Adapting Biochemical Kripke Structures for Distributed Model Checking

Susmit Jha
IIT Kharagpur

R K Shyamasundar
Tata Institute of Fundamental Research

Outline

- Distributed Model Checking
- Biochemical Kripke Structures
- Bounded Hamming Distance Kripke Structures
- Polynomial Fragments for BHDKS
- Hypercube based fragments for BHDKS
- Conclusions and Future Work

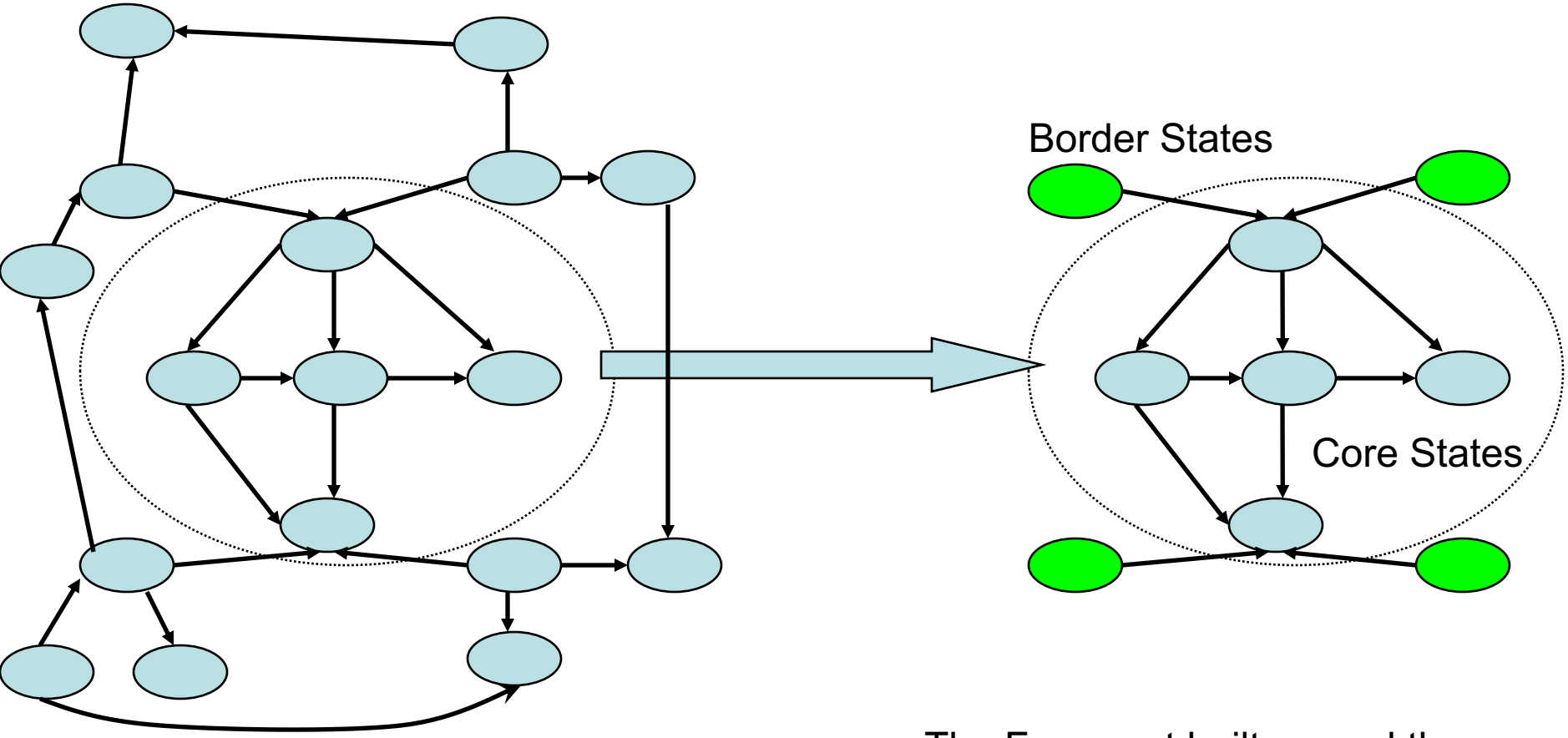
Model Checking

- The size of the state space of a system can grow exponentially in the number of atomic propositions. – State Explosion
- Several methods to tame the state explosion problem – symbolic (BDD s), partial order reduction. Distributed Model checking is a complementary technique.
- Real world model checkers do run on network of processors. DMC adds to the arsenal against state explosion.
- Assumption based DMC is a relatively recent development [Brim et al]

Assumption based Distributed Model Checking (Brim et al)

- Split the Kripke structure into smaller fragments.
- A fragment F is any set of states S of the Kripke structure (and the induced transition relation $R|_S$ by this set) and includes copies of all those states which have a transition into any state in the set S.
- Fragments are model checked individually by making assumptions on the “border” states and then communication occurs to solve inter-dependencies.

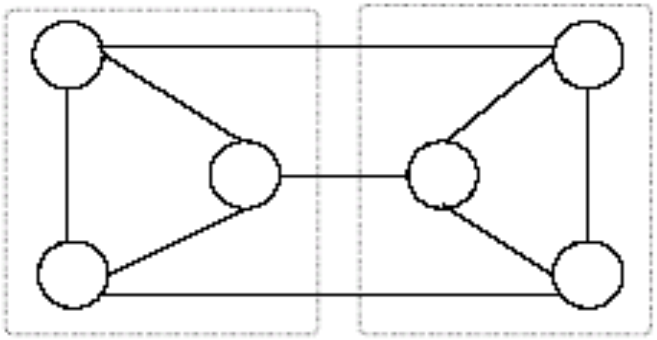
How Fragmentation Works?



Portion of a Kripke Structure:
The circled outline shows the set around which we want to form a fragment

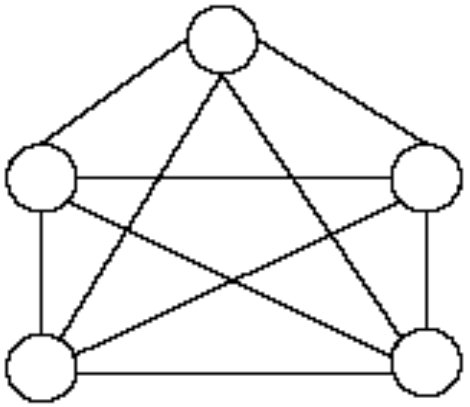
The Fragment built around the States within the circle in the Kripke Structure

Bad Cases of Assumption Based DMC



A Bad Instance with poor choice of sets

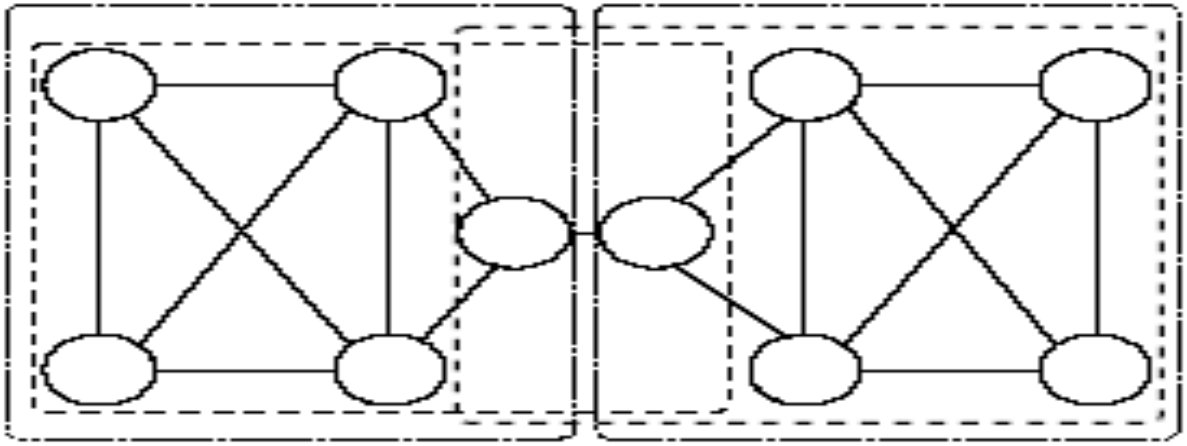
The subsets are shown by dotted boxes. For these subsets, each of the fragment will be as large as the original Kripke structure and the purpose of the distributed algorithm will fail.



5 - clique

Irrespective of the choice of our subsets, each fragment will be as large as the whole Kripke structure once again.

A good example



The dotted boxes surround the subsets used for constructing the partition. The dashed lines show the actual partitions themselves. Also, the undirected edges indicate transitions possible in both directions.

Observe that the partition was able to reduce the size of the Kripke structure rather well.

Biochemical Kripke Structures

- a biochemical reaction takes the system from a state with biochemical entities matching the left-hand side of the reaction rule, into one of the other states in which the biochemical entities of the right-hand side have been added.
- The biochemical entities which appear only in the left-hand side of the rule and not in the right-hand side may be nondeterministically present or absent in the target state.
- each biochemical entity is associated with a proposition.
- If the biochemical entity is present in a state, the associated boolean proposition is *true* else it is *false*.
- a transition occurs from one state to another by “executing” a biochemical reaction and the truth values of the boolean propositions change to reflect the biochemical entities added or removed

An Example Model

- Abstract Modeling.
  - Consider the scenario of A and B reacting to form C and D,
    - $A + B \rightarrow C + D$
  - We want to non-deterministically capture all possible scenarios:
    - A + B + ¬C + ¬D ➔ ¬A + B + C + D
    - A + B + ¬C + ¬D ➔ A + ¬B + C + D
    - A + B + ¬C + ¬D ➔ ¬A + ¬B + C + D
    - ……
    - ……
    - A + B + C + D ➔ ¬A + B + C + D
    - A + B + C + D ➔ A + ¬B + C + D
    - A + B + C + D ➔ ¬A + ¬B + C + D
    - A + B + C + D ➔ A + B + C + D

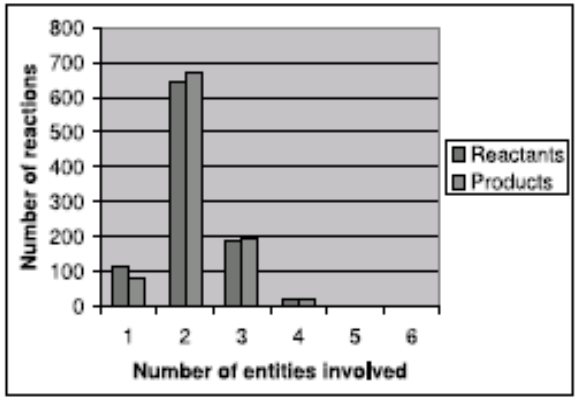
Some notes on the Kripke Structure Model

- By using this boolean abstraction, such models are capable of reasoning about all possible behaviors of the system with unknown concentration values and unknown kinetics parameters [Fages et al].
- This modeling is particularly useful for complex chemical systems like biochemical pathways where even a boolean abstraction can generate valuable results.
- It is also now well appreciated that biological models, despite their hybrid nature, indeed have many digital (boolean) controls.
- A Kripke structure is an asynchronous formalism.
  - In particular, two reactions occurring “simultaneously” can be modeled as one occurring after another because of the
    - non-deterministic modeling with respect to the reactants
    - and the asynchronous interleaving semantics of Kripke structures.

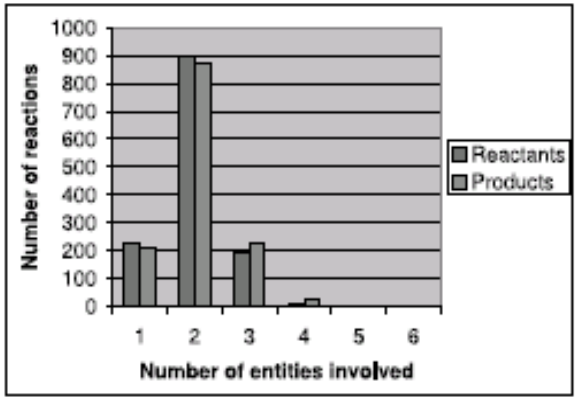
Bound on the number of chemical entities involved in a reaction

- The number of biochemical entities reacting in a chemical reaction is fairly small.
  - almost 60% of the reactions in the databases we analyzed have no more than two reactants or two products.
  - no reaction was found with more than six reactants or products in these databases of widely differing organisms.
  - explained by the fact that there is a very low probability of the interaction of more than a few entities at the atomic level.
- Contrast this with an arithmetic operation a := a×b, a system wide reset in a VLSI chip or the setting of bits in a long flag register.
  - the Kripke structure of these hardware or software systems from one state to another such that the Hamming distance between them is arbitrarily large.

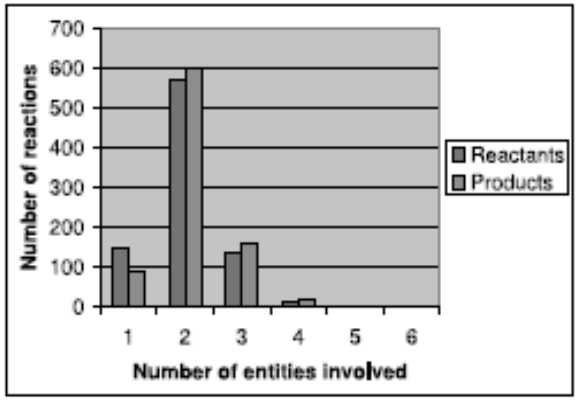
The HumanCyc, EcoCyc, AnthraCyc and YeastCyc Databases Reactions Summary



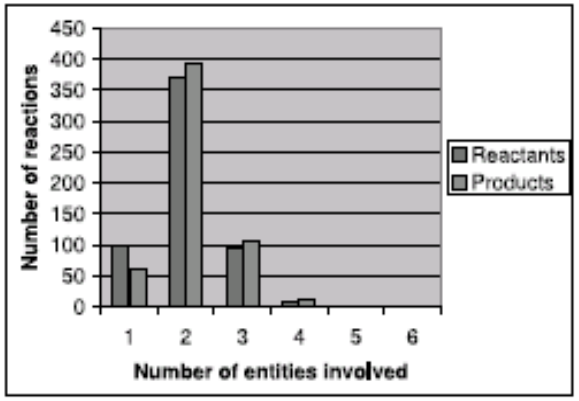
(a) HumanCyc



(b) EcoCyc



(c) AnthraCyc



(d) YeastCyc

The bar charts clearly show that most reactions have small number of reactants and products. There is no reaction having more than 6 reactants or products among some 3000 biochemical reactions in these databases.

Bounded Hamming Distance Kripke Structures

- A k - Bounded Hamming Distance Kripke structure has a transition between two states in the Kripke structure only if the Hamming distance between the propositional labels of these states is at most k.

Definition 1. *Let K = (S,R,$\mathcal{AP}$,$\mathcal{L}$,F) be a Kripke structure, where S is the set of states, R is the transition relation, $\mathcal{AP}$ is the set of atomic propositions, $\mathcal{L}$ is the labeling of states with atomic propositions, F is the set of final states, and H(x,y) denotes the Hamming distance between x and y. Then, K is called a k - Bounded Hamming Distance Kripke structure iff*

$$\forall s, s' \in S, \;\; R(s, s') \implies (H(\mathcal{L}(s), \mathcal{L}(s')) \leq k)$$

Biochemical Kripke Structures are BHDKS

Theorem 1. *A biochemical Kripke structure **K** is a* k *– Bounded Hamming Distance Kripke structure (BHDKS) for some small* k*.*

Proof Sketch:
If there is a transition from s to s′, then the system executes some reaction at state s.

Now, the reaction has at most r reactants and at most p products, where r and p are small.

When the reaction is executed, the reactants can non-deterministically be removed from the system, while the products are added to the system.

Thus, s′ can differ from s in at most k = r + p chemical entities.

Edge Density of BHDKS

Theorem 2. *A state in the k - Bounded Hamming Distance Kripke structure with* log n *number of propositions (where* n > 1*) has a degree of at most* $(\log n)^k$.

Proof.
Consider all possible neighbors N(s) of some state s in the Kripke structure.

From the definition of BHDKS, we know that s′ 2 N(s) iff H(s,s') ≤ k.

Hence, N(s) can have no more states than those which are atmost k away from s.

Thus, an upper bound = $\sum_{i=0}^{k} \left({}^{\log(n)}C_i \right)$

$$\cdot \ (\log (n)\ )^k$$

The number of transitions in a Bounded Hamming Distance Kripke structure are no more than polynomially (in the number of propositions in the Kripke structure) larger than the number of states.

Polynomial Fragments for BHDKS

Theorem 3. *Given any set T ½ S of the state space of a k - Bounded Hamming Distance Kripke structure K = (S,R) with log (n) propositions, the size of the smallest separator V of T w.r.t. S is no more than |T| . (log (n)) ^{k}.*

Proof: Straightforward from bound on edge density

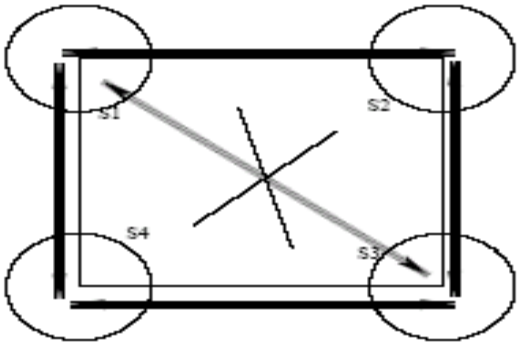
Corollary: *Given any set T ½ S of the state space of a k - Bounded Hamming Distance Kripke structure K = (S,R) with log (n) propositions, the size of the fragment associated with T is no more than |T| . (1 + (log (n)) ^{k}).*

Proof: Any set of states with its separator w.r.t. the rest of the Kripke structure contains a fragment.

This shows that the size of the state space which needs to be put at one node of the distributed computation grows only polynomially in the number of propositions in the Bounded Hamming Distance Kripke structure.

Example Hypercube Splitting

The sets S1, S2, S3 and S4 are formed as before by dividing the state space into 4 parts around 4 equidistant centers 0^{2p}, 0^p1^p, 1^{2p} and 1^p0^p.



If we take these sets as the corners of a 2-D hypercube (square), then one can show that there can be no transitions between the distributed nodes along the diagonals.

So the size of each fragment is at most 3 times the size of the core set at each node

Bound on the size of the fragment

Theorem 4. *For a BHDKS Kripke structure split uniformly around four centers* 0^{2p}, 0^p1^p, 1^{2p} *and* 1^p0^p*, there can be no transition along the diagonal as long as* $p > k$.

Proof: Suppose there is a transition from the set around 0^{2p} to the set around 1^{2p} say from x to y. Then, $H(x,y) \cdot k$.

Also, by construction, $H(x,0^{2p}) \cdot p/2$ and $H(y,1^{2p}) \cdot p/2$.

Now by triangle inequality, $H(y,0^{2p}) + H(y,1^{2p}) \, ¸ \, H(0^{2p},1^{2p})$

i.e. $H(y,0^{2p}) \, ¸ \, 2p - p/2$.

Also, by triangle inequality, $H(x,y) + H(x,0^{2p}) \, ¸ \, H(y,0^{2p})$

i.e. $H(x,y) \, ¸ \, H(y,0^{2p}) - H(x,0^{2p})$

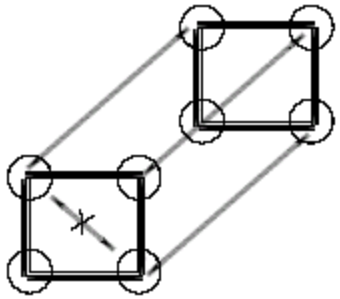
i.e. $H(x,y) \, ¸ \, 2p - p/2 - p/2$

i.e. $H(x,y) \, ¸ \, p$

Thus, as long as $p > k$, there can be no transitions along the diagonal.

HyperCube Splitting of State Space

- Consider a state space split into 2^l parts in a l - Dimensional hypercube – the centers of each of the 2^l split state spaces P_i are uniformly distributed.



Theorem 5. *For a* k*-BHDKS Kripke structure with (log (n)) propositions split uniformly around* 2l *centers* 0^{lp}, $0^{(l-1)p}\,1^{p}$, ... $0^{p}1^{(l-1)p}$, 1^{lp} *(where p = (log (n) / l)) and* p > k*, there can be no transition along any of the diagonals of this* l *- dimensional hypercube.*

No transitions exist among non-adjacent distributed nodes

Suppose that there is a transition from the set around θ to the set around δ say from x to y . Then, $H(x,y) \cdot k$.

Also, δ and θ are along some diagonal and not adjacent. So, $H(\theta,\delta) \neg 2p$.

Also, by construction, $H(x,\theta) \cdot p/2$ and $H(y,\delta) \cdot p/2$.

By triangle inequality $H(y,\theta) + H(y,\delta) \neg H(\theta,\delta)$

i.e. $H(y,\theta) \neg 2p - p/2$

(assuming the worst case that δ and θ are as close as possible without being neighbors in the l -dimensional hypercube) .

Also, by triangle inequality, $H(x,y) + H(x,\theta) \neg H(y,\theta)$

i.e. $H(x,y) \neg H(y,\theta) - H(x,\theta)$

i.e. $H(x,y) \neg p$

Size of the fragment in Hypercube fragmentation

Corollary 2. *The size of the separator of the set associated with each distributed node in the* l*-Dimensional hypercube is at most* l *times the size of the largest possible core set at each node i.e.* $(l / 2^{l}) . n$

Proof.
Each node in the l-Dimensional hypercube has transitions only to the neighbouring nodes in the hypercube.

In an l-dimensional hypercube, there are l neighbours.

By construction, each neighbour has no more than $(n / 2^{l})$ *core* states.

Corollary *The size of the fragment associated with each node in the l-Dimensional hypercube is at most* (l + 1) *times the size of the largest possible core set at each node i.e.*$((l + 1) / 2^{l}) . n.$

Experimental Results

Sl No	Database	Radius of the Frag-ment	Number of States in the core	Maximum number of states in the fragment	Ratio of fragment to the core
1	HumanCyc	8	60321482688944611644	58218118459069712450424	965
2	HumanCyc	9	7459853563127158123804	719888188817241356451515 6	965
3	HumanCyc	10	829547867699812679324780	800431570432559915098596984	964
4	HumanCyc	11	83785702021492624364150540	80835123199556021465682097364	964
5	HumanCyc	12	77503169484011783042367978 60	74764684646408464350771370760 96	964
6	EcoCyc	8	215766787047246662253	286658426283477266973032	1328
7	Ecocyc	9	31310453270114925645193	41591878653908316107275044	1328
8	Ecocyc	10	40860575706621402650205 69	54270306998590744772102849 60	1328
9	Ecocyc	11	48438928429446296001103101 7	64326583496672658366811053520 8	1327
10	Ecocyc	12	52597289383826851902453164625	69838841881773224220828914800104	1327

The ratio indicates that the fragment is only a small multiple of the size of the core

Experimental Results

Sl No	Benchmark	Hamming Diameter	Size of core	Size of border	Fraction of core to fragment size
1	Circadian oscillations	2	10	59	0.1449
2	Circadian oscillations	3	51	127	0.2865
3	Circadian Oscillations	4	140	149	0.48445
4	Circadian Oscillations	5	251	102	0.7110
5	Circadian Oscillations	6	333	41	0.8904
6	Cell Division Cycle	2	7	25	0.2187
7	Cell Division Cycle	3	29	48	0.3766
8	Cell Division Cycle	4	71	63	0.5299
9	Cell Division Cycle	5	126	59	0.6811
10	Cell Division Cycle	6	179	41	0.8136

Table 2. Fragmentation results for the CMBSLib Benchmark: http://contraintes.inria.fr/CMBSlib/

The ratio indicates that the fragment is only a small multiple of the size of the core. In this case, the ratios are even more favorable.

Conclusions and Future Work

- We have seen that biochemical Kripke structures can be divided into fragments as small as polynomial in the number of atomic propositions present in the Kripke structure.
- The hypercube algorithm tends to distribute the entire exponential state space in a uniform manner, and one may raise the question as to the benefit of this exercise when the reachable state space is small.
- Our explicit distributed construction of the state space partitioning assumes that there is a number close to log n which has factors that can be used as l - the dimension of the embedding hypercube.
- The choice of the hypercube in which the system is embedded and the assignment of different embeddings onto the same hypercube (by changing the order of propositions in the state space) needs to be studied.
- Bounded Hamming Distance Kripke structures are also very suitable for Bounded Model Checking.
  - It is an interesting challenge to exploit the locality in transitions to derive SAT heuristics for BHDKS.
- Can the insight be used to guide design for verifiability of control paths in circuits ?

Acknowledgement

- ITPAR (Indo-Italian Promotion of Advanced Research) for visit to University of Trento
- Indian Academy of Sciences Summer Fellowship