared to converge to the needle itself.

The number of total points searched,

(not the number of individuals,)

is almost similar as a random search.





(Cont'd)

Alas, how to reach the needle longer than 25-bit?

$< II >$

Reduced Even-Parity Boolean Function.

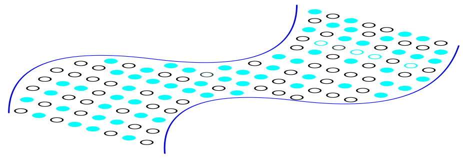
A discussion

(Yu & Miller, 2002)
"Finding Needles in Haystacks is not Hard with Neutrality."

vs.

(Collins, 2005)
"Finding Needles in Haystacks is Harder with Neutrality."

Reduced Even-n-Parity Problem

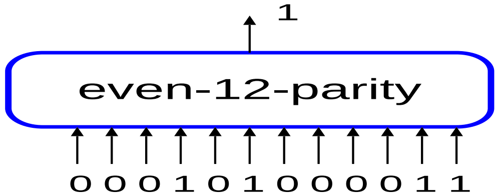


Even-parity Boolean function only with XOR and EQ

$\Downarrow$

A commonly used benchmark in GP, GA etc.

(Cont'd)



with

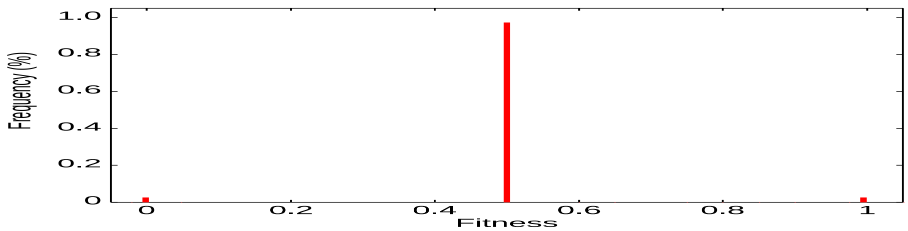
EQ

0 0	1
0 1	0
1 0	0
1 1	1

XOR

0 0	0
0 1	1
1 0	1
1 1	0

A strange fitness distribution



(1) 100%, (2) 50%, or (3) 0% outputs are correct.

⇒ No information of "How good are candidates?"

Yu & Miller's reported success (2002)

by *Cartesian Genetic Programming*

for $n = 12$

48 successes out of 100 runs each with 10,000 iterations
while
none of 4,000,000 randomly created candidates failed

Yu & Miller's assumption

The success is due to what they call a *neutral-mutation-on-intron.*

Yu & Miller's CGP

(1) Create 100 gates of XOR or EQ at random.

(2) Select randomly one output gate out of those 100 created.

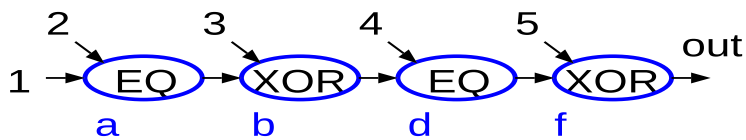
(3) From one gate to the next, set the two input connections to either of the output line previously set gate or one of the n input lines at random.

An example genotype of CGP for Even-5-Parity

a b c d e f

((EQ 1 2)(XOR 3 a) (EQ a 2) (EQ 4 b) (XOR 1 4) (XOR d 5))

⇓



Neutral mutation on introns

Some genes do not contribute to phenotype ⇒ *intron*

And, so does a mutation on intron ⇒ *neutral*

Hence
Size of function is flexible!

But the question is, as they claimed, "Does it enhance the search?"

Collins' doubt (2005)

"Unable to repeat the same performance as reported."

$\Downarrow$

He tried, instead,

"a random sampling which uses a fully expressed genotype."

$\Downarrow$

Better than Yu & Miller's result.

Collins' random sampling

(1) Create 100 random gates either from XOR or EQ.

(2) Set the output to the 100-th gate.

(3) For each gate:

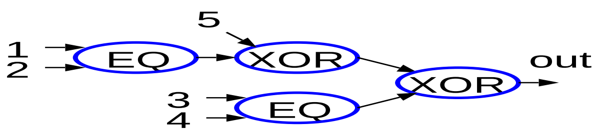
  (i) Set the type of the gate to either XOR or EQ at random.

  (ii) Set one of the input connection of the gate to the previous gate.

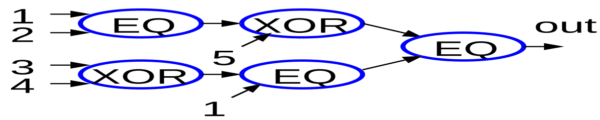
  (iii) Set the other input connection to a randomly selected input line.

(4) Repeat (3) up to N.

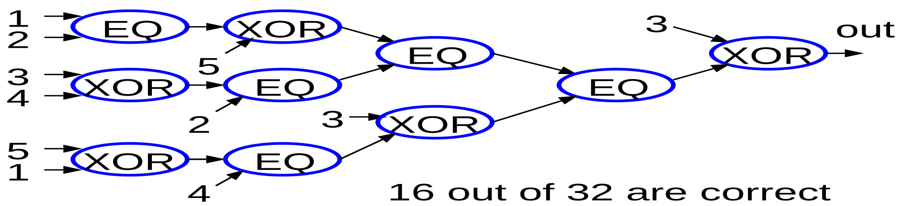
Collins' algorithm excludes many candidates such as...



16 out of 32 are correct



16 out of 32 are correct



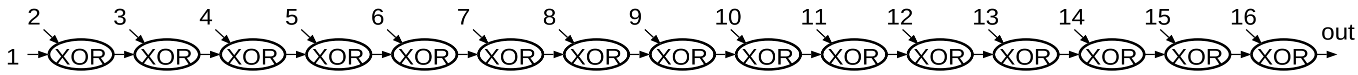
16 out of 32 are correct

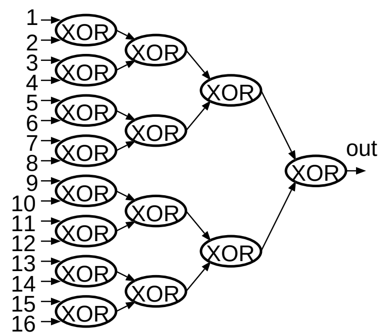
(Cont'd)

Anyway, their solutions found are both limited to $n < 13$.

On the other hand...

We human can create an even-n-parity for however large n.

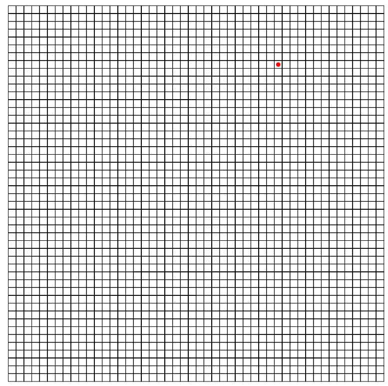




$< III >$

A Robot Navigation in 2-D grid.

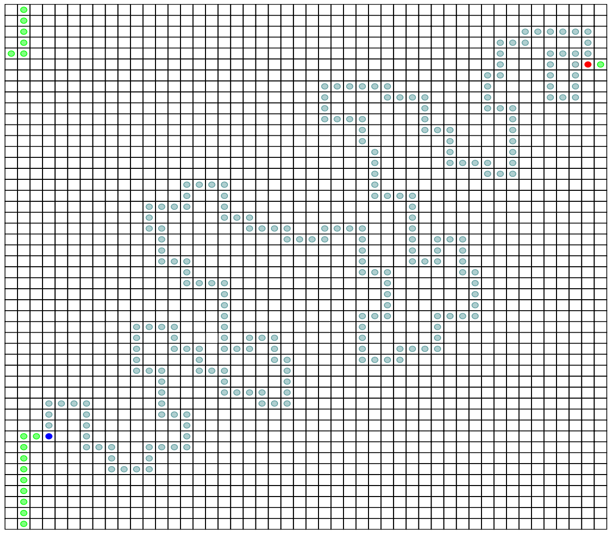
Finding a needle hidden in a huge 2-dimensional grid



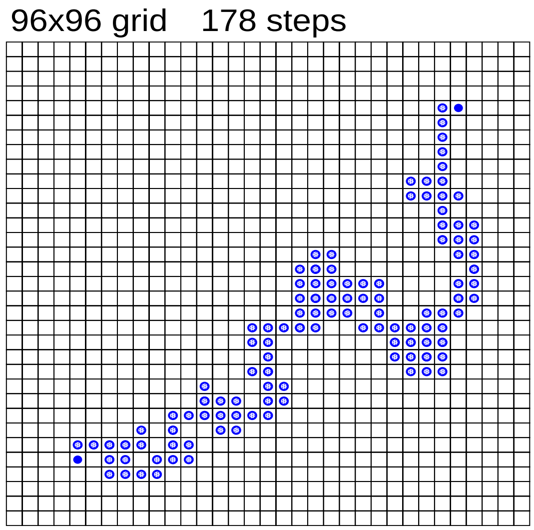
Hinton & Nowlan's scenario

Can parachuters reach the needle in a pastoral
if allowed random walks around their falling place?

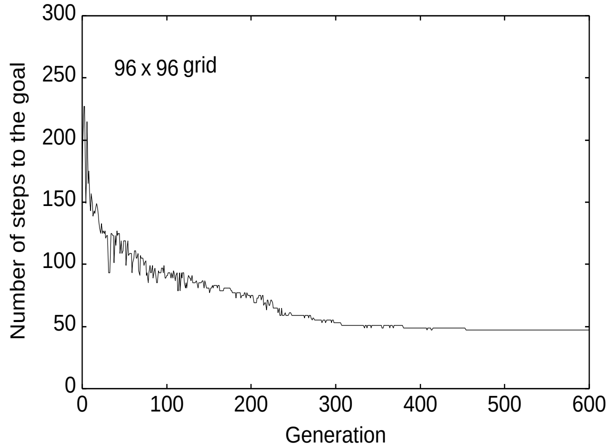

Let's try, here, a random walk, instead

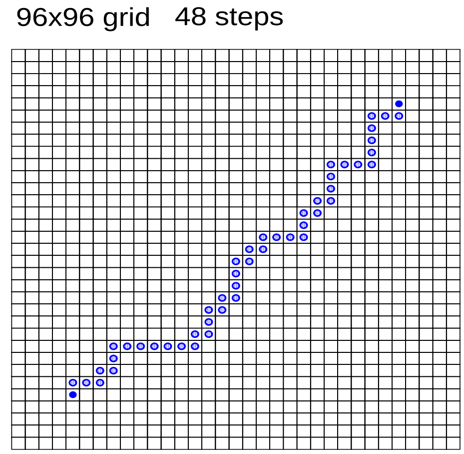


Can random-walker reach the needle?

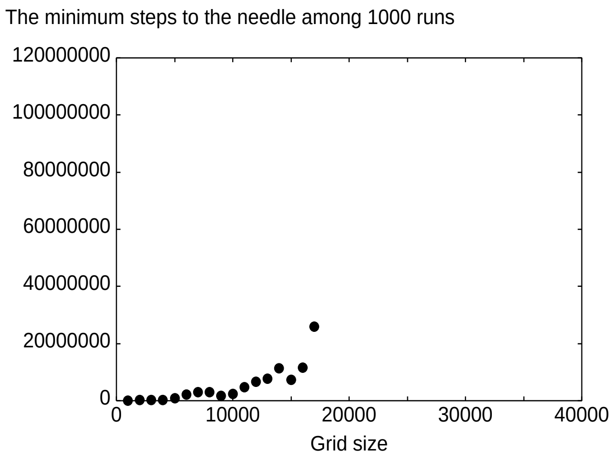


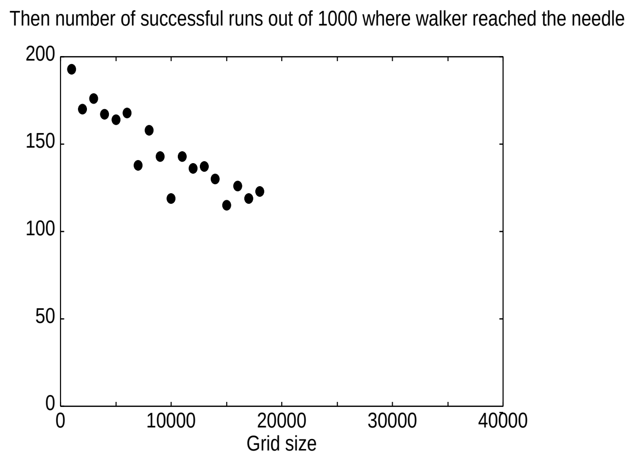
Can a learning enhance efficiency?





A limit in grid size





< IV >

Quantum Computation

Find the needle from no-structured huge database

"Find x such that $P(x) = 1$"

when

only x from N data fulfills $P(x) = 1$ while all others do not.

(Cont'd)

$N/2$ queries on average, and $(N-1)$ on worst case.
are necessary.

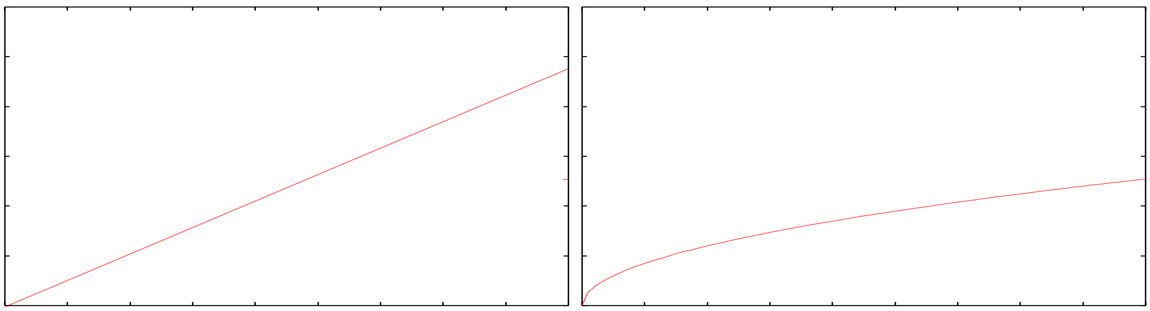
(E.g. imagine breaking a PIN code.)

$\Downarrow$

We need $O(N)$ *random-trial-&-errors* or *one-by-one-searches.*

Grover's quantum search (1997)

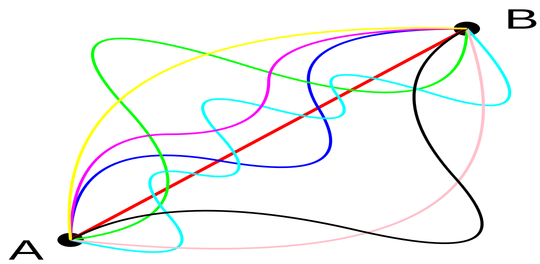
A speed up from $O(N)$ to $O(\sqrt{N})$.

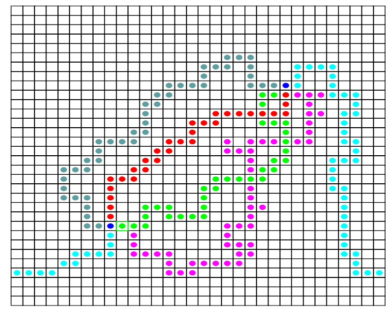


A strange path of quantum computation

When a particle goes from A to B,

it takes all possible paths at the same time.





What are differences?

1.

Measurement cannot be done
before completing computation.

(Cont'd)

2.

A Function, like $P(x)$, is applied

to *superposed* states simultaneously

not a *single* state one-by-one.

(Cont'd)

3.
We must know *when to stop.*

While in a *classical computer*, we could continue the run after conversion, e.g., in GA.

Grover's algorithm

1. Compute $P(x)$ for all possible x.
   (Done simultaneously, and cannot be measured at this moment.)
2. Rotate the state of needle $\pi/2$ radian while not all others.
   (Also done simultaneously.)
3. Repeat 1. - 2. $\pi\sqrt{N}/8$ iterations.
   (One must know when to stop.)
4. Measure all states.
   (The needle is measured with a high probability close to 1.)

Bit & Qubit

classical computer (1 0 1 1)

vs.

quantum computer (|1> |0> |1> |1>)

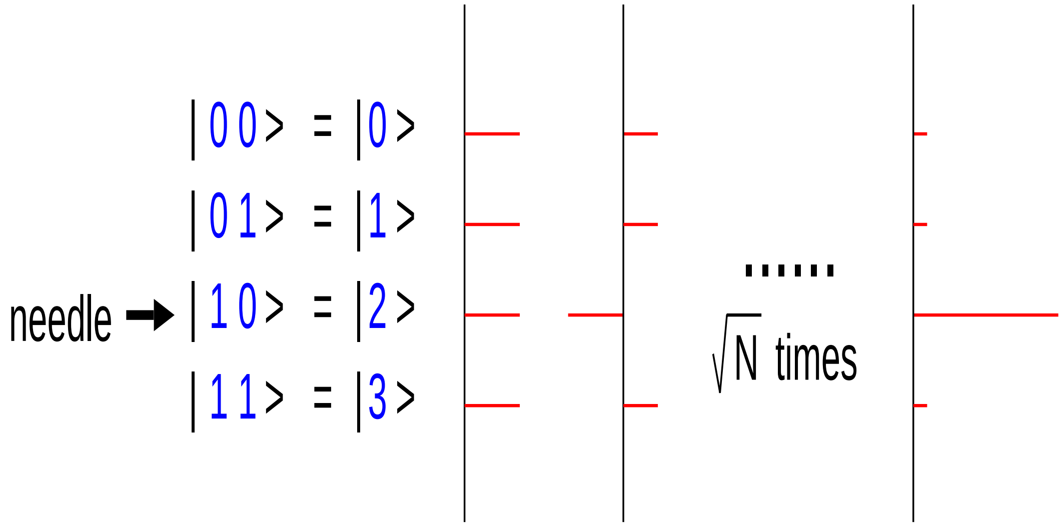
Qubit is a superposition of two basis

$$|0\rangle = a|0\rangle + b|1\rangle$$

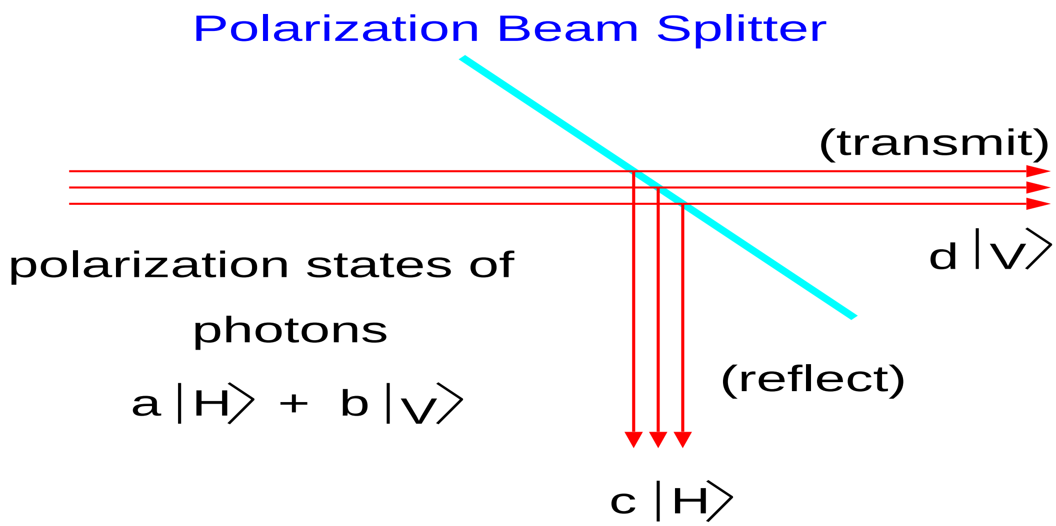
and

$$|1\rangle = c|0\rangle + d|1\rangle$$

where $a, b, c, d \in Z$, and e.g., $\|c\|^2$ is a probability for $|1\rangle$ to be $|0\rangle$.


needle
|00> = |0>
|01> = |1>
|10> = |2>
|11> = |3>
......
√N times

A toy implementation by photons



A quantum walk on the line

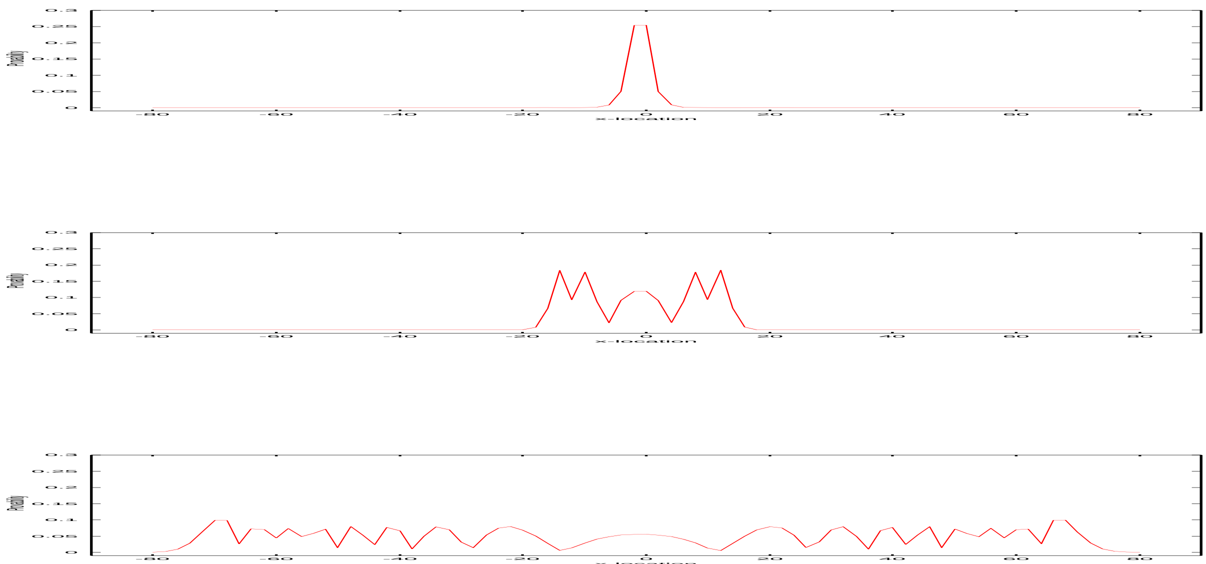
$$p(n,t) = \|\psi_L(n,t)\|^2 + \|\psi_R(n,t)\|^2$$

where

$$\psi_L = \frac{1+(-1)^{n+t}}{2} \int_{-\pi}^{\pi} \frac{dk}{2\pi} (1 + \frac{\cos k}{\sqrt{1+\cos^2 k}}) \exp\{-i(\omega_k t + kn)\}$$

$$\psi_R = \frac{1+(-1)^{n+t}}{2} \int_{-\pi}^{\pi} \frac{dk}{2\pi} \frac{\exp(ik)}{\sqrt{1+\cos^2 k}}) \exp\{-i(\omega_k t + kn)\}$$

A propagation of wave function



Concluding Remarks

- Searching for a needle in a hay is open & important issue.
- We need real efficient approaches.
- Not yet so far a real efficient one.

- We have to avoid an effect of
  like-to-hear-what-we-would-like-to-hear.
  as all of today's cases seem to have it more or less.