

# **Solving the Generalized Form of the Game of Set Efficiently**

Steven Takeshita

Adviser: Zachary Kincaid

## **Abstract**

*This document is intended to serve as a sample you can use for independent work reports. We provide some guidelines on content and formatting. They are not required, but they might be helpful.*

## **1. Motivation and Goal**

The Game of Set was created in 1974 and published in 1991. The game consists of  $3^4 = 81$  unique cards. Each card has 4 properties and one of three values. A valid set of three cards is one in which for each property, they all either have the same or different value. At the beginning of the game, 12 cards are shown, and players must locate valid sets. Once a set is found or no set exists, three new cards are added. The most number of sets collected at the ends win. Therefore, a natural extension of this problem is to determine how fast can we find sets. The goal of my project is to create a solver that locates a set efficiently in practice for a game with  $p$  properties and  $v$  values. The Game of Set is an interesting problem for dynamic algorithms in that when three new cards are added, the algorithm should be able to build off of previous knowledge. A way to utilize this past information could be helpful in applying similar principles to other problems that could lend itself to dynamic programming.

## **2. Problem Background and Related Work**

The Game of Set has been researched thoroughly for education and as a motivating example for cap sets, but little research exists about solving a generalized version. Chaudhuri et al (2003) proved that the generalized version of Set is in fact NP-Complete through a reduction from perfect-Dimensional Matching, a known NP-Hard problem [?].

This paper proves computational complexity bounds, but does not solve the problem in practice. Solvers do exist on the internet and can be readily found [?]. Some solvers even employ image processing to identify sets [?]. However, the solvers are constrained to have four properties and three values. Therefore, the solver I will build will be able to solve the generic case of Set in which it has  $p$  properties and  $v$  values.

## 2.1. Problem Definition

A deck will contain  $v^p$  possible cards where the cards have  $p$  properties and  $v$  values. Initially,  $v * p$  cards will be outputted as the starting layout. Identify an arbitrary set of  $v$  cards, in which for all properties the values are either the same or all different. Remove this set of size  $v$  and count it as another set found, or if no set exists, add  $v$  more cards from the remaining  $v^p - v$  cards. Find a total of  $n$  sets, where  $n \leq v^{p-1}$  and each time a set is found  $v$  more cards will be immediately added. Therefore, there are 2 update functions that the algorithms must handle. One to add  $v$  new cards to the board and the other to remove  $v$  cards that formed a set.

By removing the first set of cards seen, this will simulate real gameplay, in which the goal is to find sets as fast as possible. By finding a generalized  $n$  number of sets, as cards are added and removed, the dynamic algorithm will hopefully see some speed up. This further simulates gameplay in which the goal is to find a collection of sets and it is likely that there might not exist sets or sets will be found causing new cards to be added.

## 3. Approach

Generalizing the Game of Set has been done academically, but no practical solution to locate sets exists. A reduction to SMT creates a solver to theoretically always verify whether a set exists, but this does not verify a set within reasonable time if we consider the the poly-time transformation and the fact that it needs to recalculate a set every time new cards are added, not using information from previous rounds to verify it faster. Therefore, my approach will use previous knowledge of the absence of sets to quickly discover sets in a new showing of cards. Set's changing probabilities of discovering a Set also lends itself to the dynamic algorithm approach. Norvig (2017) showed

that on a fresh layout of the game with 12 cards, the ratio of games in which there exists a Set to those that do not contain a set, is 29:1. But as the game plays on, the ratio drops to 15:1, a 50% less chance of finding a set. For 15 cards, the ratio drops to 4% of its original ratio [?]. Therefore, as the game is iteratively played in which Sets are taken out, the game becomes immensely harder. With a dynamic algorithm approach, the solver I build should be able to beat the SMT solver especially with this added difficulty as Sets are removed from the board.

## 4. Reduction to SMT

### 4.1. Set Up

There exists  $v * p$  cards forming a starting board  $B$ , but in general there will be  $n$  cards on a board at any moment. These cards can be represented as vectors where all entries  $b_{i,j} \in \{0, 1, \dots, v-1\}$ :

$$B = \begin{bmatrix} b_{1,1} \\ b_{1,2} \\ \vdots \\ b_{1,p} \end{bmatrix} \begin{bmatrix} b_{2,1} \\ b_{2,2} \\ \vdots \\ b_{2,p} \end{bmatrix} \dots \begin{bmatrix} b_{n,1} \\ b_{n,2} \\ \vdots \\ b_{n,p} \end{bmatrix} \quad (1)$$

The SMT Solver will then attempt to find a set of  $v$  cards. The satisfying set can be denoted as vectors where  $k_{i,j} \in \{0, 1, \dots, v-1\}$ :

$$K = \begin{bmatrix} k_{1,1} \\ k_{1,2} \\ \vdots \\ k_{1,p} \end{bmatrix} \begin{bmatrix} k_{2,1} \\ k_{2,2} \\ \vdots \\ k_{2,p} \end{bmatrix} \dots \begin{bmatrix} k_{v,1} \\ k_{v,2} \\ \vdots \\ k_{v,p} \end{bmatrix} \quad (2)$$

## 4.2. All Different or All Same Constraint

To correctly identify satisfying, distinct sets that exist from the given cards, we need to create three sets of constraints. The first set of constraints will only be satisfied when the cards found  $K$  represent a set, where for all properties, they have either the same or all different value. This constraint can be written as two cases for each property of the cards:

**The values are all the same for a given property i:**

$$(k_{1,i} = k_{2,i}) \wedge (k_{2,i} = k_{3,i}) \wedge \dots \wedge (k_{v-1,i} = k_{v,i}) \quad (3)$$

$$\bigwedge_{m=1}^{v-1} k_{m,i} = k_{m+1,i} \quad (4)$$

**The values are all different for a given property i:**

$$\begin{aligned} &((k_{1,i} \neq k_{2,i}) \wedge (k_{1,i} \neq k_{3,i}) \wedge \dots \wedge (k_{1,i} \neq k_{v,i})) \\ &\wedge ((k_{2,i} \neq k_{3,i}) \wedge (k_{2,i} \neq k_{4,i}) \wedge \dots \wedge (k_{2,i} \neq k_{v,i})) \wedge \dots \wedge (k_{v-1,i} \neq k_{v,i}) \end{aligned} \quad (5)$$

$$\bigwedge_{m=1}^{v-1} \bigwedge_{j=m+1}^v k_{m,i} \neq k_{j,i} \quad (6)$$

Therefore, we can write more concisely that for all properties of the cards, the values must all be the same or all different:

$$\bigwedge_{i=1}^p \left( \left( \bigwedge_{m=1}^{v-1} \bigwedge_{j=m+1}^v k_{m,i} \neq k_{j,i} \right) \vee \left( \bigwedge_{m=1}^{v-1} k_{m,i} = k_{m+1,i} \right) \right) \quad (7)$$

### 4.3. The Cards Must Be From The Board

The second set of constraints is that the cards selected in the set  $K$  must all be from the board  $B$ , which is of size  $n$ . This constraint can be encoded into the SMT solver as:

A given card  $i$  from  $K$  must be from the board  $B$ :

$$\begin{aligned}
 &((k_{i,1} = b_{1,1}) \wedge (k_{i,2} = b_{1,2}) \wedge \dots \wedge (k_{i,p} = b_{1,p})) \vee \\
 &((k_{i,1} = b_{2,1}) \wedge (k_{i,2} = b_{2,2}) \wedge \dots \wedge (k_{i,p} = b_{2,p})) \vee \dots \vee \\
 &((k_{i,1} = b_{n,1}) \wedge (k_{i,2} = b_{n,2}) \wedge \dots \wedge (k_{i,p} = b_{n,p})) \quad (8)
 \end{aligned}$$

Therefore, all cards  $i \in K$  must be from the board  $B$ :

$$\bigwedge_{i=1}^v \left( \bigvee_{j=1}^n \left( \bigwedge_{m=1}^p k_{i,m} = b_{j,m} \right) \right) \quad (9)$$

### 4.4. All Distinct Cards

We constrain the possible cards in the set to be from  $B$ , but this includes duplicates. A possible set could be three of the exact same cards and would satisfy the above constraints but does not represent a real set in the game. Therefore, the last constraint is that the cards selected to be in  $K$ , must all be distinct cards. The cards of a set are considered all distinct if for any two cards, they have at least one property that has a differing value. The constraint can be written as:

A given card,  $i$ , differs with respect to at least one property compared to all cards coming after  $i$ :

$$\begin{aligned}
 &((k_{i,1} \neq k_{i+1,1}) \vee (k_{i,2} \neq k_{i+1,2}) \vee \dots \vee (k_{i,p} \neq k_{i+1,p})) \wedge \\
 &((k_{i,1} \neq k_{i+2,1}) \vee (k_{i,2} \neq k_{i+2,2}) \vee \dots \vee (k_{i,p} \neq k_{i+2,p})) \wedge \dots \wedge \\
 &((k_{i,1} \neq k_{v,1}) \vee (k_{i,2} \neq k_{v,2}) \vee \dots \vee (k_{i,p} \neq k_{v,p})) \quad (10)
 \end{aligned}$$

Written more succinctly:

$$\bigwedge_{i=1}^{v-1} \left( \bigwedge_{j=i+1}^v \left( \bigvee_{m=1}^p k_{i,m} \neq k_{j,m} \right) \right) \quad (11)$$

#### 4.5. Condensing Sets

The satisfying assignment can be achieved through any permutation of cards of a set. Therefore, the SMT solver will be forced to check every permutation of cards for a set, when in reality, all permutations of a given set are equivalent because order does not matter in a set of cards. This will break the symmetry in the encoding and speed up the process of the SMT solver, as the SMT solver will not need to search branches of its search tree that are duplicates of another branch. Therefore, we need to encode this relation of forcing a single set for all permutations of cards. Therefore, we can say that for every possible set, the cards need to be in sorted order by their first property, and if the values are all the same, then by the sorted by the second property, and so forth if the cards in the set have equivalent value for a given property. This can be encoded as follows:

$$\begin{aligned} & (k_{1,1} \leq k_{2,1} \leq \dots \leq k_{v,1}) \wedge ((k_{1,2} \leq k_{2,2} \leq \dots \leq k_{v,2}) \vee \neg(k_{1,1} = k_{2,1} = \dots = k_{v,1})) \\ & \wedge ((k_{1,3} \leq k_{2,3} \leq \dots \leq k_{v,3}) \vee \neg(k_{1,2} = k_{2,2} = \dots = k_{v,2}) \vee \neg(k_{1,1} = k_{2,1} = \dots = k_{v,1})) \\ & \wedge \dots \wedge ((k_{1,p} \leq k_{2,p} \leq \dots \leq k_{v,p}) \vee \neg(k_{1,p-1} = k_{2,p-1} = \dots = k_{v,p-1}) \vee \dots \vee \neg(k_{1,1} = k_{2,1} = \dots = k_{v,1})) \end{aligned} \quad (12)$$

Written more compactly:

$$\bigwedge_{i=1}^p \left( (k_{1,i} \leq k_{2,i} \leq \dots \leq k_{v,i}) \bigvee_{j=1}^{i-1} (\neg(k_{1,j} = k_{2,j} = \dots = k_{v,j})) \right) \quad (13)$$

By building these four constraints (7, 9, 11, 12) from a given board, we can reduce finding an arbitrary set to SMT and use a solver to locate a set efficiently.

#### 4.6. Update Functions

The solver to find sets must also support two update functions. The first update function of removing cards can be easily encoded into the SMT constraints. Let set  $V$  be an arbitrary set that was located by the SMT solver. To remove the set from the board, we can add new constraints such that the new set to be found cannot be equal to any of the cards found in  $V$ , where  $V$  contains  $v$  cards. Again, to be a distinct card, at least one property must be differing. This can be written as:

$$\forall k \in K, \forall v \in V (k_1 \neq v_1 \vee k_2 \neq v_2 \vee \dots \vee k_p \neq v_p) \quad (14)$$

A given card  $i$  from  $K$  must not be equivalent to a card in  $V$ :

$$\begin{aligned} & ((k_{i,1} \neq v_{1,1}) \vee (k_{i,2} \neq v_{1,2}) \vee \dots \vee (k_{i,p} \neq v_{1,p})) \wedge \\ & ((k_{i,1} \neq v_{2,1}) \vee (k_{i,2} \neq v_{2,2}) \vee \dots \vee (k_{i,p} \neq v_{2,p})) \wedge \dots \wedge \\ & ((k_{i,1} \neq v_{v,1}) \vee (k_{i,2} \neq v_{v,2}) \vee \dots \vee (k_{i,p} \neq v_{v,p})) \end{aligned} \quad (15)$$

All cards  $i \in K$  must not be equivalent to a card in  $V$ :

$$\bigwedge_{i=1}^v \left( \bigwedge_{j=1}^v \left( \bigvee_{m=1}^p k_{i,m} \neq v_{j,m} \right) \right) \quad (16)$$

For the second update function in which new cards may be added to the deck, we will need to create a new set of constraints (7, 9, 11) from above. Therefore, this represents a full reduction to SMT from the Game of Set.

## 5. Implementation

The simulations for the randomizer and solvers were all written in python. Considering python's lightweight feel and ease of iterating over all variables in a list, makes it the natural extension for representing a deck and board of cards. Therefore, python made the code easier to write up and allowed for helpful language defined shortcuts.

### 5.1. Randomizer

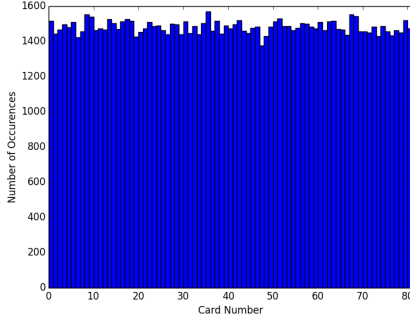
The randomization algorithm for creating a randomized deck is very important in terms of creating a real life scenario gameplay. Therefore, the algorithm must uniformly distribute all of the cards ( $v * p$ ) initially so that no solver can have an unfair advantage knowing what cards are more likely to come out. To do so, I used a couple of important libraries to ensure randomness and to plot the distribution in order to guarantee its uniform distribution. I used the Random library in python to be able to generate random indices to pull for the beginning board ( $v * p$ ) and for the next  $v$  cards when sets are found. I assumed the Random library to be perfectly random. To ensure that the next  $v$  cards are always random from the set, I employed the use of a modified version of Fisher Yates Shuffle, creating a Las Vegas algorithm. Fisher Yates Shuffle works by iteratively randomly picking indices from a finite sequence and produces an unbiased permutation [?]. To convert the algorithm for use on a deck, I use rejection sampling to determine new cards to add to the board. I create a randomized card and if it is not in the board already and not in the removed cards set, I add it. If it is, I generate another card until I have created  $v * p$  cards that are all distinct. Since the algorithm uses a randomized strategy but will always a board, whose cards are all distinct and from the deck with equal probabilities, this is known as a Las Vegas algorithm. To support the update function of removing cards, I append those cards to a set of removed cards. Therefore, when  $v$  new cards are to be added, I generate  $v$  new random cards through the above algorithm to be added to the board. This yields a randomized  $v * p$  starting boards and unbiased drawing of  $v$  cards from the deck.

To confirm that this in fact created random beginning boards and new cards, I graphed the distribution of the beginning board for 1000 trials with  $v = 10$  and  $p = 10$ . I then used numpy to

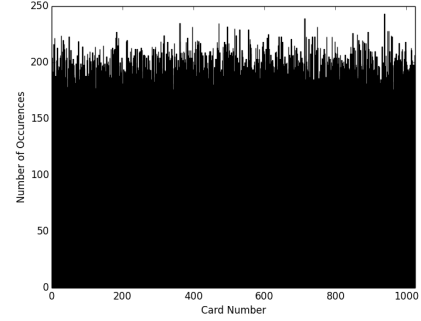


and matplotlib to count all the occurrences of each card and then graph them as a histogram. See the figure below.

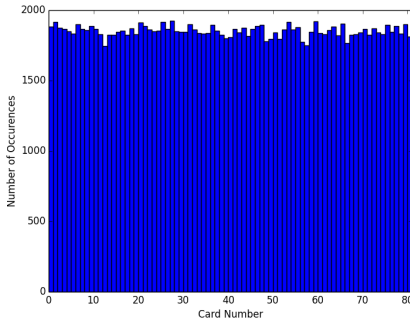
As in Figure 1 and 2, we can see that in both cases, the distribution of cards that were selected form a uniform distribution. Therefore, this confirms that the implementation of the Fisher Yates shuffle in creating a perfectly random board and draws was successful.



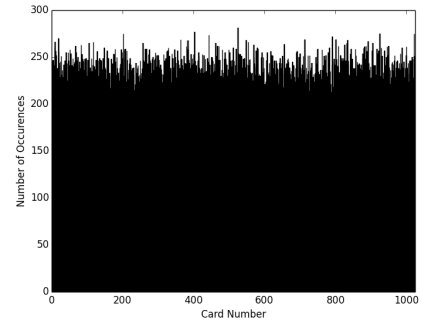
**Figure 1: 3 properties, 4 values, 10,000 trials Initial Board Distribution**



**Figure 2: 4 properties, 5 values, 10,000 trials Initial Board Distribution**



**Figure 3: 3 properties, 4 values, 10,000 trials Board Distribution after One Draw**



**Figure 4: 4 properties, 5 values, 10,000 trials Initial Board Distribution after One draw**

## 5.2. SMT Solver Implementation

For the SMT Solver implementation, I used Z3, a high performance theorem prover created by Microsoft Research. The software creates an easy way to create variables and constraints in Python. Each card that needs to be found represented  $p$  variables. I coded the above constraints (7, 9, 11) in Z3 to be only all satisfied if there exists a set within the board.

To optimize the SMT Solver implementation, I queue up the set to be removed and add the constraints defined in constraint 14 to run through the SMT Solver again to find another set to be removed in case there existed many sets already on the starting board. Once the SMT Solver no longer says that with the constraints and variables it is satisfied, I then know to add  $v$  new cards (or more if the board had a lot of sets and therefore the SMT solver was able to find my sets without having to update the constraints) to hopefully create a new set. I will only do this once no sets exist on the board after removing so many cards to hopefully provide marginal speedup to the program. Also, the cards need to be deleted from the board and therefore, I will call upon the Randomizer method to delete all the cards found by the SMT before no sets existed on the board. Therefore, this will make sure that the dynamic algorithm doesn't have too much of a lead when removing cards. Now, I will add  $v$  new cards and rebuild the constraints (7, 9, 11) and run the SMT solver to hopefully find another set of cards.

This iterative sequence will take place  $n$  times to find  $n$  sets within a game and the SMT solver will return these sets for validation of a legal set.

### **5.3. Dynamic Algorithm Implementation**

if the value number is 1 then all cards are considered a set if the value number is 2 then all pairs of cards are considered a set because for all combos they must differ or be the same by definition of having only two values. Therefore, only two cards comprise a set.

### **5.4. Testing**

For both of these implementation, I will need to implement testing to ensure that the sets found by the solvers are indeed sets and that the algorithm correctly removed the sets from the board.

**5.4.1. Testing Satisfying Set** To confirm that for all properties the cards from the set have either the same value or all different value, I iterated over all properties and all cards values of that property to check whether when put into a set data structure that the length of the set was either 1, meaning that all the values were the same, or the same length as before, meaning all values were distinct elements.

To check that all the boards must be from the board, I had to store all iterations of the board, as new cards could be added and sets would be removed. Therefore, I used a set data structure to store all the cards, and updated this set as new cards were added to the board. Finally, I iterated over each set found and checked that it was included in the board at some point and therefore meant it was taken from the board.

To confirm that the last constraint of all distinct cards was satisfied for each possible set, I asserted that the length of the set was equivalent to the length of the set when converted into a set data structure. Therefore, if they are equal, all elements in the set are therefore distinct and make up all different cards.

Finally, the solver has to handle finding many sets, and to ensure that the solver does not find the same set  $n$  times, I assert that when converted to a long list of cards that make up all  $n$  sets found, that the length of this list converted to a set data structure should be the same as the length of the long list. Therefore, it would show that all  $n * v$  cards are all distinct.

**5.4.2. Testing Correct Deletion** The constraints created for the SMT solver also relies on the fact that the cards from a satisfying set must be deleted from the board at some point. Therefore, as a check to ensure that the corresponding cards were deleted from the board set, I iterated over all cards and asserted that they were not in board at the termination of the program. This is also important for the Dynamic solver because it must output the correct board as well at the end from finding  $n$  sets.

## 6. Results

## 7. Conclusion

### 7.1. Paper Formatting

There are no minimum or maximum length limits on IW reports. We are including this template because we think it will be helpful for citing things properly and for including figures into formatted text. If you are using  $\text{\LaTeX}$  [?] to typeset your paper, then we strongly suggest that you start from the template available at <http://iw.cs.princeton.edu> – this document was prepared with that template.

If you are using a different software package to typeset your paper, then you can still use this document as a reasonable sample of how your report might look. Table 1 is a suggestion of some formatting guidelines, as well as being an example of how to include a table in a Latex document.

Field	Value
Paper size	US Letter 8.5in $\times$ 11in
Top margin	1in
Bottom margin	1in
Left margin	1in
Right margin	1in
Body font	12pt
Abstract font	12pt, italicized
Section heading font	14pt, bold
Subsection heading font	12pt, bold

**Table 1: Formatting guidelines.**

**Please ensure that you include page numbers with your submission.** This makes it easier for readers to refer to different parts of your paper when they provide comments.

We highly recommend you use bibtex for managing your references and citations. You can add bib entries to a references.bib file throughout the semester (e.g., as you read papers) and then they will be ready for you to cite when you start writing the report. If you use bibtex, please note that the references.bib file provided in the template example includes some format-specific incantations at the top of the file. If you substitute your own bib file, you will probably want to include these incantations at the top of it.

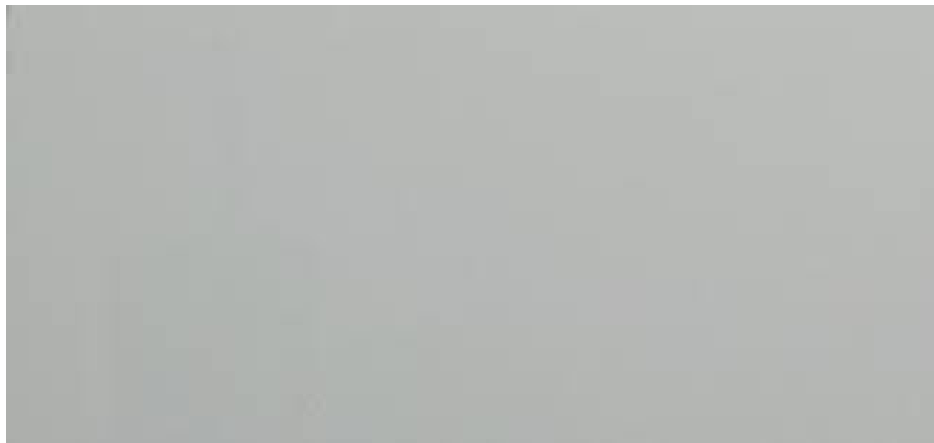
## 7.2. Citations and Footnotes

There are various reasons to cite prior work and include it as references in your bibliography. For example, If you are improving upon prior work, you should include a full citation for the work in the bibliography [?, ?]. You can also cite information that is used as background or explanation[?]. In addition to citing scholarly papers or books, you can also create bibtex entries for webpages or other sources. Many online databases allow you to download a premade bibtex entry for each paper you access. You can simply copy-paste these into your references.bib file.

Sometimes you want to footnote something, such as a web site.<sup>1</sup> Note that the footnote number comes after the punctuation.

### 7.3. Figures and Tables.

Figure 5 shows an example of how to include a figure in your report. Ensure that the figures and tables are legible. Please also ensure that you refer to your figures in the main text. Make sure that your figures will be legible in the expected forms that the report will be read. If you expect someone to print it out in gray-scale, then make sure the figures are legible when printed that way.



**Figure 5: This is a gray image.**

In Section 7.1, an example of a table was given. (Note that the “S” in Section is capitalized. Here’s one more example - see Table 2.

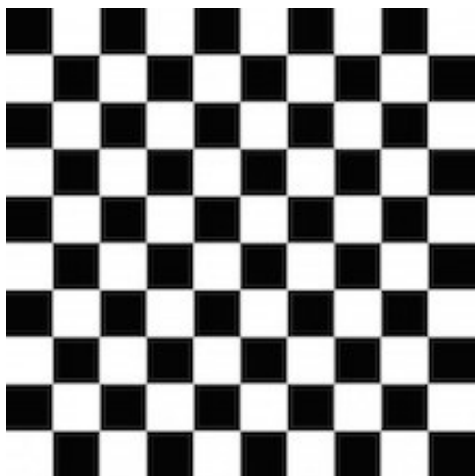
Some field	Another field
200	10000
400	20000
800	40000
1600	80000
3200	160000
6400	320000

**Table 2: Some data in a table.**

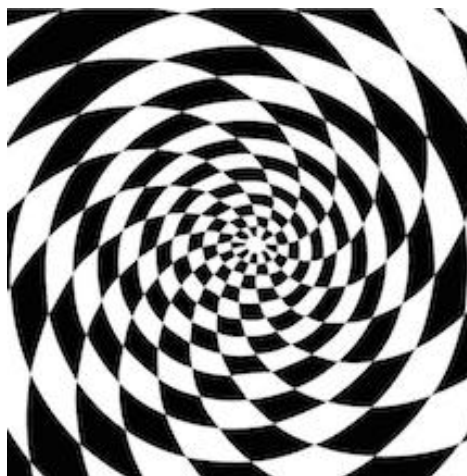
Here’s an example that shows how you can have side-by-side figures - see Figure 6 and Figure 7. (Note that the the “F” in Figure is capitalized.

---

<sup>1</sup><http://www.cs.princeton.edu>



**Figure 6: Plain checkerboard.**



**Figure 7: Cool checkerboard.**

#### **7.4. Double Quotes.**

Latex double quotes are not the same as the double quote key on your keyboard. The standard way of writing quotes and double quotes in LaTeX is with “ and ” not with " and ".

Now that may be confusing, so you may want to use the `\{quotes}` command. For example “The quick brown fox.”

#### **7.5. Main Body.**

Avoid bad page or column breaks in your main text, i.e., last line of a paragraph at the top of a column or first line of a paragraph at the end of a column. If you begin a new section or sub-section near the end of a column, ensure that you have at least two (2) lines of body text on the same column.

### **8. Outline**

The following is a possible outline for your paper.

#### **8.1. Introduction**

- Motivation and Goal (The goal of this project is...)
- Overview of challenge and previous work
- Approach

- Summary of implementation
- Summary of results
- (optional) Roadmap: The remainder of this paper is organized as follows....

## **8.2. Problem Background and Related Work**

- Survey of prior work with similar goals
- For each previous approach, explain what has been done and why it does not meet your goal

## **8.3. Approach**

- Key novel idea
- Why it is a good idea

## **8.4. Implementation**

- System overview (flow chart of key steps?)
- Subsection for each step or issue you addressed
  - Problem statement
  - Possible approaches
  - Chosen approach and why
  - Implementaton details

## **8.5. Evaluation**

- Experiment design...
- Data...
- Metrics...
- Comparisons...
- Qualitative results...
- Quantitative results...

## **8.6. Summary**

- Conclusions...
- Limitations...
- Future work...

## **9. Ethics**

Your independent work report should abide by the basic standards of scholarly ethics and by the Princeton Honor Code. If you have any doubts about how to cite other work, how to quote or include text or images from other works, or other issues, please discuss them with your project adviser or with the IW coordinators.