HITACHI
Inspire the Next

An Ising Computing to Solve Combinatorial Optimization Problems

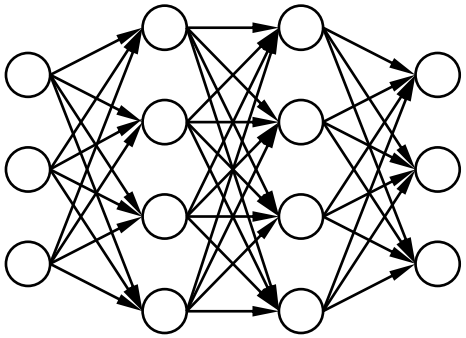
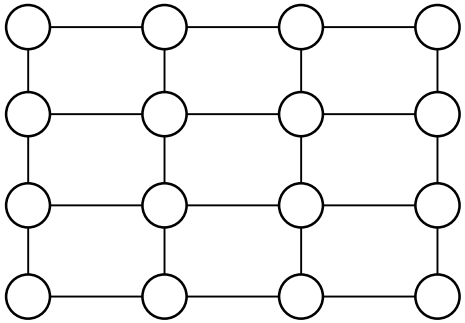
Masanao Yamaoka
Research & Development group, Hitachi, Ltd., Tokyo, Japan

10/19/2017
5th Berkeley Symposium on Energy Efficient Electronic Systems & Steep Transistors Workshop

Difference between NN and Ising model

HITACHI
Inspire the Next

- Nodes connected by bi-directional edges
- Optimization problem is main target

	Neural network	Ising model (Boltzman machine)
		
Connection	Directional (feedfoward NN)	Bi-directional
Purpose	Learning	Combinatorial optimization problems

Outline

HITACHI
Inspire the Next

- Background
- New-paradigm computing
- CMOS Ising computer
- Measurement results
- For practical application

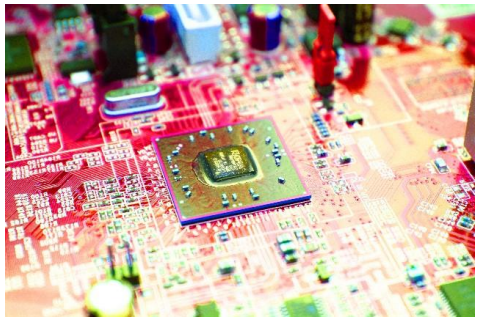
Importance of optimization

HITACHI
Inspire the Next

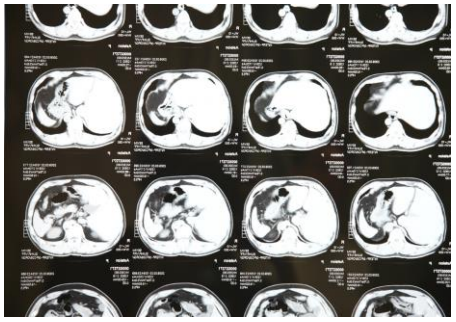
- Optimization processing used in various industries
- In IoT era, data volume large in both edge and cloud

Example areas that needs optimization processing

Logistics operation



VLSI design



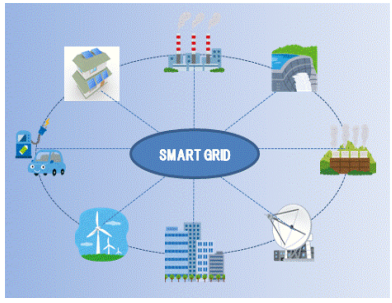
Medical diagnosis with images

Management strategy

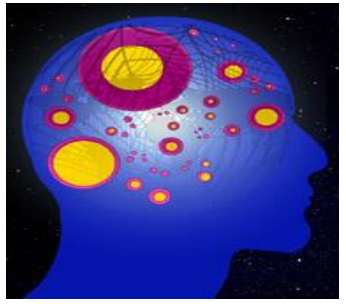
Route selection

Scheduling at wide-scale disaster

Learning plan customizing



Smart-grid control

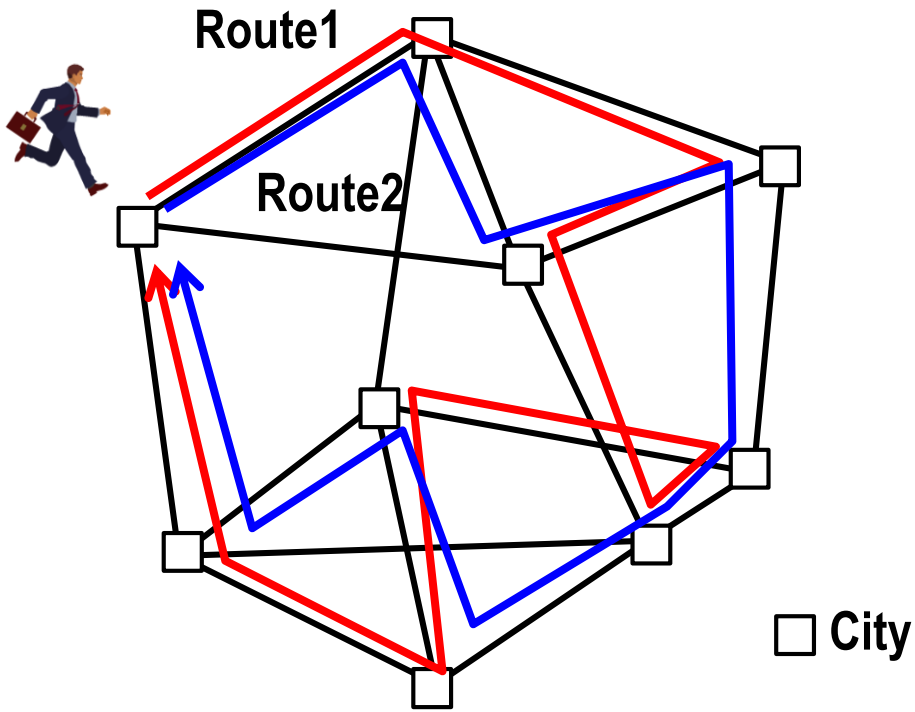


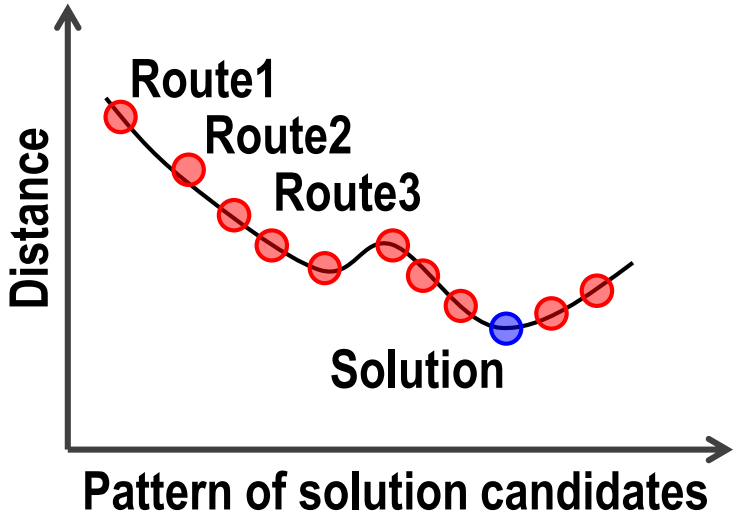
Robotics control

Combinatorial optimization problem (1)

HITACHI
Inspire the Next

- Example: travelling salesman problem
- Searching the shortest route that visits each city exactly once and returns to the origin city with list of cities and distances between cities

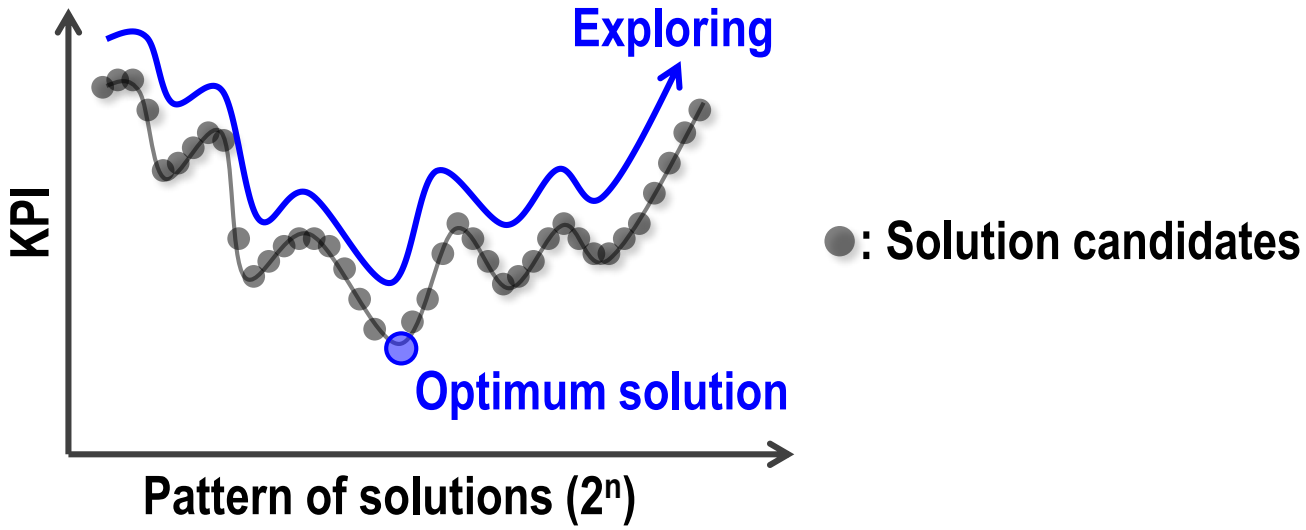




Combinatorial optimization problem (2)

HITACHI
Inspire the Next

- Problem to explore an optimum solution for minimum or maximum KPI in given conditions
- Difficult to solve by conventional computers due to enormous candidates of solution with large number of parameters

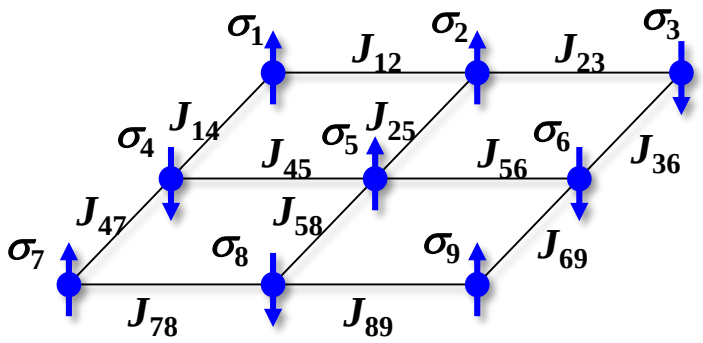


KPI: Key Performance Indicator

Ising model

HITACHI
Inspire the Next

- Ising model: expressing behavior of magnetic spins, upper or lower directions
- Spin status updated by interaction between spins to minimize system energy



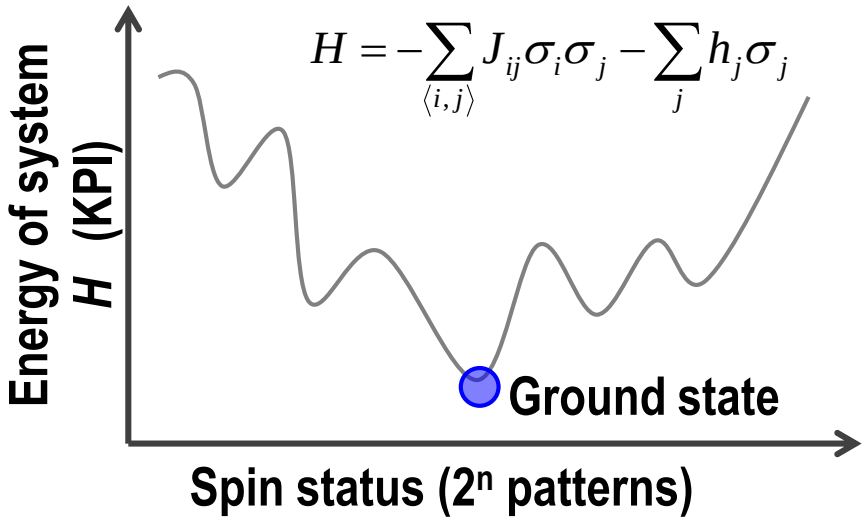
$$H = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i\sigma_j - \sum_j h_j\sigma_j$$

- H: Energy of Ising model
- σ_i: Spin status (+1/-1)
- J_{ij}: Interaction coefficient
- h_j: External magnetic coefficient

Computing with Ising model

HITACHI
Inspire the Next

- Shape of landscape of Ising model energy same as KPI plot of combinatorial optimization problem
- By mapping original problems to Ising model, optimum solution acquired as ground state of model



n: number of spins

Correspondence of parameters

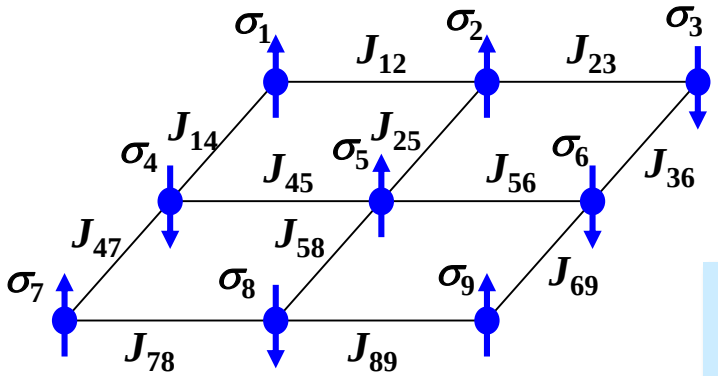
Ising model	Optimization problems
Energy H	KPI
Spin status σ_i	Control parameters
Interaction coefficient J_{ij}	Input data (sensor data, etc)

CMOS Ising computing

HITACHI
Inspire the Next

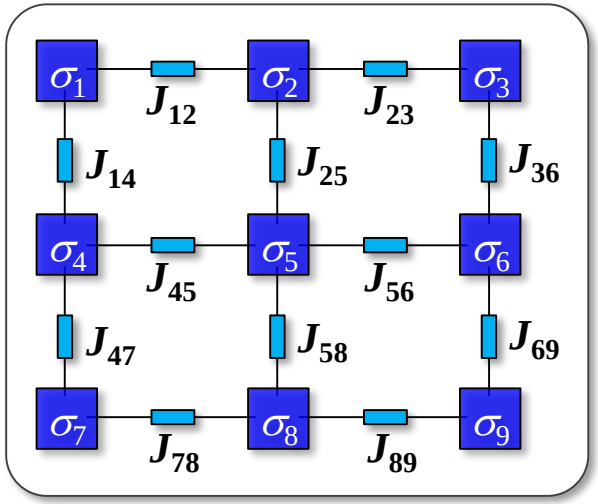
- Mimicking Ising model with CMOS circuits
- Spin status updated by interaction with adjacent spins

Ising model



CMOS circuit mimicking Ising model

CMOS circuits



Ising model	CMOS
Spin status σ_i, +1/-1	Memory "1"/"0"
Interaction J_{ij}	Interaction coefficient: memory Interaction operation: digital circuits

Rule of spin status update

HITACHI
Inspire the Next

- Spin status updated to lower Ising model energy
- Coefficient +: same direction
  Coefficient - : opposite direction
- Majority of adjacent spins effect accepted

$$H = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i\sigma_j - \sum_j h_j\sigma_j$$

Spin status:
$J_{ij} > 0$: same direction
$J_{ij} < 0$: opposite direction

Spin update rules

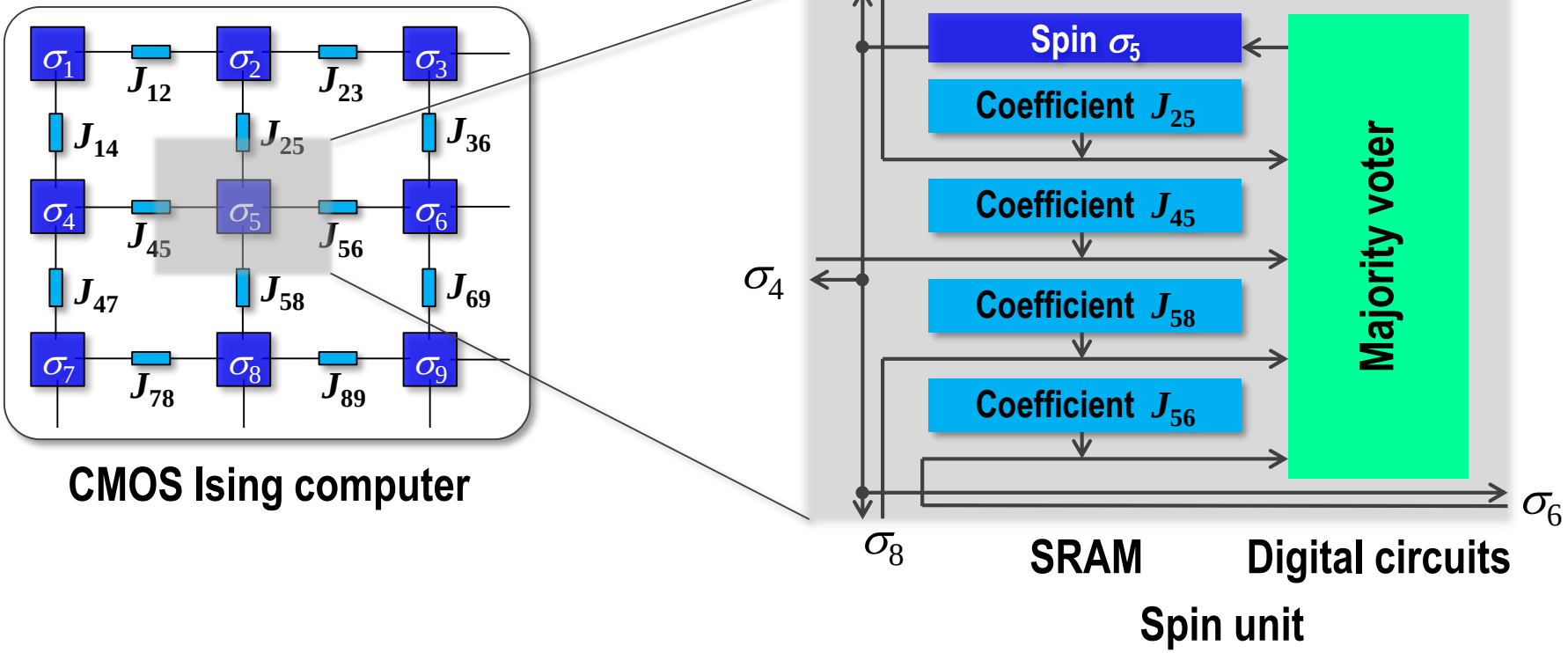
Next spin status:
in case a>b, σ_5=+1
in case a<b, σ_5 =-1
in case a=b, σ_5 =+1 or -1

a=number of (+1, +1) or (-1, -1)
b=number of (+1, -1) or (-1, +1)
(value from adjacent spin, coefficient)

Operation of CMOS Ising computer

HITACHI
Inspire the Next

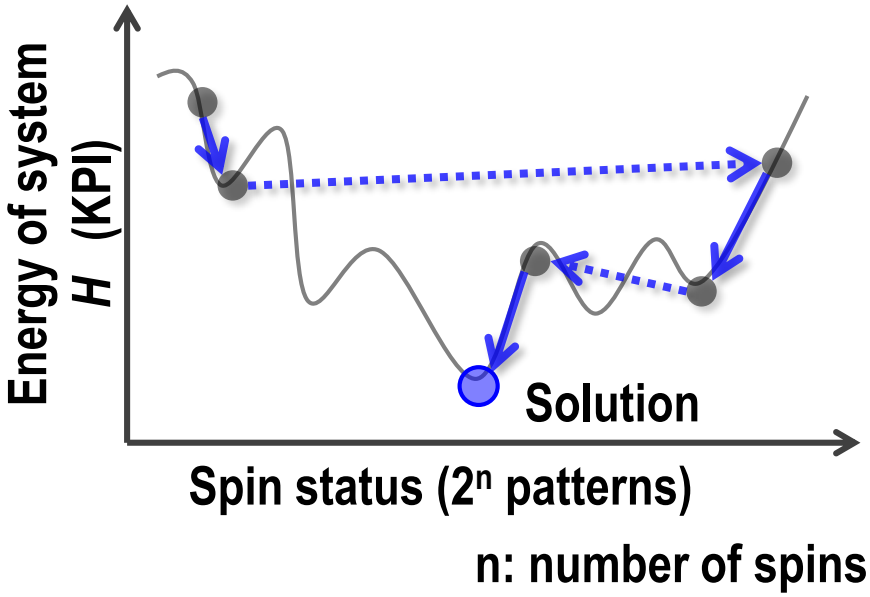
- Spin-update rule achieved by majority voter
- Each spin has update circuit and updated in parallel
- Short calculation time even when number of spins is large



CMOS annealing

HITACHI
Inspire the Next

- Only using adjacent spin interaction, spin status stuck at local minimum status
- To avoid local minimum sticking, random status transition used
- Optimum solution not always acquired



■ Transition to lower energy (adjacent spin interaction) →

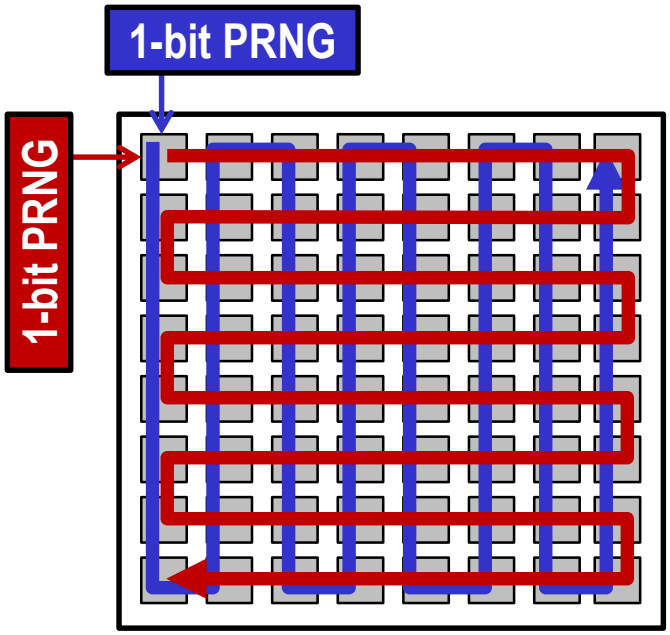
■ Avoidance of local minimum (random transition) ⇢

Random transition: random numbers

HITACHI
Inspire the Next

- Two sequences of 1-bit random numbers input, propagated, and evaluated at spin interaction
- Only two PRNGs provide randomness to whole chip

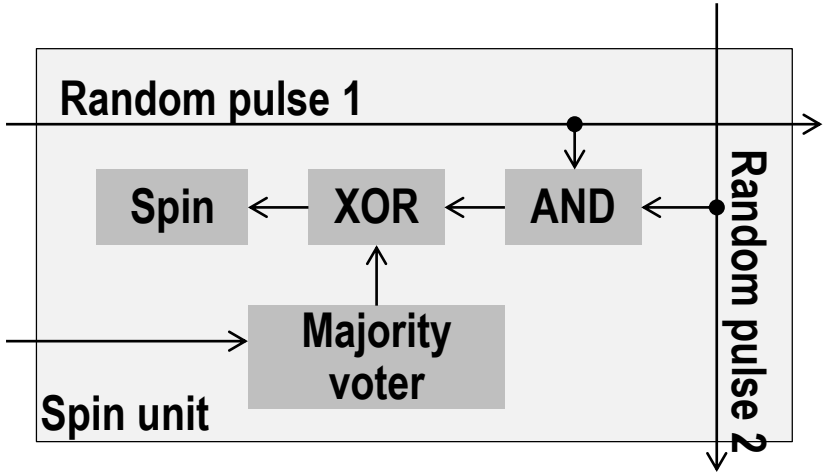
Zigzag paths



PRNG: Pseudo Random Number Generator

Spin inversion using random pulse

- Output of the interaction circuit will be inverted if both of the random pulses are '1'



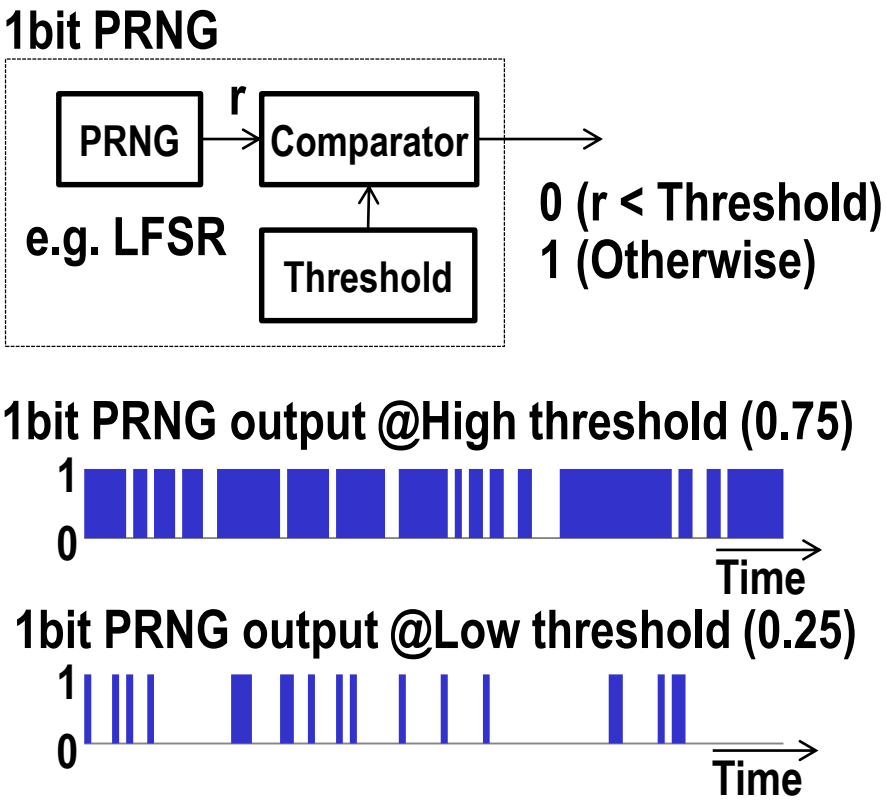
M. Hayashi et al., "An Accelerator Chip for Ground-state Searches of the Ising Model with Asynchronous Random Pulse Distribution," 6th International Workshop on Advances in Networking and Computing

Random pulse control

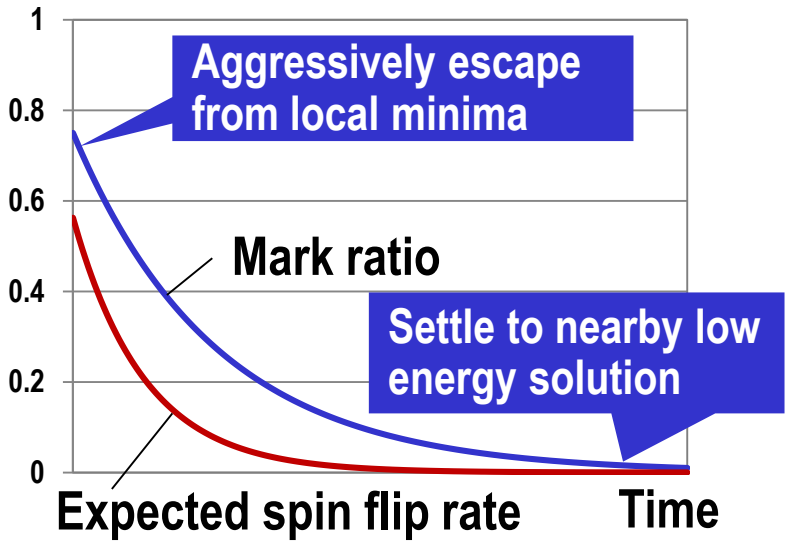
HITACHI
Inspire the Next

- Spin flip rate gradually lowered for annealing
- Spin flip rate controlled by mark ratio of 1bit PRNGs

Controlling mark ratio of 1bit PRNG



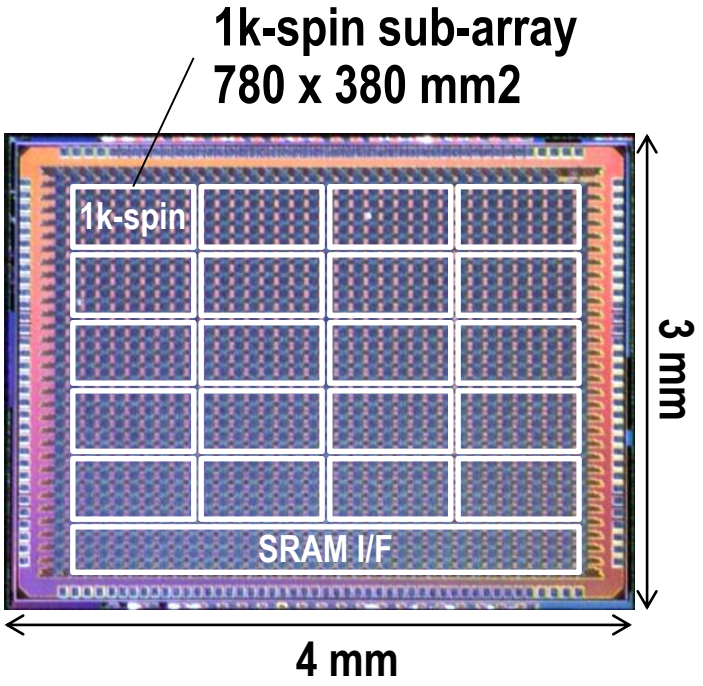
Mark ratio during ground-state search



➢Decreasing mark ratio has similar effect to cooling schedule in Simulated Annealing

Fabrication results: Ising chip

HITACHI
Inspire the Next



Items	Values
Number of spins	20k (80 x 128 x 2)
Process	65 nm
Chip area	4x3=12 mm^2
Area of spin	11.27 x 23.94 =270 μm^2
Number of SRAM cells	260k bits Spin value: 20k bits Interaction coefficient: 240k bits
Memory IF	100 MHz
Interaction speed	100 MHz
Operating current of core circuits (1.1 V)	Write: 2.0 mA Read: 6.0 mA Interaction: 44.6 mA

1st generation prototype

HITACHI
Inspire the Next

- Ising chips installed on computing node
- FPGA installed to control Ising chip
- Accessed via LAN cable from PCs/servers

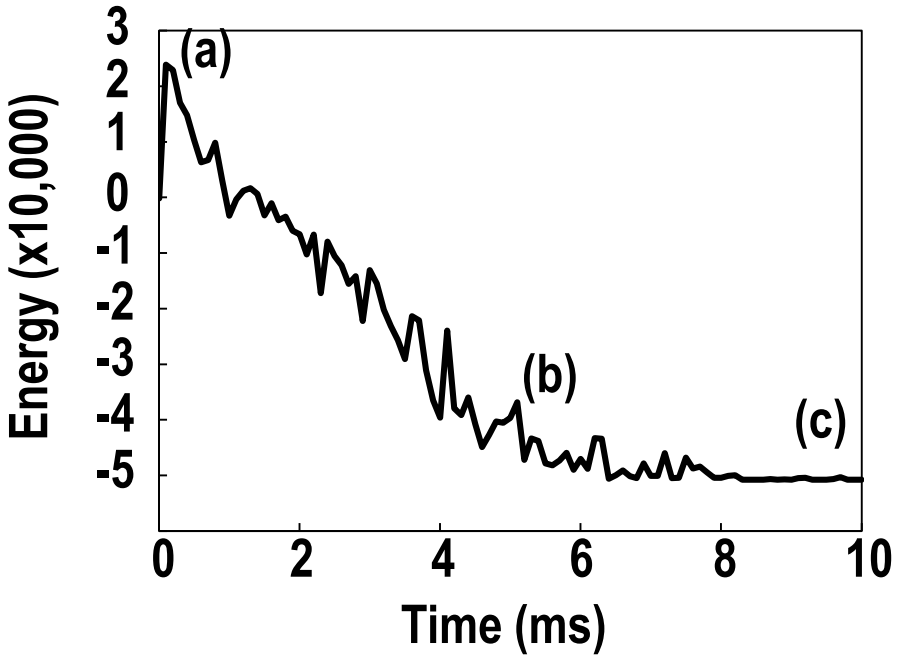
Computing node

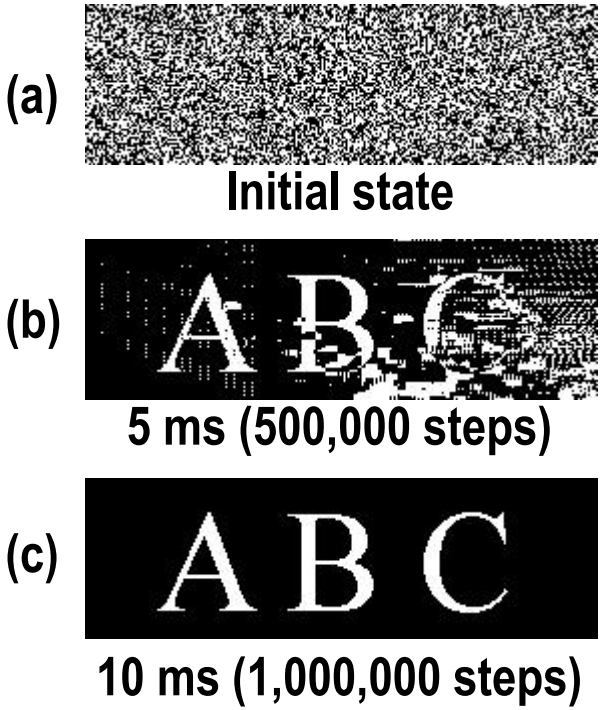
Configuration	2U rack mount
Operation frequency	100 MHz
Number of spins	40k (2chips)
OS	Linux

Measurement results with random numbers

HITACHI
Inspire the Next

- MAX-cut problem, NP-complete problem, with 20k-spin solved
- Coefficient values set "ABC" appeared at optimum
- Optimum solution not always acquired

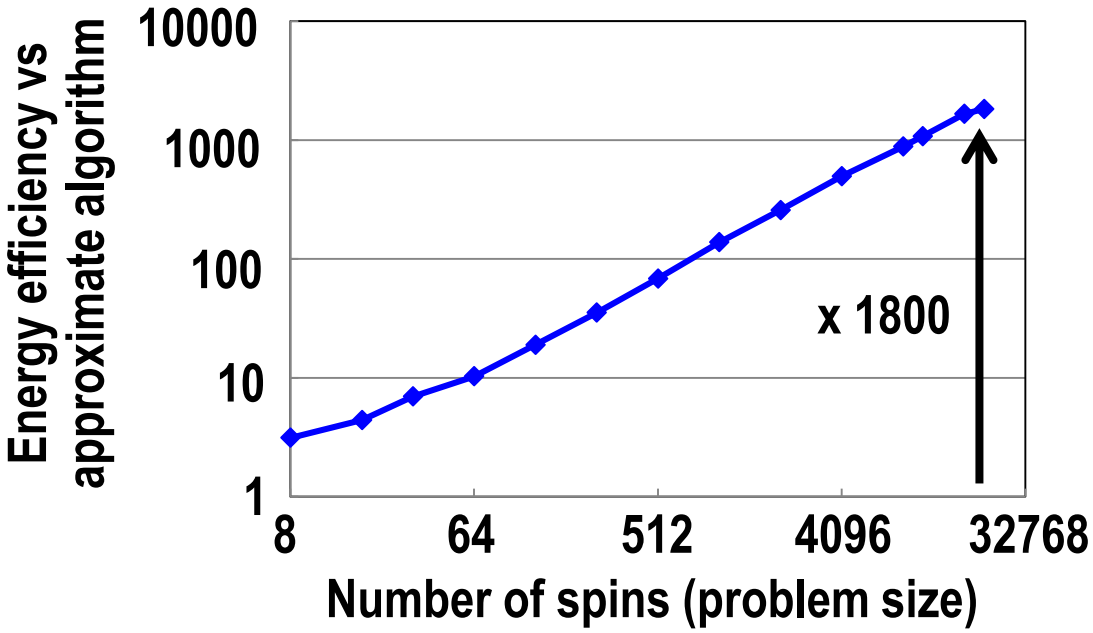




Operating energy

HITACHI
Inspire the Next

- 1,800 times higher energy efficiency for 20k spin problem than approximation algorithm on CPU



Conditions:

Randomly generated Maximum-cut problems, energy for same accuracy solution
Ising chip: VDD=1.1 V, 100-MHz interaction, best solution among 10-times trial is selected.
Approximation algorithm: SG3(*) is operated on Core i5, 1.87 GHz, 10 W/core.

(*): Sera Kahruman et al., "On Greedy Construction Heuristics for the Max-Cut Problem," International Journal on Computational Science and Engineering, Volume 3, Number 3/2007, pp. 211-218, 2007.

Comparison of Ising computers

HITACHI
Inspire the Next

- Room-temperature operation for easy use
- Higher scalability by CMOS process

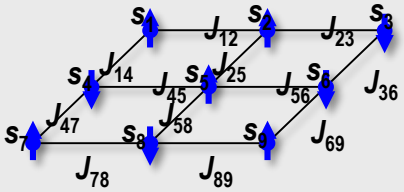
	CMOS annealing	Quantum annealing [1]
Approach	Ising computing	
	Semiconductor (CMOS)	Superconductor
Coefficient value	2 bit (+/- 1 or 0)	4 bit
Accuracy	Sufficient for application	Better
Temperature	Room temperature	15 mK
Power	0.05 W	15,000 W with cooling
Scalability (Num. of spins)	20k (65nm) Scalable with CMOS	2048

[1] http://www.dwavesys.com/

How to solve "real" problems

HITACHI
Inspire the Next

● Development of application and software techniques required for Ising computer

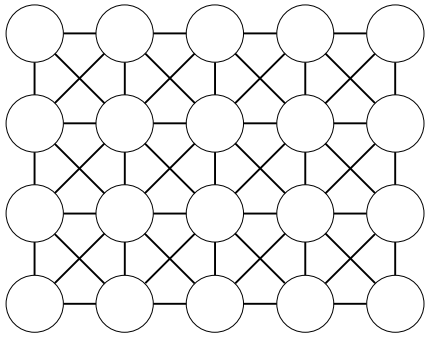
Systems	Traffic	Supply chain	Power grid
Application	Reduction of traffic jam	Reduction of logistics cost	Stable power supply
Software	↓	Mapping to Ising-model energy functions	↓
Problem Mapping	Shortest path problem $H(\sigma)$ = ...	Traveling salesman $H(\sigma)=\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{a=1}^{N}D_{ij}\sigma_{ia}\sigma_{j,a+1 \bmod N}$ $+\gamma\left[\sum_{a=1}^{N}\left(1-\sum_{i=1}^{N}\sigma_{ia}\right)^2+\sum_{i=1}^{N}\left(1-\sum_{a=1}^{N}\sigma_{ia}\right)^2\right]$	Max-flow problem $H(\sigma)$ = ...
	↓	Embedding to Ising hardware	↓
Graph embedding	Ising model on computer		J_{12}=x, J_{23}=y, ...
Hardware	Ising computer		

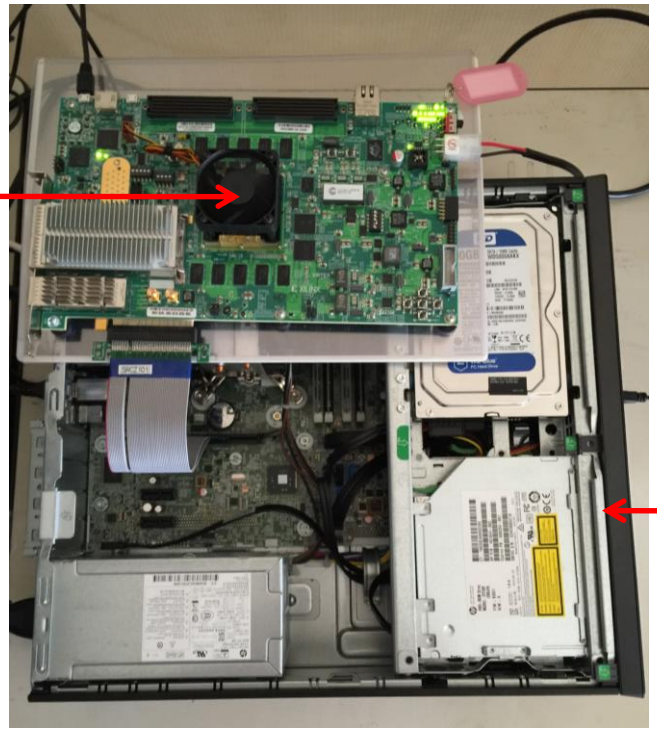
2nd generation prototype with FPGA

HITACHI
Inspire the Next

- For application and software development, 2nd generation FPGA prototype used
- Various structures for trials (various topology, various bit number of coefficient)

Reconfigurable FPGA used to trial various structures (King's graph)

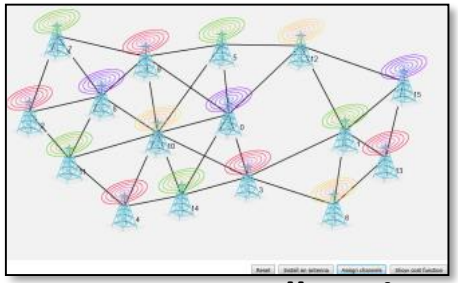




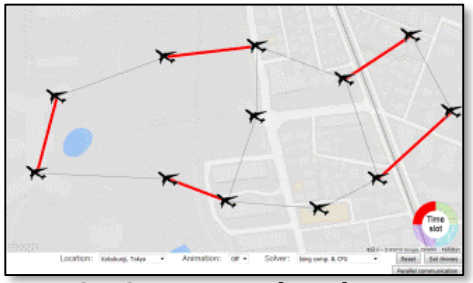
Control PC
- FPGA control
- Software

Demo of CMOS Ising computer

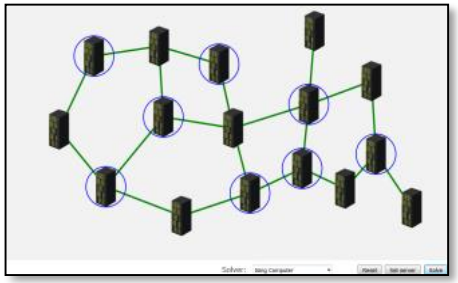
HITACHI
Inspire the Next



1. Frequency allocation for wireless radio



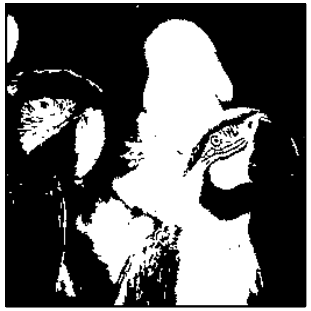
2. Communication order allocation



3. Server security

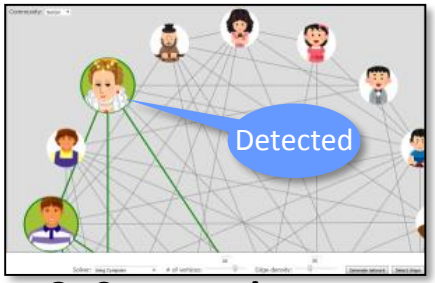
4. Image inpainting

5. Exploring explosion material

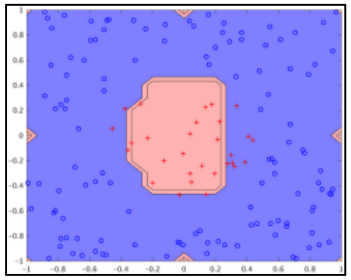


6. Noise reduction

7. Facility allocation



8. Community core detection

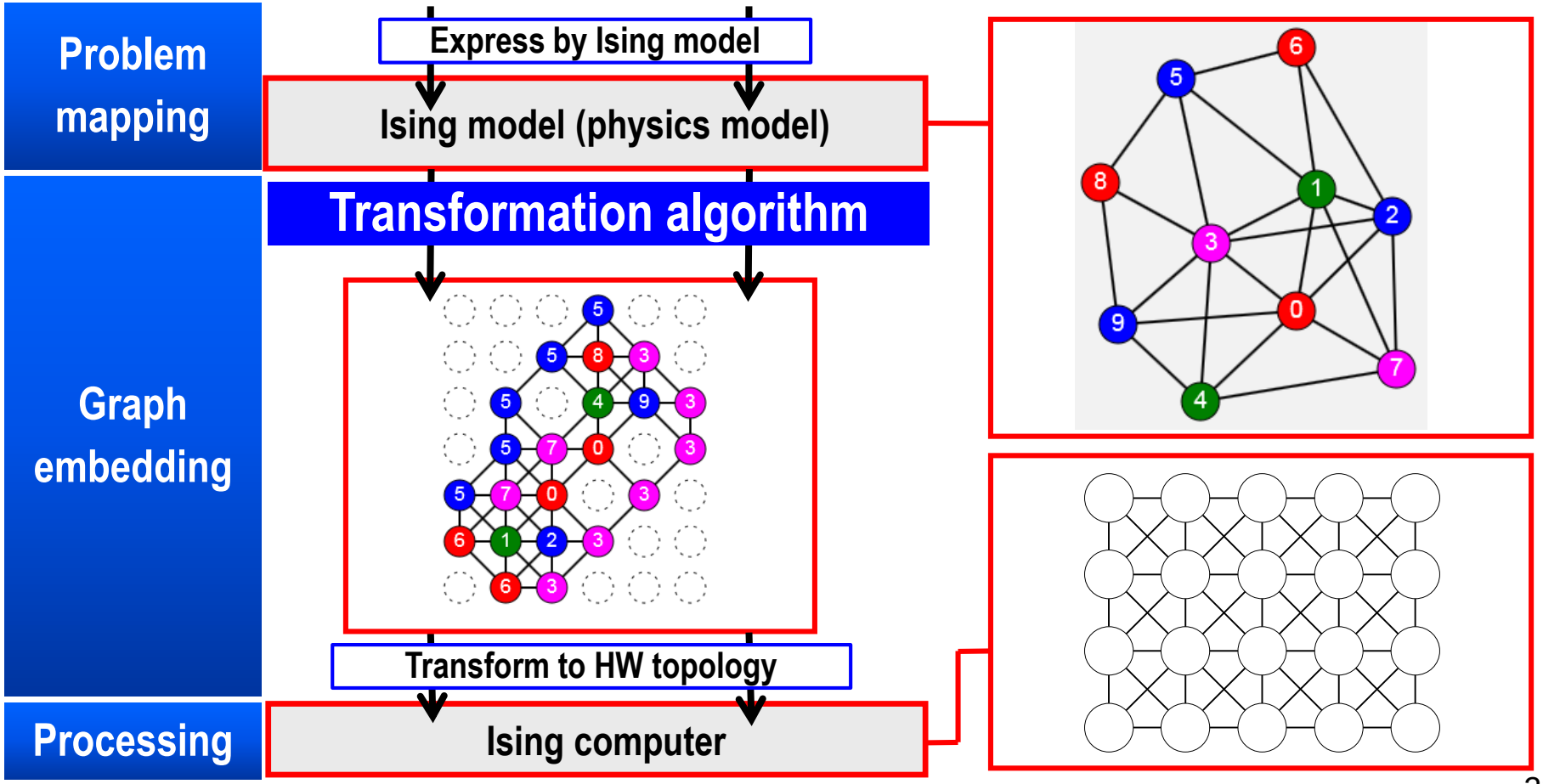


9. Machin learning (Boosting)

Graph embedding (software technique)

HITACHI
Inspire the Next

- Complex graph embedded to simple graph
- Graph embedding to use hardware efficiently



Summary

HITACHI
Inspire the Next

- New computing is necessary for system optimization with large amount of data.
- CMOS Ising computing for combinatorial optimization problem is proposed.
- 20k-spin prototype chip achieves 1,800 times higher power efficiency.
- Software techniques and related hardware techniques are necessary for practical application.

HITACHI

Inspire the Next