Genome Wide Epigenetic Modifications as a Shared Memory Consensus Problem

Sabrina Rashid[1], Gadi Taubenfeld[2], Ziv Bar-Joseph[1,3]

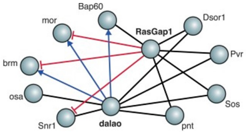
[1] Computational Biology Department, Carnegie Mellon University.
[2] School of Computer Science, The Interdisciplinary Center.
[3] Machine Learning Department, Carnegie Mellon University.

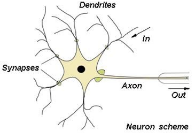
Motivation

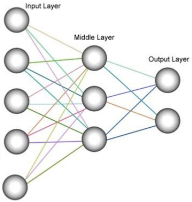
- Like large-scale computing platforms, cellular/ molecular systems are mostly distributed consisting of entities that interact, coordinate, and reach decisions without central control.

Motivation

- Like large-scale computing platforms, cellular/ molecular systems are mostly distributed consisting of entities that interact, coordinate, and reach decisions without central control.
- However, most prior biologically inspired distributed computing methods rely on message passing.







Objective

- Here we model the process of genome wide epigenetic modifications as a shared memory system.

Objective

- Here we model the process of genome wide epigenetic modifications as a shared memory system.
- Formulate a particular consensus problem for the shared memory model from DNA epigenome.

Objective

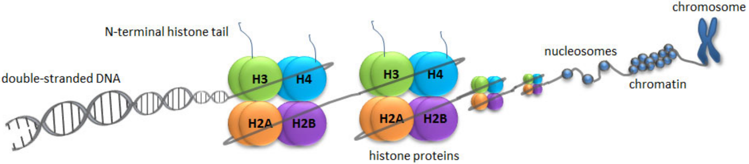
- Here we model the process of genome wide epigenetic modifications as a shared memory system.
- Formulate a particular consensus problem for the shared memory model from DNA epigenome.
- Present possible algorithms to achieve consensus with biologically plausible assumptions.

Objective

- Here we model the process of genome wide epigenetic modifications as a shared memory system.
- Formulate a particular consensus problem for the shared memory model from DNA epigenome.
- Present possible algorithms to achieve consensus with biologically plausible assumptions.
- Prove convergence of the algorithms both analytically and in simulations.

Epigenetics

- Epigenetics refer to the post translational modifications of the histone proteins on which the DNA is wrapped [1].

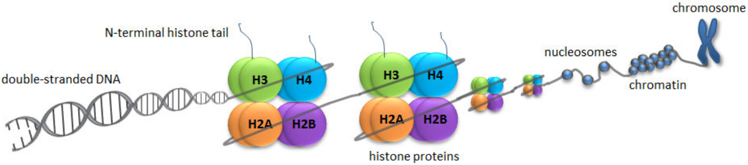


---

[1] Figure courtesy: whatisepigenetics.com

Epigenetics

- Epigenetics refer to the post translational modifications of the histone proteins on which the DNA is wrapped [1].



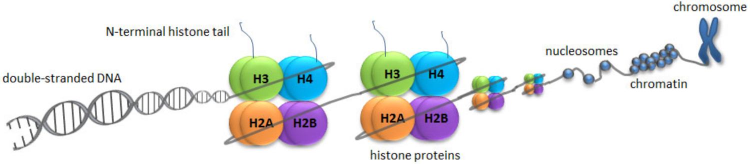
- Such modifications play an important role in regulating gene expression and chromatin states.

---

[1] Figure courtesy: whatisepigenetics.com

Epigenetics

- Epigenetics refer to the post translational modifications of the histone proteins on which the DNA is wrapped [1].



- Such modifications play an important role in regulating gene expression and chromatin states.
- Hence these modifications are highly regulated and consistent across large stretches of the genome.

---

[1] Figure courtesy: whatisepigenetics.com

Epigenetic Modifiers

- Histone modifiers can be broadly categorized into three classes [2]:
  - Readers
  - Writers
  - Erasures

---

[2] Craig L Peterson and Marc-Andre Laniel. Histones and histone modifications. Current Biology, 14(14):R546R551, 2004.

Epigenetic Modifiers

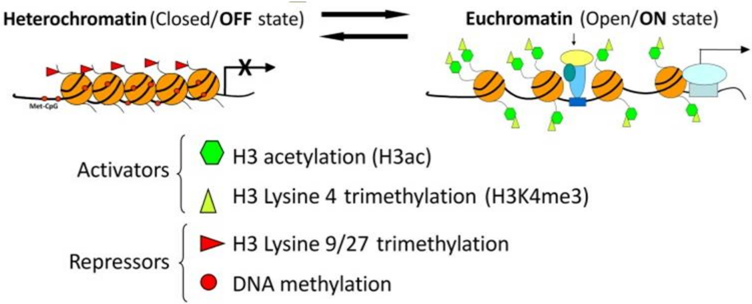
- Histone modifiers can be broadly categorized into three classes [2]:
  - Readers
  - Writers
  - Erasures

---

[2] Craig L Peterson and Marc-Andre Laniel. Histones and histone modifications. Current Biology, 14(14):R546R551, 2004.

Epigenetic Modifiers

- Histone modifiers can be broadly categorized into three classes [2]:
  - Readers
  - Writers
  - Erasures



---

[2] Craig L Peterson and Marc-Andre Laniel. Histones and histone modifications. Current Biology, 14(14):R546R551, 2004.

Modeling Assumptions

- Restrict histone modifications to two types 0 or 1: writers (W0/W1) and erasers (E0/E1).

Modeling Assumptions

- Restrict histone modifications to two types 0 or 1: writers (W0/W1) and erasers (E0/E1).
- Regulation of the modifiers results in the same histone marks appearing for a stretch of DNA, i.e., *local consensus.*

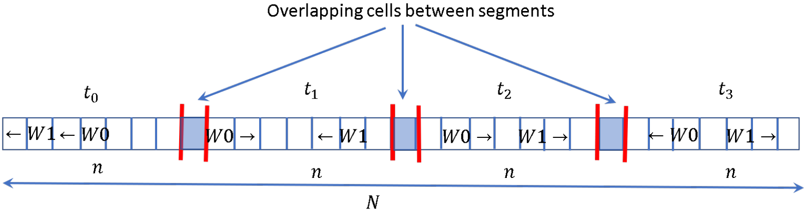
Modeling Assumptions

- Restrict histone modifications to two types 0 or 1: writers (W0/W1) and erasers (E0/E1).
- Regulation of the modifiers results in the same histone marks appearing for a stretch of DNA, i.e., *local consensus.*
- Stretch of DNA region with N total histones $\leftrightarrow$ shared memory array of size N.

Modeling Assumptions

- Restrict histone modifications to two types 0 or 1: writers (W0/W1) and erasers (E0/E1).
- Regulation of the modifiers results in the same histone marks appearing for a stretch of DNA, i.e., *local consensus.*
- Stretch of DNA region with N total histones ↔ shared memory array of size N.
- $N_i \in \{V(Empty), 0, 1\}$
- State transition rules:
    - Allowed: V to 0 (1) by a W0 (W1) writer.
    - Allowed: 0 (1) to V by an E0 (E1) eraser.
    - Not allowed: 0 (1) to 1 (0).

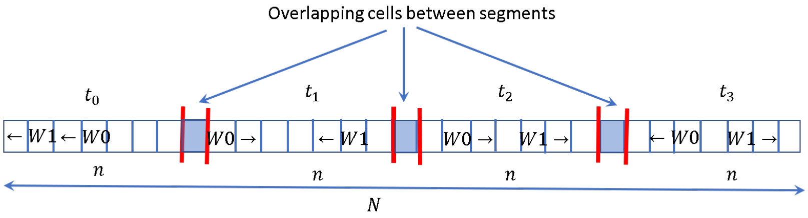
Shared Memory Array



---

[3] Zhonghui Tang et al. Ctcf-mediated human 3d genome architecture reveals chromatin topology for transcription. Cell, 163(7):16111627, 2015

Shared Memory Array

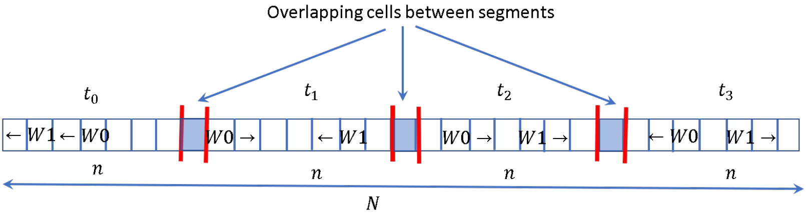


- In biology, this is achieved using CTCF gene and the fact that proteins often scan DNA in a local manner [3].

---

[3] Zhonghui Tang et al. Ctcf-mediated human 3d genome architecture reveals chromatin topology for transcription. Cell, 163(7):16111627, 2015

Shared Memory Array



- In biology, this is achieved using CTCF gene and the fact that proteins often scan DNA in a local manner [3].
- Multiple W0 (W1) writers/ E0 (E1) erasers could get assigned to the same segment.

---

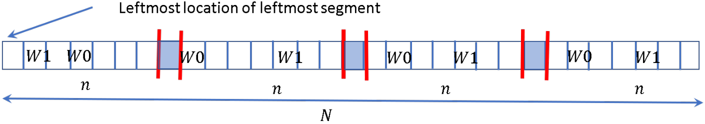
[3] Zhonghui Tang et al. Ctcf-mediated human 3d genome architecture reveals chromatin topology for transcription. Cell, 163(7):16111627, 2015

Consensus Write-erase Problem

- The requirements of the consensus write-erase problem are that there exist a *decision value* $v \in \{0, 1\}$ such that,
  - *Agreement*: The value of each one of the N cells is eventually v, and does not change thereafter.
  - *Validity*: At least one v-writer exists.
  - The solution must be symmetric, i.e., the solution should not favor one of the two possible decision values.

Naïve Solution

- Writers of the leftmost segment compete to write the leftmost cell of the segment (assume v is the written value).
- The v-writers continue writing v into all the cells of the leftmost segment.
- $v-$writers of the other segments wait until leftmost overlapped cell of their segment is written.
- Continue writing the rest of the segments.



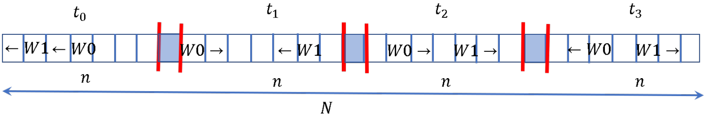
Naïve Solution

- Assume a priori agreement on which side is the left side.
- Inefficient O(N) solution. Erasers do not participate.

Proposed Solution

The algorithm is inspired by the "Game of Life". Let $v \in \{0, 1\}$.

- Starting location and direction of traveling for each processor is chosen randomly.



- *Rule for a v-writer*:
  - Sees an empty cell, it writes v and moves on to the next cell.
  - Otherwise, sees a non-empty cell, it moves on to the next cell.
- *Rule for a v-eraser*:
  - Sees the value v which is preceded by the value $1 - v$ (*collision*), it erases the v and moves on.
  - Otherwise, it just moves on.

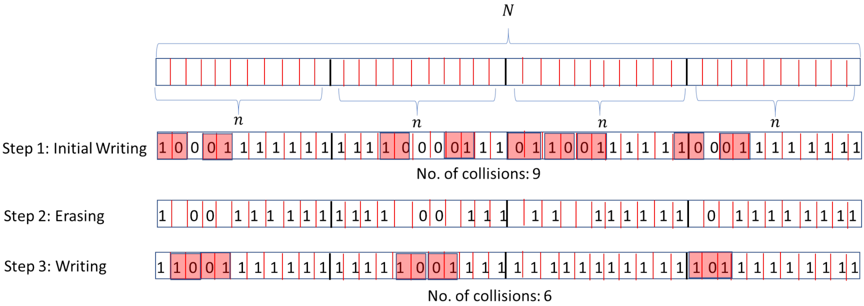
Alternate Solution

In the full paper, we also consider a more efficient variant of this algorithm in which writers must spin (wait) when a collision is noticed.

Analytical and Simulation Results

- Algorithm proceeds in rounds, a write step followed by an erase step is denoted as round.
- To prove convergence, we compute the expected number of collisions in each round.
- We show that the expected number of collisions decreases in each round given $|W0| \neq |W1|$, hence guaranteeing convergence.

Execution Round



Step 1: Initial Writing

No. of collisions: 9

Step 2: Erasing

Step 3: Writing

No. of collisions: 6

Analytical and Simulation Results

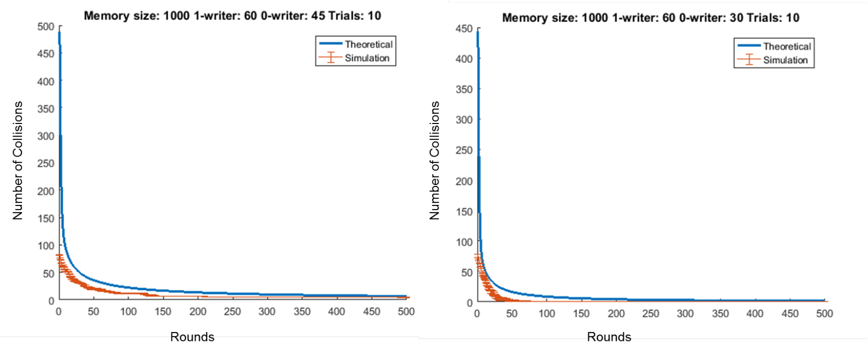


Figure: Plots of expected number of collisions (Theoretical) and number of collisions from simulation vs execution rounds. a) High competition among the processors, $|W1|/|W0| = 1.33$, b) Low competition among the processors, $|W1|/|W0| = 2$.

Summary

- We formulated a consensus write-erase problem for distributed shared memory inspired from genome wide epigenetic modifications.
- The presented algorithms give important insight into the governing mechanisms of the histone marks.
- We show both theoretically and in simulations that our proposed algorithm indeed leads to consensus.

Future works

- Further improve the lower bounds given the simulation results.
- Find supporting experimental data to prove that a similar algorithm is really used in nature.

Acknowledgment

Ziv Bar-Joseph

Gadi Taubenfeld

&

BDA organizers.