Quantum Finite State Transducers

Rūsiņš Freivalds*
Institute of Mathematics and Computer Science,
University of Latvia, Raiņa bulvāris 29, LV–1459, Riga, Latvia.

Andreas Winter†
SFB 343, Fakultät für Mathematik, Universität Bielefeld,
Postfach 100131, D–33501 Bielefeld, Germany.

October 31, 2000

Abstract

We introduce *quantum finite state transducers (qfst)*, and study the class of relations which they compute. It turns out that they share many features with probabilistic finite state transducers, especially regarding undecidability of emptiness (at least for low probability of success). However, like their 'little brothers', the quantum finite automata, the power of qfst is incomparable to that of their probabilistic counterpart. This we show by discussing a number of characteristic examples.

1 Introduction

The issue of this work is to introduce and to study the computational model of quantum finite state transducers. These can be understood as finite automata which compute a relation between strings, instead of a decision (which we read as a binary valued function). After the necessary definitions (section 2), the relation to quantum finite automata is clarified (section 3), then decidability questions are addressed (section 4): it is shown that emptiness of the computed relation is undecidable both for qfst and pfst. However, we present some evidence that it may be decidable for high probability of success. Next, the relation between deterministic and probabilistic transducers is explored (section 5), and in section 6 qfst and pfst are compared.

We feel our extension of quantum automata studies to this new model justified by the following quote from D. Scott [9]:

> *'The author (along with many other people) has come recently to the conclusion that the functions computed by the various machines are more important – or at least more basic – than the sets accepted by these devices. (...) In fact by putting the functions first, the relationship between various classes of sets becomes much clearer'.*

*Email: `rusins@paul.cclu.lv`. Research supported by contract IST–1999–11234 (QAIP) from the European Commision, and grant no. 96.0282 from the Latvian Council of Science. This work was carried out during the first author's stay at Bielefeld University in October 2000.

†Email: `winter@mathematik.uni-bielefeld.de`. Research supported by SFB 343 "Diskrete Strukturen in der Mathematik" of the Deutsche Forschungsgemeinschaft.

2 Definitions

We start be reviewing the concept of probabilistic finite state transducer. For a finite set X we denote by X^* the set of all finite strings formed from X, the empty string is denoted ε.

Definition 1 *A* probabilistic finite state transducer (pfst) *is a tuple* $T = (Q, \Sigma_1, \Sigma_2, V, f, q_0, Q_{\mathrm{acc}}, Q_{\mathrm{rej}})$, *where Q is a finite set of states, Σ_1, Σ_2 is the input/output alphabet, $q_0 \in Q$ is the initial state, and $Q_{\mathrm{acc}}, Q_{\mathrm{rej}} \subset Q$ are (disjoint) sets of accepting and rejecting states, respectively. (The other states, forming set Q_{non}, are called non–halting). The transition function $V : \Sigma_1 \times Q \to Q$ is such that for all $a \in \Sigma_1$ the matrix $(V_a)_{qp}$ is stochastic, and $f_a : Q \to \Sigma_2^*$ is the output function. If all matrix entries are either* 0 *or* 1 *the machine is called a* deterministic finite state transducer (dfst).

The meaning of this definition is that, being in state q, and reading input symbol a, the transducer prints $f_a(q)$ on the output tape, and changes to state p with probability $(V_a)_{qp}$, moving input and output head to the right. After each such step, if the machine is found in a halting state, the computation stops, accepting or rejecting the input, respectively.

To capture this formally, we introduce the *total state* of the machine, which is an element

$$(P_{\mathrm{NON}}, P_{\mathrm{ACC}}, p_{\mathrm{rej}}) \in \ell^1(Q \times \Sigma_2^*) \oplus \ell^1(\Sigma_2^*) \oplus \ell^1(\{\mathrm{REJ}\}),$$

with the natural norm

$$\|(P_{\mathrm{NON}}, P_{\mathrm{ACC}}, p_{\mathrm{rej}})\| = \|P_{\mathrm{NON}}\|_1 + \|P_{\mathrm{ACC}}\|_1 + |p_{\mathrm{rej}}|.$$

At the beginning, the total state is $((q_0, \epsilon), \mathbf{0}, 0)$ (where we identify an element of $Q \times \Sigma_2^*$ with its characteristic function). The computation is represented by the (linear extensions of the) transformations

$$T_a : ((q, w), P_{\mathrm{ACC}}, p_{\mathrm{rej}}) \mapsto \left(\left(\sum_{p \in Q_{\mathrm{non}}} (V_a)_{qp} p, w f_a(q) \right), P'_{\mathrm{ACC}}, p'_{\mathrm{rej}} \right),$$

of the total state, for $a \in \Sigma_1$, with

$$P'_{\mathrm{ACC}}(x) = \begin{cases} P_{\mathrm{ACC}}(x) + \sum_{p \in Q_{\mathrm{acc}}} (V_a)_{qp} & \text{if } x = w f_a(q), \\ P_{\mathrm{ACC}}(x) & \text{else,} \end{cases}$$

and $p'_{\mathrm{rej}} = p_{\mathrm{rej}} + \sum_{p \in Q_{\mathrm{rej}}} (V_a)_{qp}$.

For a string $x_1 \dots x_n$ the map T_x is just the concatenation of the T_{x_i}. Observe that all the T_a conserve the probability.

Implicitely, we add initial and end marker symbols (‡, \$) at the input, with additional stochastic matrices $V_{\ddagger}$ and $V_{\$}$, executed only at the very beginning, and at the very end. We assume that $V_{\$}$ puts no probability outside $Q_{\mathrm{acc}} \cup Q_{\mathrm{rej}}$.

By virtue of the computation, to each input string $v \in \Sigma_1^*$ there corresponds a probability distribution $T(\cdot|v)$ on the set $\Sigma_2^* \cup \{\mathrm{REJ}\}$:

$$T(\mathrm{REJ}|v) := T_{\ddagger v \$}((q_0, \epsilon), \mathbf{0}, 0)[\mathrm{REJ}]$$

is the probability to reject the input v, whereas

$$T(w|v) := T_{\ddagger v \$}((q_0, \epsilon), \mathbf{0}, 0)[w]$$

is the probability to accept, after having produced the output w.

Definition 2 *Let $\mathcal{R} \subset \Sigma_1^* \times \Sigma_2^*$.*

For $\alpha > 1/2$ we say that T computes the relation $\mathcal{R}$ with probability α *if for all v, whenever $(v,w) \in \mathcal{R}$, then $T(w|v) \geq \alpha$, and whenever $(v,w) \notin \mathcal{R}$, then $T(w|v) \leq 1-\alpha$*

For $0 < \alpha < 1$ we say that T computes the relation $\mathcal{R}$ with isolated cutpoint α *if there exists $\epsilon > 0$ such that for all v, whenever $(v,w) \in \mathcal{R}$, then $T(w|v) \geq \alpha + \epsilon$, but whenever $(v,w) \notin \mathcal{R}$, then $T(w|v) \leq \alpha - \epsilon$.*

The following definition is modelled after the previous one and the one for quantum finite state automata [7]:

Definition 3 *A* quantum finite state transducer (qfst) *is a tuple $T = (Q, \Sigma_1, \Sigma_2, V, f, q_0, Q_{\mathrm{acc}}, Q_{\mathrm{rej}})$, where Q is a finite set of states, Σ_1, Σ_2 is the input/output alphabet, $q_0 \in Q$ is the initial state, and $Q_{\mathrm{acc}}, Q_{\mathrm{rej}} \subset Q$ are (disjoint) sets of accepting and rejecting states, respectively. The transition function $V : \Sigma_1 \times Q \to Q$ is such that for all $a \in \Sigma_1$ the matrix $(V_a)_{qp}$ is unitary, and $f_a : Q \to \Sigma_2^*$ is the output function.*

Like before, implicitely matrices $V_{\ddagger}$ and $V_{\$}$ are assumed, $V_{\$}$ carrying no amplitude from Q_{non} to outside $Q_{\mathrm{acc}} \cup Q_{\mathrm{rej}}$. The computation proceeds as follows: being in state q, and reading a, the machine prints $f_a(q)$ on the output tape, and moves to the superposition $\sum_p (V_a)_{qp} p$ of internal states. Then a measurement of the orthogonal decomposition $E_{\mathrm{non}} \oplus E_{\mathrm{acc}} \oplus E_{\mathrm{rej}}$ (with the subspaces $E_i = \mathrm{span}\ Q_i \subset l^2(Q)$, which we identify with their respective projections) is performed, stopping the computation with accepting the input on the second outcome (while observing the output), with rejecting it on the third.

Here, too, we define total states: these are elements

$$(|\psi_{\mathrm{NON}}\rangle, P_{\mathrm{ACC}}, p_{\mathrm{rej}}) \in \ell^2(Q \times \Sigma_2^*) \oplus \ell^1(\Sigma_2^*) \oplus \ell^1(\{\mathrm{REJ}\}),$$

with norm

$$\|(|\psi_{\mathrm{NON}}\rangle, P_{\mathrm{ACC}}, p_{\mathrm{rej}})\| = \||\psi_{\mathrm{NON}}\rangle\|_2^2 + \|P_{\mathrm{ACC}}\|_1 + |p_{\mathrm{rej}}|.$$

At the beginning the total state is $(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0)$, the total state transformations, for

$$|\psi\rangle = \sum_{q \in Q} |q\rangle \otimes |\omega_q\rangle, \qquad \text{with } |\omega_q\rangle = \sum_{w \in \Sigma_2^*} \alpha_{qw} |w\rangle,$$

are (for $a \in \Sigma_1$)

$$T_a : (|\psi\rangle, P_{\mathrm{ACC}}, p_{\mathrm{rej}}) \mapsto \left(E_{\mathrm{non}} \sum_q V_a |q\rangle \otimes |\omega_q f_a(q)\rangle, P'_{\mathrm{ACC}}, p'_{\mathrm{rej}} \right),$$

where $|\omega_q f_a(q)\rangle = \sum_w \alpha_{qw} |w f_a(q)\rangle$, and

$$P'_{\mathrm{ACC}}(x) = P_{\mathrm{ACC}}(x) + \left\| E_{\mathrm{acc}} \sum_{q,w \text{ s.t. } x = w f_a(q)} \alpha_{qw} V_a |q\rangle \right\|_2^2,$$

$$p'_{\mathrm{rej}} = p'_{\mathrm{rej}} + \left\| E_{\mathrm{rej}} \sum_q V_a |q\rangle \otimes |\omega_q f_a(q)\rangle \right\|_2^2.$$

Observe that, again, the T_a preserve the norm. Hence the distributions $T(\cdot|v)$ are defined, and so are the concepts of computation with probability α or with isolated cutpoint α.

Notice the physical benefits of having the output tape: whereas for finite automata a superposition of states means that the amplitudes of the various transitions are to be added, this is no longer true for transducers if we face a superposition of states *with different output tape content.* I.e. the entanglement of the internal state with the output may prohibit certain interferences. This will be a crucial feature in some of our later constructions.

3 Quantum Finite Automata and Quantum Transducers

The definition of qfst is taylored in such away that by throwing away the output tape and the output function, we get a quantum finite automaton. One, however, with distinct acceptance and rejection properties, as compared to the qfst.

Nevertheless, the decision capabilities of qfst equal those of quantum finite automata:

Theorem 4 *A language L is accepted by a 1–way quantum finite automaton with probability bounded away from 1/2 if and only if the relation $L \times \{0\} \cup \overline{L} \times \{1\}$ is computed with isolated cutpoint.*

Proof: First observe that for finite automata (probabilistic and quantum), recognizability with an isolated cutpoint is equivalent to recognizability with probability bounded away from 1/2 (by "shifting the cutpoint": just add in the ‡–step possibilities to accept or reject right away with certain probabilities). We have to exhibit two constructions:

Let there be given a quantum finite automaton. We may assume that it is such that $V_\$$ is a permutation on Q. [This can be forced by duplicating each $q \in Q_{\rm acc} \cup Q_{\rm rej}$ – q' – and modifying the transition function as follows: denote by σ the map interchanging q with q' for $q \notin Q_{\rm non}$, and being the identity on $Q_{\rm non}$. Define a unitary U such that for $q \in Q_{\rm non}$

$$U|q\rangle = \sum_p (V_\$)_{qp} |\sigma p\rangle,$$

and $U|q\rangle = |q\rangle$ for $q \in Q_{\rm acc} \cup Q_{\rm rej}$. Now let

$$V'_\ddagger := UV_\ddagger, \qquad V'_\$:= \sigma, \qquad V'_a := UV_aU^{-1}.$$

It is easily checked that this automaton behaves exactly like the initial one]. Construct a qfst as follows: its states are $Q \cup \widehat{Q}$, with $\widehat{Q} = \{\hat{q} : q \in Q_{\rm acc} \cup Q_{\rm rej}\}$ being the accepting states, and no rejecting states. Let the transition function be W with

$$\begin{aligned} W_a|q\rangle &= V_a|q\rangle \text{ for } q \in Q_{\rm non}, \text{ but} \\ W_a|q\rangle &= |\hat{q}\rangle \text{ for } q \in Q_{\rm acc} \cup Q_{\rm rej}. \end{aligned}$$

Since $V_\$$ is the permutation σ on Q, we may define

$$W_\$ = \begin{cases} |\widehat{\sigma q}\rangle & \text{for } \sigma q \in Q_{\rm acc} \cup Q_{\rm rej}, \\ |\sigma q\rangle & \text{for } \sigma_q \in Q_{\rm non}. \end{cases}$$

Finally, let the output function be (for $q \in Q$)

$$f_a(q) = \begin{cases} \epsilon & \text{for } q \in Q_{\rm non}, \\ 0 & \text{for } q \in Q_{\rm acc}, \\ 1 & \text{for } q \in Q_{\rm rej}, \end{cases} \qquad f_\$(q) = \begin{cases} 0 & \text{for } \sigma q \in Q_{\rm acc}, \\ 1 & \text{for } \sigma q \in Q_{\rm rej}, \end{cases}$$

and ϵ otherwise. It can be checked that it behaves in the desired way.

Given a qfst, construct a quantum finite automaton as follows: its states are $Q \times \Sigma_2^{\leq t}$, where the second component represents the tape content up to $t = 1 + \max_{a,q} |f_a(q)|$ many symbols. Initial state is (q_0, ϵ). Observe that by definition of the T_a amplitude that once is shifted onto output tapes of length larger than 1 is never recovered for smaller lengths. Hence we may as well cut such branches by immediate rejection: the states in $Q \times \Sigma_2^{\geq 2}$ are all rejecting, and so are $(Q_{\mathrm{acc}} \cup Q_{\mathrm{rej}}) \times \{1\}$. The accepting states are $Q_{\mathrm{acc}} \times \{0\}$.

The transition function is given by

$$W_a|q, w\rangle := \sum_{p \in Q} (V_a)_{qp} |p, w f_a(q)\rangle$$

(for $a = \$$ this is followed by mapping $|p, \epsilon\rangle$ to a rejecting state, while leaving the other halting states alone), i.e. the automaton performs like the qfst on the elements of Q, and uses the second component to simulate the output tape. One can check that this construction behaves in the desired way. □

4 Decidability questions

As is well known, the emptiness problem for the language accepted by a deterministic (or nondeterministic) finite automaton is decidable. Since the languages accepted by probabilistic and quantum finite automata with bounded error are regular [8, 7], these problems are decidable, too.

For finite state transducers the situation is more complicated: In [5] it is shown that the emptiness problem for deterministic and nondeterministic fst is decidable. In contrast we have

Theorem 5 *The emptiness problem for pfst computing a relation with probability* $2/3$ *is undecidable.*

Likewise, the emptiness problem for qfst computing a relation with probability $2/3$ *is undecidable.*

Proof: By reduction to the Post Correspondence Problem: let an instance $(v_1, \dots, v_k)$, $(w_1, \dots, w_k)$ of PCP be given (i.e. $v_i, w_i \in \Sigma^+$). It is to be decided whether there exists a sequence $i_1, \dots, i_n$ such that

$$v_{i_1} \cdots v_{i_n} = w_{i_1} \cdots w_{i_n}.$$

Construct the following qfst: it has states q_0, q_v, q_w, and q_{rej}. The initial transformation produces a superposition of $q_v, q_w, q_{\mathrm{rej}}$, each with amplitude $1/\sqrt{3}$. The unitaries U_i are all identity, but the output function is defined as $f_i(q_x) = x_i$, for $x \in \{v, w\}$. The endmarker maps q_v, q_w to accepting states. It is clear that $i_1, \dots, i_n$ is a solution iff $(i_1 \dots i_n, v_{i_1} \cdots v_{i_n})$ is in the relation computed with probability $2/3$.

By replacing the unitaries by stochastic matrices (with entries the squared moduli of the corresponding amplitudes) the same applies to pfst.

Since it is well known that PCP is undecidable, it follows that there can be no decision procedure for emptiness of the relation computed by the constructed pfst, or qfst, respectively. □

Remark 6 *Undecidable questions for quantum finite automata were noted first for "$1\frac{1}{2}$–way" automata, i.e. ones which move only to the right on their input, but may also keep their position on the tape. In [1] it is shown that the equivalence problem for these is undecidable. The same was proved for 1–way–2–tape quantum finite automata in [3].*

Conjecture 7 *The emptiness problem for probabilistic and quantum fst computing a relation with probability* 0.99 *is decidable.*

The emptiness problem for probabilistic and quantum fst computing a relation with a single–letter input alphabet, with probability $1/2 + \epsilon$ *is decidable.*

To prove this, we would like to apply a packing argument in the space of all total states, equipped with the above metric. However, this fails because of the infinite volume of this space (for finite automata it is finite, see [8] and [7]). In any case, a proof must involve the size of the gap between the upper and the lower probability point, as the above theorem shows that it cannot possibly work with gap 1/3.

Still, we can prove:

Theorem 8 *If the relation* $\mathcal{R}$ *is computed by a pfst or a qfst with an isolated cutpoint, then* $\mathrm{Range}(\mathcal{R}) = \{y : \exists x\ (x, y) \in \mathcal{R}\}$ *is a recursive set (so, for each specific output, it is decidable if it is ever produced above the threshold probability).*

Proof: Let the cutpoint be α, with isolation radius δ, and let $y = y_1 \dots y_n \in \Sigma_2^*$.

Define $Y = \{y_1 \dots y_i : 0 \le i \le n\}$, the set of prefixes of y. Consider the *output–y–truncated total state*, which is an element

$$(|\widetilde{\psi}\rangle, \widetilde{P}_{\mathrm{ACC}}, \widetilde{p}_{\mathrm{rej}}) \in \ell^2(Q \times Y) \oplus \ell^1(Y) \oplus \ell^1(\{\mathrm{REJ}\}) < \ell^2(Q \times \Sigma_2^*) \oplus \ell^1(\Sigma_2^*) \oplus \ell^1(\{\mathrm{REJ}\}).$$

It is obtained from $(|\psi\rangle, P_{\mathrm{ACC}}, p_{\mathrm{rej}})$ – with $|\psi\rangle = \sum_{q,w} \alpha_{qw} |q\rangle \otimes |w\rangle$ – by defining

$$\begin{aligned} |\widetilde{\psi}\rangle &= \sum_{q \in Q, w \in Y} \alpha_{qw} |q\rangle \otimes |w\rangle, \\ \widetilde{P}_{\mathrm{ACC}} &= P_{\mathrm{ACC}}|_Y, \\ \widetilde{p}_{\mathrm{rej}} &= p_{\mathrm{rej}} + \sum_{q \in Q, w \notin Y} |\alpha_{qw}|^2 + \sum_{w \notin Y} P_{\mathrm{ACC}}(w). \end{aligned}$$

Let us denote this transformation by J. Now observe that in the total state evolution of the qfst probability once put outside Y never returns, and likewise, amplitude once put outside $Q \times Y$ never returns (compare proof of theorem 4). Formally, this is reflected in the relation

$$JT_{ab}(|\widetilde{\psi}\rangle, \widetilde{P}_{\mathrm{ACC}}, \widetilde{p}_{\mathrm{rej}}) = JT_b JT_a(|\widetilde{\psi}\rangle, \widetilde{P}_{\mathrm{ACC}}, \widetilde{p}_{\mathrm{rej}}).$$

Hence, if we want to know if $T(y|x) \ge \alpha + \delta$ for some x, we may concentrate on the space of output–y–truncated total states, which is finite dimensional, and its transformation functions $\widetilde{T}_a = JT_a$.

It is easily seen that there is a constant γ such that for all truncated total states s, t and all $w \in \Sigma_1^*$

$$\|\widetilde{T}_w s - \widetilde{T}_w t\| \le \gamma \|s - t\|.$$

Hence, for $x, x', w \in \Sigma_1^*$, if

$$\|\widetilde{T}_{\ddagger x}(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0) - \widetilde{T}_{\ddagger x'}(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0)\| < \delta/\gamma,$$

then

$$\|\widetilde{T}_{\ddagger xw\$}(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0) - \widetilde{T}_{\ddagger x'w\$}(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0)\| < \delta.$$

Because of the cutpoint isolation we find that either both or none of $(x, y), (x', y)$ is in $\mathcal{R}$. Now, because of caompactness of the set of truncated total states reachable from the starting state, it follows that there is a constant c such that for all $x \in \Sigma_1^*$ of length $|x| \geq c$ one can write $x = vx_0w$, with $|x_0| < c$, such that

$$\|\widetilde{T}_{\ddagger vx_0}(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0) - \widetilde{T}_{\ddagger v}(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0)\| < \delta/\gamma.$$

Hence

$$\|\widetilde{T}_{\ddagger vw\$}(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0) - \widetilde{T}_{\ddagger x\$}(|q_0\rangle \otimes |\epsilon\rangle, \mathbf{0}, 0)\| < \delta,$$

and thus, if x had produced y with probability at least $\alpha + \delta$, so had the shorter string vw. This means that we only have to consider input strings of length up to c to decide whether $y \in \mathrm{Range}(\mathcal{R})$. □

Remark 9 *Obviously, this reasoning applies to pfst, too.*

5 Deterministic vs. Probabilistic Transducers

Unlike the situation for finite automata, pfst are strictly more powerful than their deterministic counterparts:

Theorem 10 *For arbitrary $\epsilon > 0$ the relation*

$$\mathcal{R}_1 = \{(0^m 1^m, 2^m) : m \geq 0\}$$

can be computed by a pfst with probability $1 - \epsilon$. It cannot be computed by a dfst.

Proof: The idea is essentially from [4]: for a natural number k choose initially an alternative $j \in \{1, \dots, k-1\}$, uniformly. Then do the following: repeatedly read k 0's, and output j 2's, until the 1's start (remember the remainder modulo k), then repeatedly read k 1's, and output $k - j$ 2's. Compare the remainder modulo k with what you remembered: if the two are equal, output this number of 2's and accept, otherwise reject.

It is immediate that on input $0^m 1^m$ this machine outputs 2^m with certainty. However, on input $0^m 1^{m'}$ each 2^n receives probability at most $1/k$.

That this cannot be done deterministically is straightforward: assume that a dfst has produced $f(m)$ 2's after having read m 0's. Because of finiteness there are k, l such that after reading k 1's (while n_0 2's were output) the internal state is the same as after reading l further 1's (while n 1's are output). So, the output for input $0^m 1^{k+rl}$ is $2^{f(m)+n_0+rn}$, and these pairs are either all accepted or all rejected. Hence they are all rejected, contradicting aceptance for $m = k + rl$. □

By observing that the random choice at the beginning can be mimicked quantumly, and that all intermediate computations are in fact reversible, we immediately get

Theorem 11 *For arbitrary $\epsilon > 0$ the relation $\mathcal{R}_1$ can be computed by a qfst with probability $1 - \epsilon$.* □

Note that this puts qfst in contrast to quantum finite automata: in [2] it was shown that if a language is recognized with probability strictly exceeding 7/9 then it is possible to accept it with probability 1, i.e. reversibly deterministically.

Theorem 12 *The relation*

$$\mathcal{R}_2 = \{(w2w, w) : w \in \{0,1\}^*\}$$

can be computed by a pfst and by a qfst with probability $2/3$.

Proof: We do this only for qfst (the pfst is obtained by replacing the unitaries involved by the stochastic matrices obtained by computing the squared moduli of the entries): let the input be $x2y$ (other forms are rejected). With amplitude $1/\sqrt{3}$ each go to one of three 'subprograms': either copy x to the output, or y, or reject without output. □

6 ... vs. Quantum Transducers

After seeing a few examples one might wonder if everything that can be done by a qfst can be done by a pfst. That this is not so is shown as follows:

Theorem 13 *The relation*

$$\mathcal{R}_3 = \{(0^m 1^n 2^k, 3^m) : n \neq k \wedge (m = k \vee m = n)\}$$

can be computed by a qfst with probability $4/7 - \epsilon$*, for arbitrary* $\epsilon > 0$*.*

Theorem 14 *The relation* $\mathcal{R}_3$ *cannot be computed by a pfst with probability bounded away from* $1/2$*. In fact, not even with an isolated cutpoint.*

Proof (of theorem 13): For a natural number l construct the following transducer: from q_0 go to one of the states q_c, $[j, b]$ ($j \in \{1, \dots, l-1\}$, $b \in \{1, 2\}$), with amplitude $\sqrt{3/7}$ for q_c and with amplitude $\sqrt{2/(7l)}$ each, for the others. Then proceed as follows (we assume the form of the input to be $0^m 1^n 2^k$, others are rejected): for q_c output one 3 for each 0, and finally accept. For $[j, b]$ repeatedly read l 0's and ouput j 3's (remember the remainder $m \mod l$). Then repeatedly read l b's and output $l - j$ 3's (output nothing on the $(3 - b)$'s). Compare the remainder with the one remembered, and reject if they are unequal, otherwise output this number of 3's. Reading \$ perform the following unitary:

$$(j \mapsto j') \otimes \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Accepting are all $[j', 2]$, rejecting are all $[j', 1]$.

Now assume that the input does not occur as the left member in the relation: this means either $m \neq k$ and $m \neq n$, or $m = n = k$. In the first case all the outputs in each of the b–branches of the program are of different length, so get amplitude $\sqrt{2/(7l)}$. The final step combines at most two of them, so any output is accepted with probability at most $4/(7l)$. The second case is more interesting: in all branches the amplitude is concentrated on the output 3^m. The rotation $V_\$$ however is made such that the amplitude on $[j', 2]$ cancels out, so we end up in a rejecting state $[j', 1]$. In total, any output is accepted with probability at most $3/7 + \epsilon$.

On the other hand, if the input occurs as the left member in the relation, exactly one of the two b–branches of the program concentrates all amplitude on output 3^m, whereas the other spreads it to l different lengths. This means that the output 3^m is accepted with probability at least $(l - 1) \cdot 1/(7l)$, and others are accepted with probability at most $1/(7l)$ each. In total, the output 3^m is accepted with probability at least $4/7 - \epsilon$, all others are accepted with probability at most $3/7 + \epsilon$. □

Proof (of theorem 14): See appendix A. □

In general however, computing with isolated cupoint is strictly weaker than with probability bounded away from 1/2 (observe that for finite automata, probabilistic and quantum, recognizability with an isolated cutpoint is equivalent to recognizability with probability bounded away from 1/2, see theorem 4):

Theorem 15 *The relation*

$$\mathcal{R}_4 = \{(0^m 1^n a, 4^l) : (a = 2 \rightarrow l = m) \wedge (a = 3 \rightarrow l = n)\}$$

can be computed by a pfst and by a qfst with an isolated cutpoint, but not with a probability bounded away from $1/2$*.*

Proof: First the construction (again, only for qfst): initially branch into two possibilities c_0, c_1, each with amplitude $1/\sqrt{2}$. Assume that the input is of the correct form (otherwise reject), and in state c_i output one 4 for each i, ignoring the $(1-i)$'s. Then, if $a = 2+i$, accept, if $a = 3-i$, reject. It is easily seen that 4^l is accepted with probability $1/2$ if $(0^m 1^n a, 4^l) \in \mathcal{R}_4$, and with probability 0 otherwise.

That this cannot be done with probability above $1/2$ is clear intuitively: the machine has to produce some output (because of memory limitations), but whether to output 4^m or 4^n it cannot decide until seeing the last symbol. Formally, assume that $|m-n| > 4t$, with $t = \max_{a,q} |f_a(q)|$. If

$$T_{\ddagger 0^m 1^n 2\$}((q_0, \epsilon), \mathbf{0}, 0)[4^m] = T(4^m | 0^m 1^n 2) \geq 1/2 + \delta,$$

necessarily

$$T_{\ddagger 0^m 1^n}((q_0, \epsilon), \mathbf{0}, 0)[4^m] + T_{\ddagger 0^m 1^n}((q_0, \epsilon), \mathbf{0}, 0)[Q_{\mathrm{non}} \times 4^{[m-2t, m+2t]}] \geq 1/2 + \delta.$$

But this implies

$$T_{\ddagger 0^m 1^n}((q_0, \epsilon), \mathbf{0}, 0)[4^n] + T_{\ddagger 0^m 1^n}((q_0, \epsilon), \mathbf{0}, 0)[Q_{\mathrm{non}} \times 4^{[n-2t, n+2t]}] \leq 1/2 - \delta,$$

hence

$$T_{\ddagger 0^m 1^n 3\$}((q_0, \epsilon), \mathbf{0}, 0)[4^n] = T(4^n | 0^m 1^n 3) \leq 1/2 - \delta.$$

□

To conlude from these examples, however, that quantum is even better than probabilistic, would be premature:

Theorem 16 *The relation*

$$\mathcal{R}_5 = \{(wx, x) : w \in \{0,1\}^*, x \in \{0,1\}\}$$

cannot be computed by a qfst with an isolated cutpoint. (Obviously it is computed by a pfst with probability 1*).*

Proof: The construction of a dfst computing the relation is straightforward. To show that no qfst doing this job exists, we recall from [7] that $\{0,1\}^* 0$ is not recognized by a 1–way quantum finite automaton with probability bounded away from $1/2$, and use theorem 4 for this language. □

7 Conclusion

We introduced quantum finite state transducers, and showed some of their unique properties: undecidability of the emptiness problem, as opposed to deterministic finite state transducers and finite automata, and incomparability of their power to that of probabilistic and deterministic finite state transducers. As open questions we would like to point out primarily our conjecture 7. Another interesting question is whether a relation computed by a qfst with probability *sufficiently close to* 1 can be computed by a pfst. This would be the closest possible analog to the "7/9–theorem" from [2].

A Appendix: Proof of theorem 14

Let $\mathcal{R}_3$ be computed by a pfst T with isolated cutpoint α. The following construction computes it with probability bounded away from 1/2: assuming $\alpha \leq 1/2$ (the other case is similar), let $p = \frac{1/2-\alpha}{1-\alpha}$. Run one of the following subprograms probabilistically: with probability p output one 3 for each 0, and ignore the other symbols (we may assume that the input has the form $0^m 1^n 2^k$), with probability $1-p$ run T on the input. It is easily seen that this new pfst computes the same relation with probability bounded away from 1/2.

Hence, we may assume that T computes $\mathcal{R}$ with probability $\varphi > 1/2$, from this we shall derive a contradiction. The state set Q together with any of the stochastic matrices V_0, V_1, V_2 is a Markov chain. We shall use the classification of states for finite Markov chains (see [6]): for V_i Q is partitioned into the set R_i of *transient* states (i.e. the probability to find the process in R_i tends to 0) and a number of sets S_{ij} of *ergodic* states (i.e. once in S_{ij} the process does not leave this set, and all states inside can be reached from each other, though maybe only by a number of steps). Each S_{ij} is divided further into its *cyclic* classes $C_{ij\nu}$ ($\nu \in \mathbb{Z}_{d_{ij}}$), V_i mapping $C_{ij\nu}$ into $C_{ij\nu+1}$. By considering sufficiently high powers V_i^d (e.g. product of all the periods d_{ij}) as transition matrices, all these cyclic sets become ergodic, in fact, V_i^d restricted to each is *regular.*

Using only these powers amounts to concentrating on input of the form $\mathbf{0}^m \mathbf{1}^n \mathbf{2}^k$, with $\mathbf{i} = i^d$, which we will do from now on. Relabelling, the ergodic sets of $V_\mathbf{i} = V_i^d$ will be denoted S_{ij}. Each has its unique equilibrium distribution, to which every initial one converges: denote it by π_{ij}. Furthermore, there are limit probabilities $a(j_0)$ to find the process $V_\mathbf{0}$ in S_{0j_0} after long time, starting from q_0. Likewise, there are limit probabilities $b(j_1|j_0)$ to find the process $V_\mathbf{1}$ in S_{1j_1} after long time, starting from π_{0j_0}, and similarly $c(j_2|j_1)$. So, by the law of large numbers, for large enough m, n, k the probability that $V_\mathbf{0}$ has passed into S_{0j_0} after $\sqrt{m}$ steps, after which $V_\mathbf{1}$ has passed into S_{1j_1} after $\sqrt{n}$ steps, after which $V_\mathbf{2}$ has passed into S_{2j_2} after $\sqrt{k}$ steps, is arbitrarily close to $P(j_0, j_1, j_2) = a(j_0)b(j_1|j_0)c(j_2|j_1)$. (Note that these probabilities sum to one).

As a consequence of the ergodic theorem (or law of large numbers), see [6], ch. 4.2, in each of these events $J = (j_0, j_1, j_2)$ the probable number of 3's written after the final \$, is linear in m, n, k:

$$T(3^{[(1-\delta)\lambda_J(m,n,k),(1+\delta)\lambda_J(m,n,k)]}|\mathbf{0}^m\mathbf{1}^n\mathbf{2}^k, J) \to 1,$$

as $m, n, k \to \infty$, with

$$\lambda_J(m,n,k) = \alpha_J m + \beta_J n + \gamma_J k,$$

and non–negative constants $\alpha_J, \beta_J, \gamma_J$.

Since we require that for $k \neq m$

$$T(3^{dm}|\mathbf{0}^m\mathbf{1}^m\mathbf{2}^k) \geq \varphi,$$

it is necessary that for a set $\mathcal{A}$ of events $J = (j_0, j_1, j_2)$

$$\alpha_J + \beta_J = d, \ \gamma_J = 0, \ \text{with } P(\mathcal{A}) \geq \varphi.$$

In fact, as for $J \notin \mathcal{A}$

$$T(3^{dm}|\mathbf{0}^m\mathbf{1}^m\mathbf{2}^k, J) \to 0$$

for certain sequences $m, k \to \infty$, we even have

$$\sum_{J\in\mathcal{A}} P(J)T(3^{dm}|\mathbf{0}^m\mathbf{1}^m\mathbf{2}^k, J) \geq \varphi - o(1).$$

For $J \in \mathcal{A}$ it is obvious that the transducer outputs *no more* 3's, once in S_{2j_2}. But this implies that for m, k large enough, $T(3^{dm}|\mathbf{0}^m\mathbf{1}^m\mathbf{2}^k, J)$ is arbitrarily close to $T(3^{dm}|\mathbf{0}^m\mathbf{1}^m\mathbf{2}^m, J)$, hence

$$T(3^{dm}|\mathbf{0}^m\mathbf{1}^m\mathbf{2}^m) \geq \varphi - o(1),$$

which implies that

$$T(3^{dm}|\mathbf{0}^m\mathbf{1}^m\mathbf{2}^m) \geq \varphi,$$

contradicting $(0^{dm}1^{dm}2^{dm}, 3^{dm}) \notin \mathcal{R}_3$. □

References

[1] M. Amano, K. Iwama, "Undecidability on Quantum Finite Automata", in Proc. 31st STOC, 1999, pp. 368–375.

[2] A. Ambainis, R. Freivalds, "1–way quantum finite automata: strengths, weaknesses, and generalizations", in Proc. 39th FOCS, 1998, pp. 332–341.

[3] R. Bonner, R. Freivalds, R. Gailis, "Undecidability of 2–tape quantum finite automata", in *Quantum Computation and Learning. Proceedings of an international workshop. Sundbyholms Slott, Sweden, 27–29 May, 2000*, R. Bonner and R. Freivalds (eds.), Malardalen University, 2000, pp. 93–100.

[4] R. Freivalds, "Language recognition using finite probabilistic multitape and multihead automata", Problems Inform. Transmission, vol. 15, no. 3, 1979, pp. 235–241.

[5] E. Gurari, *Introduction to the Theory of Computation*, Computer Science Press, 1989.

[6] J. G. Kemeny, J. L. Snell, *Finite Markov Chains*, Van Nostrand, Princeton, 1960.

[7] A. Kondacs, J. Watrous, "On the power of quantum finite state automata", in Proc. 38th FOCS, 1997, pp. 66–75.

[8] M. O. Rabin, "Probabilistic Automata", Information and Control, vol. 6, 1963, pp. 230–245.

[9] D. Scott, "Some definitional suggestions for automata theory", J. of Comput. and Syst. Science, 1967, pp. 187–212.