A Decomposition Theorem for Probabilistic Transition Systems

Oded Maler[*]
VERIMAG
Miniparc ZIRST
38330 Montbonnot
France
Oded.Maler@imag.fr

July 12, 1995

[*] A preliminary version of this paper appeared in: STACS 93 *Proc. 10th Annual Symposium on Theoretical Computer Science* (P. Enjalbert, A. Finkel and K.W. Wagner Eds.), Lecture Notes in Computer Science, Vol. 665, pp. 586–594, Springer-Verlag, Berlin 1993. The results presented in this paper have been obtained while the author was with INRIA/IRISA, Rennes, France.

Abstract

In this paper we prove that every finite Markov chain can be decomposed into a cascade product of a Bernoulli process and several simple permutation-reset *deterministic* automata. The original chain is a state-homomorphic image of the product. By doing so we give a positive answer to an open question stated in [Paz71] concerning the decomposability of probabilistic systems. Our result is based on the observation that in probabilistic transition systems, "randomness" and "memory" can be separated so as to allow the non-random part to be treated using common deterministic automata-theoretic techniques. The same separation technique can be applied to other kinds of non-determinism as well.

1 Preliminaries

The object of our study is a probabilistic input-output state-transition system. Its definition is not new and has appeared under various names in the past (e.g., [Arb68, Paz71, Sta72]).

Definition 1 (Probabilistic Transition Systems) *A probabilistic transition system (PTS) is a quadruple $\mathcal{A} = (X, Q, Y, p)$ where X is the input alphabet, Q is the state-space, Y is the output alphabet and $p : Q \times X \times Q \times Y \rightarrow [0, 1]$ is the input-transition-output probability function satisfying for every $q \in Q, x \in X$:*

$$\sum_{(q',y) \in Q \times Y} p(q, x, q', y) = 1$$

The intuitive meaning of this definition is that whenever $\mathcal{A}$ is in a state q and reads the input x it will move to state q' and emit y with probability $p(q, x, q', y)$. Throughout this paper we will consider only finite Q, X, and Y. Several well-known models can be considered as degenerate variants of PTSs where either X or Q are singletons, $|Y| \leq |Q|$ or some additional constraints are imposed upon p. We will mention a few of these:

- A *Markov chain*: X is a singleton, $Y = Q$ and $p(q, x, q', y) > 0$ only if $q' = y$. The intuitive meaning is that the behavior of the chain depends only on the passage of time, and the observable output coincides with the internal state. In this case we will refer to the transition probability (also known as transition matrix) as $p(q, q')$.

- A *probabilistic automaton*: a Markov chain with a non-singleton input alphabet. In the Markovian terminology this is a controlled process where the input letter determines which of the several transition matrices will be applied at each step.

- A *deterministic input-output automaton*: for every $q \in Q$, $x \in X$ there exists exactly one $q' \in Q$, $y \in Y$ such that $p(q, x, q', y) = 1$. In this case we can express p using a transition function $\delta : Q \times X \rightarrow Q$ and an output function $\gamma : Q \times X \rightarrow Y$. When the output is suppressed, i.e., $\gamma(q, x) = q$, we have a probabilistic automaton with a $0 - 1$ transition matrix.

- An *acceptor*: $Y = \{0,1\}$. In the deterministic case $\mathcal{A}$ is said to accept all input sequences that produce output sequences ending with 1. In the probabilistic case it accepts all the input sequences such that the expected value of their corresponding last output is above some threshold. If we suppress the input we get what is also known as a partially-observable Markov chain.

- A *Bernoulli process*: both X and Q are singletons. In this case the system has no memory and no input and it produces its output according to a fixed probability distribution.

2 Homomorphisms between PTSs

One of the most important notions concerning transition systems is the notion of homomorphism. A system $\mathcal{A}_2$ is homomorphic to $\mathcal{A}_1$ if, in some sense, $\mathcal{A}_2$ approximates $\mathcal{A}_1$. This notion is very well developed and studied in the context of deterministic systems but its application to probabilistic systems is a bit more subtle. We will consider here only *state* homomorphism, that is, homomorphism between two PTSs having *the same* input and output alphabets. These definitions can be extended to mappings between the input and output alphabets of the two systems.

Definition 2 (PTS Homomorphism) *Given two PTSs $\mathcal{A}_1 = (X, Q_1, Y, p_1)$ and $\mathcal{A}_2 = (X, Q_2, Y, p_2)$, a (state) homomorphism from $\mathcal{A}_1$ to $\mathcal{A}_2$ is a surjective function $\varphi : Q_1 \rightarrow Q_2$ such that for every $(q_2, x, q_2', y) \in Q_2 \times X \times Q_2 \times Y$ and every $q_1 \in \varphi^{-1}(q_2)$ we have*

$$p_2(q_2, x, q_2', y) = \sum_{q_1' \in \varphi^{-1}(q_2')} p_1(q_1, x, q_1', y)$$

We denote this fact by $\mathcal{A}_2 \leq_\varphi \mathcal{A}_1$. Two systems are isomorphic if φ is a bijection.

Intuitively this definition means that $\mathcal{A}_2$ can be constructed by partitioning Q_2 into blocks in such a way that the transition probabilities between the blocks are consistent with the transition probabilities between their elements (this is also termed the *lumpability condition* in the Markovian terminology). It can be seen that in the case of $0-1$ probabilities

this notion coincides with the familiar notion of automaton homomorphism, namely $\varphi(\delta(q,x)) = \delta'(\varphi(q), x)$.
An essential property of homomorphisms is their transitivity, that is, if $\mathcal{A}_2$ approximates $\mathcal{A}_1$ and $\mathcal{A}_3$ approximates $\mathcal{A}_2$ then $\mathcal{A}_3$ approximates $\mathcal{A}_1$.

Claim 1 (Transitivity of Homomorphism) *If $\mathcal{A}_2 \leq_\varphi \mathcal{A}_1$ and $\mathcal{A}_3 \leq_\psi \mathcal{A}_2$ then $\mathcal{A}_3 \leq_\theta \mathcal{A}_1$ where $\theta = \psi\varphi$.*

Proof: We will give the proof for Markov chains for reasons of clarity – the generalization to input-output PTSs is straightforward. Let $\mathcal{A}_1 = (Q, p_1)$, $\mathcal{A}_2 = (R, p_2)$ and $\mathcal{A}_3 = (S, p_3)$ be three chains satisfying the premise of the claim. We want to show that for every $s, s' \in S$ and every $q \in \theta^{-1}(s)$ we have

$$p_3(s,s') = \sum_{q' \in \theta^{-1}(s')} p_1(q,q')$$

But $\theta(q) = s$ if for some $r \in R$, $\varphi(q) = r$ and $\psi(r) = s$. Thus for every $s \in S$

$$\theta^{-1}(s) = \bigcup_{r \in \psi^{-1}(s)} \varphi^{-1}(r)$$

Thus we have to prove that for every $r \in \psi^{-1}(s)$ and $q \in \varphi^{-1}(r)$

$$p_3(s,s') = \sum_{r' \in \psi^{-1}(s')} \left(\sum_{q' \in \varphi^{-1}(r')} p_1(q,q') \right)$$

But since $\mathcal{A}_2 \leq_\varphi \mathcal{A}_1$ we can replace, for every $q \in \varphi^{-1}(r)$, the expression in the parentheses by $p_2(r,r')$ and obtain

$$p_3(s,s') = \sum_{r' \in \psi^{-1}(s')} p_2(r,r')$$

which, in turn, follows from $\mathcal{A}_3 \leq_\psi \mathcal{A}_2$. ▪

3 Composition of PTSs

Two PTSs can be connected together such that the output of the first is the input of the second, or formally:

Definition 3 (Cascade Product) *Given two PTSs $\mathcal{A}_1 = (X, Q_1, Z, p_1)$ and $\mathcal{A}_2 = (Z, Q_2, Y, p_2)$, their cascade product is $\mathcal{A}_1 \circ \mathcal{A}_2 = (X, Q, Y, p)$ where $Q = Q_1 \times Q_2$ and for every $(q_1, q_2), (q_1', q_2') \in Q$, $x \in X$ and $y \in Y$:*

$$p((q_1, q_2), x, (q_1', q_2'), y) = \sum_{z \in Z} p_1(q_1, x, q_1', z) \cdot p_2(q_2, z, q_2', y)$$

This definition can be extended to a family $\mathcal{A}_1, \ldots, \mathcal{A}_k$ of PTSs such that the input alphabet of $\mathcal{A}_{i+1}$ is the output alphabet of $\mathcal{A}_i$. The product defined this way is associative so the notation $\mathcal{A}_1 \circ \mathcal{A}_2 \circ \ldots \circ \mathcal{A}_k$ is well-defined. One can see that this definition reduces to the common notion of cascade product when both systems are deterministic, $Z = X \times Q_1$ and $Y = Q_2$. In that case we have the following well-known result ([KR65]), stating that every finite automaton can be constructed from simple building blocks:

Theorem 2 (Krohn-Rhodes Decomposition) *Every deterministic automaton $\mathcal{A}$ is inverse-homomorphic to a cascade product of simple permutation automata and reset automata.*

This theorem is beyond the scope of this paper, so we will only mention that:

1. The permutation groups of the components divide the subgroups of the transformation semigroup of $\mathcal{A}$ (which implies that counter-free automata can be decomposed into a cascade of reset automata).

2. The number of automata in the cascade is bounded by $|Q|$.

3. The number of states in the decomposition can be exponential in $|Q|$.

Additional details can be found in [Eil72, Gin68, MP90]. With respect to this theorem, the following question has been asked in [Paz71, p. 115]: *Can every Markov system be "embedded" in a nontrivial way into a cascade type interconnection of systems which have a specific simple form? In other words, is there any theorem which can be proved for Markov systems and which parallels in some way the Krohn-Rhodes theorem for the deterministic case?* In this paper we give an affirmative answer.

4 Our Result

First we will show how to decompose a finite-state PTS into an isomorphic cascade product of a Bernoulli process and a deterministic automaton. For simplicity we will consider the degenerate case of a Markov chain, and show that every such chain can be simulated by a product of two systems, the first one taking care of the randomness and the other behaving deterministically according to the outcome of the former. In other words, instead of throwing a different coin at every state, we throw each time *the same* (but a much larger) coin, whose outcome tells us which transition to take from each of the states we might be in. The probabilities of all the possible trajectories of the original chain and those of its associated decomposition are the same.

Definition 4 (Probability of Transformations) *For a set $Q = \{q_1, \ldots, q_n\}$, we let $M = Q^Q$ denote the set of all n^n transformation on Q. Equipped with the composition operation, M is a semigroup. With every Markov chain*[1] *$\mathcal{A} = (\{x^*\}, Q, Q, p)$ we associate a function $\pi : M \to [0,1]$ by letting*

$$\pi(m) = \prod_{i=1}^{n} p(q_i, m(q_i))$$

Claim 3 $\sum_{m \in M} \pi(m) = 1$.

Proof: Follows from

$$\begin{aligned} \sum_{m \in M} \pi(m) &= \textstyle\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} p(q_1, q_{i_1}) \cdot p(q_2, q_{i_2}) \ldots p(q_n, q_{i_n}) && (1) \\ &= \textstyle\prod_{i=1}^{n} \sum_{j=1}^{n} p(q_i, q_j) = \prod_{i=1}^{n} 1 && (2) \end{aligned}$$

Claim 4 (New Decomposition I) *Every Markov chain $\mathcal{A} = (\{x^*\}, Q, Q, p)$ with $|Q| = n$ is isomorphic to a cascade product of a Bernoulli generator with at most n^n outcomes and a deterministic n-state automaton.*

Proof: We define a Bernoulli process $\mathcal{B} = (\{x^*\}, \{q^*\}, M, \pi)$ and a deterministic automaton $\mathcal{A}' = (M, Q, p')$ where for all $m \in M, q \in Q$, $p'(q, m, m(q)) =$

[1] We omit the singleton input and the output (which is identical to the state) from the definition of p.

1. Their product $\mathcal{C} = \mathcal{B} \circ \mathcal{A}'$ is a Markov chain $\mathcal{C} = (\{x^*\}, \{q^*\} \times Q, Q, \bar{p})$ where $\bar{p}$ is defined as

$$\bar{p}((q^*, q), (q^*, q')) = \sum_{m \in M} \pi(m) \cdot p'(q, m, q') = p(q, q')$$

and the straightforward state bijection $\varphi((q^*, q)) = q$ is indeed a PTS isomorphism between $\mathcal{A}$ and $\mathcal{C}$. ▟

Note that $\mathcal{B}$ can be further decomposed into a direct product of n independent Bernoulli trials, each having at most n outcomes. This result extends easily to input-output PTSs: instead of an input-less Bernoulli process we will have a one-state PTS with input; we can get rid from the output by splitting states, as in the standard proof of the equivalence of Moore and Mealy machines (see [HU79]).

In order to take advantage of this decomposition result and combine it with the Krohn-Rhodes decomposition we need (a weak version of) the following:

Claim 5 *Let* $\mathcal{B} = (X, Q, Z, p)$, $\mathcal{A}_1 = (Z, R, Y, p_1)$ *and* $\mathcal{A}_2 = (Z, S, Y, p_2)$ *be PTSs. If* $\mathcal{A}_2 \leq \mathcal{A}_1$ *then* $\mathcal{B} \circ \mathcal{A}_2 \leq \mathcal{B} \circ \mathcal{A}_1$.

Proof: Without loss of generality we let $Y = Q \times S$ and thus $\mathcal{B} \circ \mathcal{A}_1 = (X, Q \times R, Q \times S, \bar{p}_1)$ and $\mathcal{B} \circ \mathcal{A}_2 = (X, Q \times S, Q \times S, \bar{p}_2)$. Based on the assumed homomorphism $\varphi : R \rightarrow S$ we construct a surjective mapping $\bar{\varphi} : Q \times R \rightarrow Q \times S$ by letting $\bar{\varphi}(q, r) = (q, \varphi(r))$. According to our definition, $\bar{\varphi}$ is a homomorphism if for every $x \in X$, $y \in Y$, $(q, s), (q', s') \in Q \times S$ and for every $(q, r) \in \bar{\varphi}^{-1}(q, s)$:

$$\bar{p_2}((q, s), x, (q', s'), y) = \sum_{(q', r') \in \bar{\varphi}^{-1}(q', s')} \bar{p_1}((q, r), x, (q', r'), y)$$

Using the definition of the product we get:

$$\sum_{z \in Z} p(q, x, q', z) \cdot p_2(s, z, s', y) = \sum_{r' \in \varphi^{-1}(s')} \sum_{z \in Z} p(q, x, q', z) \cdot p_1(r, z, r', y)$$

Since $p(q, x, q', z)$ does not depend on r we can rearrange the right hand side and get

$$\sum_{z \in Z} p(q, x, q', z) \cdot p_2(s, z, s', y) = \sum_{z \in Z} p(q, x, q', z) \cdot \sum_{r' \in \varphi^{-1}(s')} p_1(r, z, r', y)$$

which follows from the fact that φ is a homomorphism. ▟

Corollary 6 (New Decomposition II) *Every finite Markov chain is inverse homomorphic to a cascade product of a Bernoulli process and a chain of deterministic permutation-reset automata.*

Proof: Follows from the above and the Krohn-Rhodes decomposition theorem (theorem 2). ∎

An example appears in the appendix.

Remark: Our claim 4 can be improved using the following result [Paz71, pp. 11-12]: *Every probabilistic $n \times n$ matrix can be written as $\sum_{i=1}^{m} p_i A_i$ where for every i, $0 \leq p_i \leq 1$, A_i is a $0-1$ matrix, $\sum_{i=1}^{m} p_i = 1$ and $m \leq n^2$.* Hence, a Bernoulli generator and a deterministic automaton over an n^2 alphabet suffice.

5 Discussion

We have shown how the automata-theoretic framework, emphasizing the notion of communication between processes, can be used in order to decompose arbitrary probabilistic transition matrices into products of several "communicating" simple zero-one matrices.

In addition to the solution we give to an open problem, the connection we establish between every finite Markov chain and its "characteristic" deterministic automaton might be used in order to transfer various results between automata theory and the theory of stochastic processes. For example, the algebraic theory of deterministic automata and their associated semigroups is well-developed (see [Eil76], [Lal79], [Pin86]) and it will be interesting to investigate the relation between the detailed classification results concerning automata, and various properties of stochastic processes discussed in the Markovian literature ([KS60]).

Finally, it is worth mentioning that this technique works for other types of finite non-determinism as well. For example, it is possible to decompose any non-deterministic automaton with input into an inverse-homomorphic (in the appropriate sense of homomorphism) cascade consisting of a non-deterministic one-state input-output automaton and a deterministic automaton. In this way the results in [MP90] conerning the translation from counter-free automata to formulas of past temporal logic can be extended to non-deterministic automata without explicit determinization.

Acknowledgement

Various anonymous referees contributed to the style and rigor of this paper. One of them pointed out the existence of Paz's variant of our claim 4.

References

[Arb68] M.A. Arbib, *Theories of Abstract Automata*, Prentice-Hall, Englewood Cliffs, 1968.

[Eil76] S. Eilenberg, *Automata, Languages and Machines, Vol. B*, Academic Press, New York, 1976.

[Gin68] A. Ginzburg, *Algebraic Theory of Automata*, Academic Press, New York, 1968.

[HU79] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, Reading, MA, 1979.

[KS60] J.G. Kemeny and J.L. Snell, *Finite Markov Chains*, Van Nostrand, New York, 1960.

[KR65] K. Krohn and J.L. Rhodes, Algebraic Theory of Machines, I Principles of Finite Semigroups and Machines, *Transactions of the American Mathematical Society* 116, 450-464, 1965.

[Lal79] G. Lallement, *Semigroups and Combinatorial Applications*, Wiley, New York, 1979.

[MP90] O. Maler and A. Pnueli, Tight Bounds on the Complexity of Cascaded Decomposition of Automata, *Proc. 31st FOCS*, 672-682, 1990.

[Paz70] A. Paz, *Introduction to Probabilistic Automata*, Academic Press, New York, 1970.

[Pin86] J.-E. Pin, *Varieties of Formal Languages*, Plenum, New York, 1986.

[Sta72] P.H. Starke, *Abstract Automata*, North-Holland, Amsterdam, 1972.

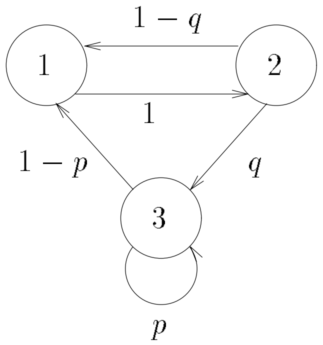


Figure 1: *A Markov chain $\mathcal{A}$.*

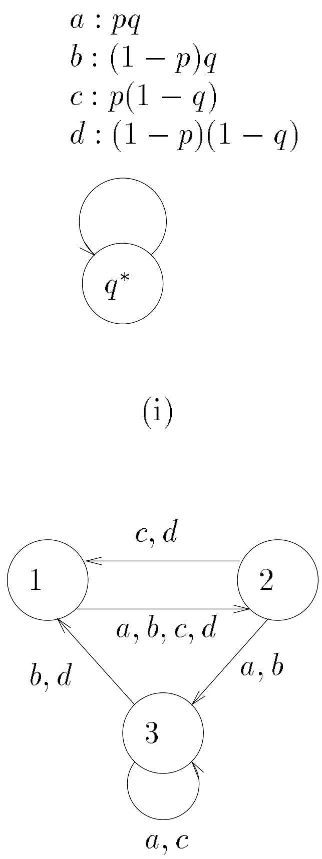
Appendix: An Example

Consider the Markov chain $\mathcal{A} = (\{x^*\}, Q, Q, p)$ with $Q = \{1, 2, 3\}$ depicted in figure 1. It is first decomposed into an isomorphic product $\mathcal{B} \circ \mathcal{A}'$ of a Bernoulli process $\mathcal{B} = (\{x^*\}, \{q^*\}, Z, \pi)$ with $Z = \{a, b, c, d\}$ and a deterministic automaton $\mathcal{A}' = (Z, Q, Q, \delta)$ where $\delta : Z \times Q \to Q$ is a deterministic transition function (see figure 2). Note that we have considered only those transformations $m \in Q^Q$ for which $\pi(m) > 0$.

By applying the Krohn-Rhodes decomposition theorem, we decompose $\mathcal{A}$ into an inverse homomorphic product $\mathcal{A}_1 \circ \mathcal{A}_2$ where $\mathcal{A}_1 = (Z, Q_1, W, \delta_1, \gamma_1)$ and $\mathcal{A}_2 = (W, Q_2, \delta_2)$ with $Q_1 = \{4, 5, 6\}$, $Q_2 = \{7, 8\}$ and $W = \{e, f, g, h\}$ – see figure 3. Note that all input symbols in both automata induce either a reset or a permutation.

Their product yields the automaton $\mathcal{C}' = (Z, Q_1 \times Q_2, \bar{\delta})$ of figure 4, which when multiplied from the left by $\mathcal{B}$ yields the chain $\mathcal{C} = (Q_1 \times Q_2, \bar{p})$ of figure 5. One can verify that the mapping $\varphi : Q_1 \times Q_2 \to Q$ defined in figure 6 which is a deterministic state-homomorphism from $\mathcal{C}'$ to $\mathcal{A}'$ is also a PTS homomorphism from $\mathcal{C}$ to $\mathcal{A}$.

Note also that the projection $\psi : Q_1 \times Q_2 \to Q_1$ is a state-homomorphism from $\mathcal{A}_1 \circ \mathcal{A}_2$ to $\mathcal{A}_1$. It is also a PTS homomorphism from $\mathcal{B} \circ \mathcal{A}_1 \circ \mathcal{A}_2$ to $\mathcal{B} \circ \mathcal{A}_1$ (see figures 7 and 8).



(i)

(ii)

Figure 2: *(i) The Bernoulli process $\mathcal{B}$ and (ii) the deterministic automaton $\mathcal{A}'$ such that $\mathcal{B} \circ \mathcal{A}'$ is isomorphic to the original Markov chain $\mathcal{A}$.*

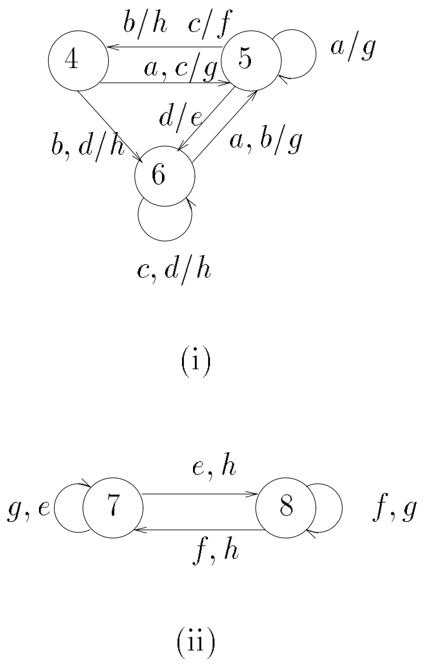


Figure 3: *The decomposition of the automaton $\mathcal{A}'$ into a cascade of deterministic permutation-reset automata (i) $\mathcal{A}_1$ and (ii) $\mathcal{A}_2$. The transition labels of the form x/y in $\mathcal{A}_1$ indicate that x is the input and y is the output.*

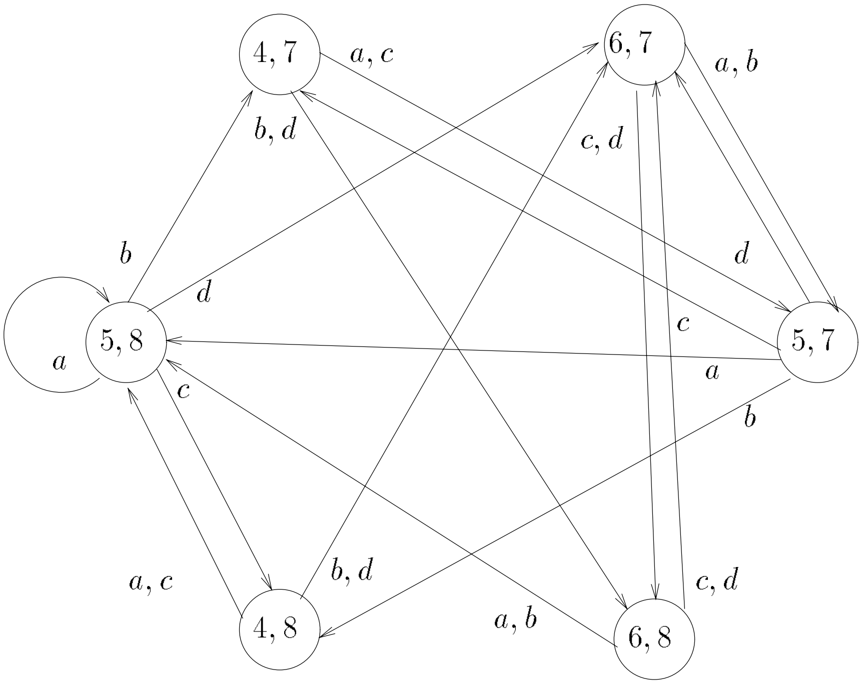


Figure 4: *The automaton* $\mathcal{C}' = \mathcal{A}_1 \circ \mathcal{A}_2$.

	$(4,7)$	$(6,7)$	$(5,7)$	$(6,8)$	$(4,8)$	$(5,8)$
$(4,7)$	0	0	p	$1-p$	0	0
$(6,7)$	0	0	q	$1-q$	0	0
$(5,7)$	$p(1-q)$	$(1-p)(1-q)$	0	0	$(1-p)q$	pq
$(6,8)$	0	$1-q$	0	0	0	q
$(4,8)$	0	$1-p$	0	0	0	p
$(5,8)$	$(1-p)q$	$(1-p)(1-q)$	0	0	$p(1-q)$	pq

Figure 5: *The Markov chain* $\mathcal{C} = \mathcal{B} \circ \mathcal{A}_1 \circ \mathcal{A}_2$ *written in a matrix form. The rows and columns are arranged according to the homomorphism from* $\mathcal{C}$ *to the original chain* $\mathcal{A}$.

$(4,7)$	1
$(4,8)$	3
$(5,7)$	2
$(5,8)$	3
$(6,7)$	1
$(6,8)$	2

Figure 6: *The homomorphism* $\varphi : Q_1 \times Q_2 \to Q$.

p	4	5	6
4	0	p	$1-p$
5	$q+p-2pq$	pq	$(1-p)(1-q)$
6	0	q	$1-q$

Figure 7: *The chain* $\mathcal{B} \circ \mathcal{A}_1$.

	$(4,7)$	$(4,8)$	$(5,7)$	$(5,8)$	$(6,7)$	$(6,8)$
$(4,7)$	0	0	p	0	0	$1-p$
$(4,8)$	0	0	0	p	$1-p$	0
$(5,7)$	$p(1-q)$	$(1-p)q$	0	pq	$(1-p)(1-q)$	0
$(5,8)$	$(1-p)q$	$p(1-q)$	0	pq	$(1-p)(1-q)$	0
$(6,7)$	0	0	q	0	0	$1-q$
$(6,8)$	0	0	0	q	$1-q$	0

Figure 8: *The Markov chain* $\mathcal{C} = \mathcal{B} \circ \mathcal{A}_1 \circ \mathcal{A}_2$ *written in a matrix form. The rows and columns are arranged according to the projection homomorphism from* $\mathcal{C}$ *to* $\mathcal{B} \circ \mathcal{A}_1$.