Combinatorial Characterization of Transducers with Bounded Variance

Sara Kropf

Alpen-Adria-Universität Klagenfurt

Joint work with Clemens Heuberger and Stephan Wagner

AofA
Strobl, June 12, 2015

Motivation

Theorem (Hwang's Quasi-Power-Theorem)

Let Ω_n be a sequence of real random variables. Suppose the moment generating function satisfies

$$\mathbb{E}(e^{\Omega_n s}) = e^{u(s)\Phi(n)+v(s)}(1+\mathcal{O}(\kappa_n^{-1}))$$

under some conditions.
Then

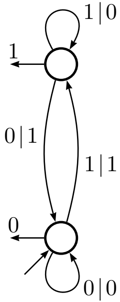
$$\mathbb{E}\Omega_n = u'(0)\Phi(n) + \mathcal{O}(1),$$
$$\mathbb{V}\Omega_n = u''(0)\Phi(n) + \mathcal{O}(1).$$

If $\sigma^2 := u''(0) \neq 0$, *then* $\frac{\Omega_n - \mathbb{E}\Omega_n}{\sqrt{\mathbb{V}\Omega_n}}$ *is asymptotically normally distributed.*

When is the variance bounded?

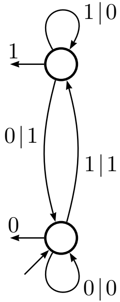
Transducers

- transducer $\mathcal{T}$ with a finite number of states



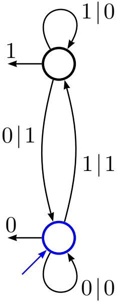
Transducers

- transducer $\mathcal{T}$ with a finite number of states
- $\text{Output}(X_n) =$ sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left



Transducers

- transducer $\mathcal{T}$ with a finite number of states
- $\text{Output}(X_n)$ = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left



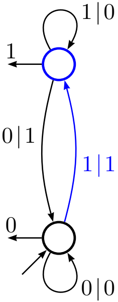
Example with $X_n = 11001$

input: 11001
output:

$\text{Output}(11001) =$

Transducers

- transducer $\mathcal{T}$ with a finite number of states
- $\text{Output}(X_n)$ = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left

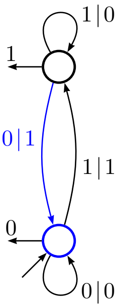


Example with $X_n = 11001$

input:	11001	$\text{Output}(11001) =$
output:	1	

Transducers

- transducer $\mathcal{T}$ with a finite number of states
- Output(X_n) = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left



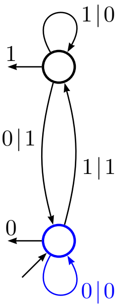
Example with $X_n = 11001$

input:	11001
output:	11

Output(11001) =

Transducers

- transducer $\mathcal{T}$ with a finite number of states
- $\text{Output}(X_n) = $ sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left

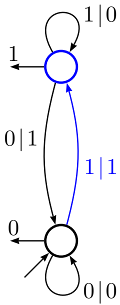


Example with $X_n = 11001$

input:	11001	Output(11001) =
output:	011	

Transducers

- transducer $\mathcal{T}$ with a finite number of states
- Output(X_n) = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left



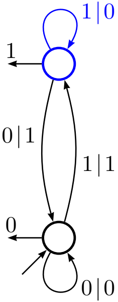
Example with $X_n = 11001$

input: 11001
output: 1011

Output(11001) =

Transducers

- transducer $\mathcal{T}$ with a finite number of states
- $\text{Output}(X_n)$ = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left



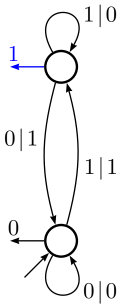
Example with $X_n = 11001$

input:	11001
output:	01011

Output(11001) =

Transducers

- transducer $\mathcal{T}$ with a finite number of states
- Output(X_n) = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left



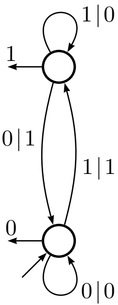
Example with $X_n = 11001$

input:	11001
output:	101011

Output(11001) =

Transducers

- transducer $\mathcal{T}$ with a finite number of states
- Output(X_n) = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left

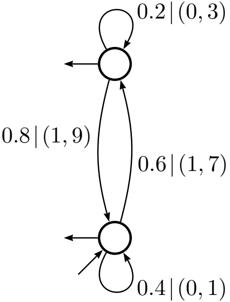


Example with $X_n = 11001$

input:	11001
output:	101011

Output(11001) = 4

Other Probability Model and Several Outputs



All results also possible for:

- inputs coming from a Markov chain
- for every transition a probability
- sum of probabilities of output transitions is 1

Some results are independent of the choice of this Markov chain.

Several simultaneous outputs.

Applications

- algorithms with finite memory usage
- many digit expansions:
  - Hamming weight
  - sum of digits function, ...
- many recursions
- motifs

Applications

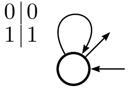
- algorithms with finite memory usage
- many digit expansions:
  - Hamming weight
  - sum of digits function, ...
- many recursions
- motifs

- completely q-additive functions
- digital sequences
- q-automatic sequences

ALPEN-ADRIA UNIVERSITÄT
KLAGENFURT I WIEN GRAZ

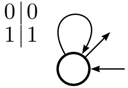
Applications

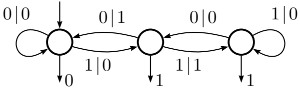
- digit sum of binary expansion



Applications

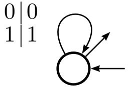
- digit sum of binary expansion
- Hamming weight of non-adjacent form (NAF):
  - digits $\{0, \pm 1\}$, base 2
  - at least one of any two adjacent digits is 0

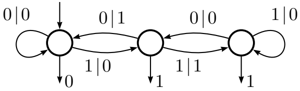


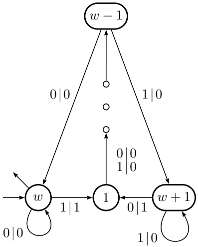


Applications

- digit sum of binary expansion
- Hamming weight of non-adjacent form (NAF):
  - digits $\{0, \pm 1\}$, base 2
  - at least one of any two adjacent digits is 0
- Hamming weight of width-w NAF:
  - digits $\{0, \pm 1, \pm 3, \dots, \pm(2^{w-1} - 1)\}$, base 2
  - at least $w - 1$ of w consecutive digits are 0







Variability Condition

Theorem (Hwang's Quasi-Power-Theorem)

Let Ω_n be a sequence of real random variables. Suppose the moment generating function satisfies

$$\mathbb{E}(e^{\Omega_n s}) = e^{u(s)\Phi(n)+v(s)}(1+\mathcal{O}(\kappa_n^{-1}))$$

under some conditions.
Then

$$\mathbb{E}\Omega_n = u'(0)\Phi(n) + \mathcal{O}(1),$$
$$\mathbb{V}\Omega_n = u''(0)\Phi(n) + \mathcal{O}(1).$$

If $\sigma^2 := u''(0) \neq 0$, then $\frac{\Omega_n - \mathbb{E}\Omega_n}{\sqrt{\mathbb{V}\Omega_n}}$ is asymptotically normally distributed.

Assume that $\mathcal{T}$ is strongly connected.
Output(X_n) satisfies all asumptions, except maybe the variability condition $\sigma^2 \neq 0$.

Bounded Variance

Theorem (Heuberger–K.–Wagner 2015)

Let $\mathcal{T}$ be strongly connected. Then the following assertions are equivalent:

1. *The asymptotic variance σ^2 is 0.*
2. *There is a constant k such that the average output of every cycle is k.*
3. *There is a constant k such that $\operatorname{Output}(X_n) = kn + \mathcal{O}(1)$.*

Bounded Variance

Theorem (Heuberger–K.–Wagner 2015)

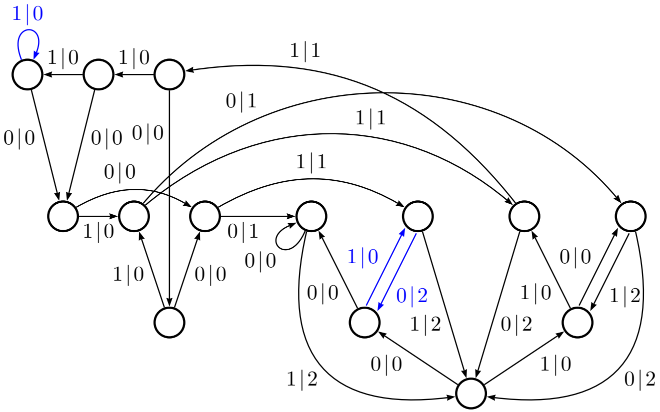
Let $\mathcal{T}$ be strongly connected. Then the following assertions are equivalent:

1. *The asymptotic variance σ^2 is 0.*
2. *There is a constant k such that the average output of every cycle is k.*
3. *There is a constant k such that $\mathrm{Output}(X_n) = kn + \mathcal{O}(1)$.*

Corollary (Heuberger–K.–Wagner 201)

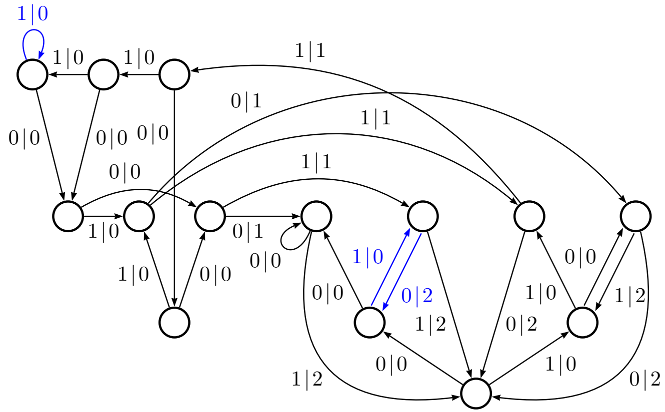
Let $\mathcal{T}$ be strongly connected, aperiodic with output alphabet $\{0, 1\}$.
Then the asymptotic variance σ^2 is 0 if and only if all output letters are the same.

Small Example



⇝ asymptotic variance $\neq 0$

Small Example



⇝ asymptotic variance $\neq 0$

Sage: $\sigma^2 = \frac{432}{2197}$

Example: τ-adic Digit Expansion

- algebraic integer τ
- joint expansion of d-dimensional vectors in $\mathbb{Z}[\tau]^d$
- redundant digit set $\mathcal{D}$ which satisfies
  - $\mathcal{D} \cap \tau\mathbb{Z}^d = \{0\}$
  - a subadditivity condition

Example: τ-adic Digit Expansion

- algebraic integer τ
- joint expansion of d-dimensional vectors in $\mathbb{Z}[\tau]^d$
- redundant digit set $\mathcal{D}$ which satisfies
  - $\mathcal{D} \cap \tau\mathbb{Z}^d = \{0\}$
  - a subadditivity condition
- input: τ-adic expansions with the irredundant digit set $\mathcal{A}$ of length $\leq n$ with equidistribution

Example: τ-adic Digit Expansion

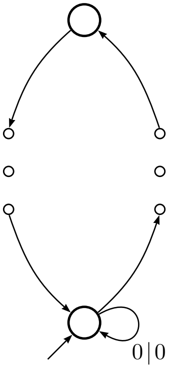
- algebraic integer τ
- joint expansion of d-dimensional vectors in $\mathbb{Z}[\tau]^d$
- redundant digit set $\mathcal{D}$ which satisfies
  - $\mathcal{D} \cap \tau\mathbb{Z}^d = \{0\}$
  - a subadditivity condition
- input: τ-adic expansions with the irredundant digit set $\mathcal{A}$ of length $\leq n$ with equidistribution

Theorem (Heigl–Heuberger 2012)

If the asymptotic variance σ^2 of the minimal Hamming weight with digit set $\mathcal{D}$ is $\neq 0$, then the minimal Hamming weight is asymptotically normally distributed.

Example: τ-adic Digit Expansion

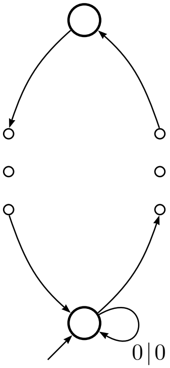
Heigl–Heuberger construct a transducer for each τ and $\mathcal{D}$:



- cycle with average output 0

Example: τ-adic Digit Expansion

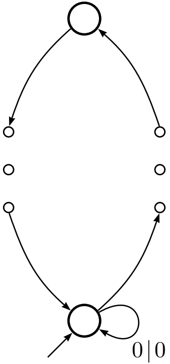
Heigl–Heuberger construct a transducer for each τ and $\mathcal{D}$:



- cycle with average output 0
- but not all minimal weights are 0
- $0 \cdots 0$ always leads to the initial state
- $\rightsquigarrow$ cycle with average output $\neq 0$

Example: τ-adic Digit Expansion

Heigl–Heuberger construct a transducer for each τ and $\mathcal{D}$:



- cycle with average output 0
- but not all minimal weights are 0
- $0 \cdots 0$ always leads to the initial state
- $\rightsquigarrow$ cycle with average output $\neq 0$
- variability condition is satisfied
- $\rightsquigarrow$ asymptotic normality

Bounded Variance

Theorem (Heuberger–K.–Wagner 2015)

Let $\mathcal{T}$ be strongly connected. Then the following assertions are equivalent:

1. *The asymptotic variance σ^2 is 0.*
2. *There is a constant k such that the average output of every cycle is k.*
3. *There is a constant k such that $\operatorname{Output}(X_n) = kn + \mathcal{O}(1)$.*

Idea of the Proof of the Theorem

$1 \Leftrightarrow 2$:

- assume: asymptotic expected value of $\mathrm{Output}(X_n)$ is 0
- probability generating function

$$A(y,z) = \sum_{l \in \mathbb{R}} \sum_{n=0}^{\infty} a_{ln} K^{-n} y^l z^n$$

  with $K = |\mathcal{A}|$ and a_{ln} = number of input words of length n with output sum l
- $A(1,z)$ has a simple dominant pole at $z = 1$

Idea of the Proof of the Theorem

$1 \Leftrightarrow 2$:

- assume: asymptotic expected value of Output(X_n) is 0
- probability generating function

$$A(y,z) = \sum_{l \in \mathbb{R}} \sum_{n=0}^{\infty} a_{ln} K^{-n} y^l z^n$$

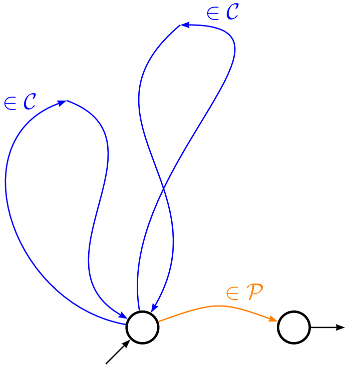
  with $K = |\mathcal{A}|$ and a_{ln} = number of input words of length n with output sum l
- $A(1,z)$ has a simple dominant pole at $z = 1$
-

$$\mathbb{E}(\text{Output}(X_n)) = [z^n]A_y(1,z) = \mathcal{O}(1)$$
$$\mathbb{V}(\text{Output}(X_n)) = [z^n]A_{yy}(1,z) + \mathcal{O}(1)$$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

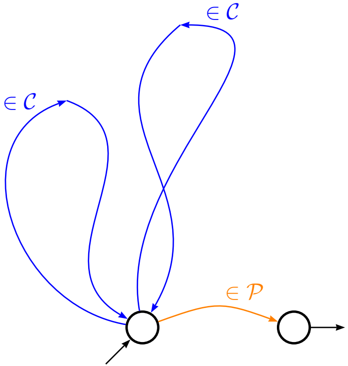
Idea of the Proof of the Theorem

Decomposition:



Idea of the Proof of the Theorem

Decomposition:

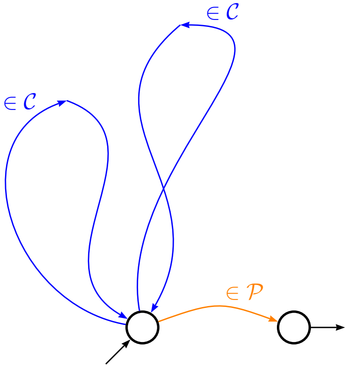


- probability generating functions

$$C(y, z), \; P(y, z)$$

Idea of the Proof of the Theorem

Decomposition:



- probability generating functions

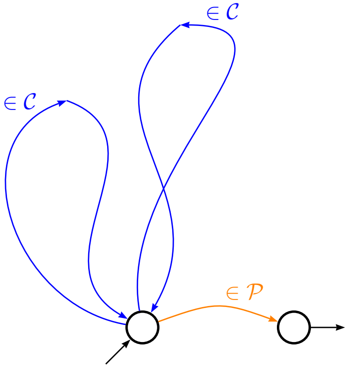
$$C(y,z),\ P(y,z)$$

- by the symbolic method:

$$A(y,z) = \frac{1}{1 - C(y,z)} P(y,z)$$

Idea of the Proof of the Theorem

Decomposition:



- probability generating functions

$$C(y,z),\ P(y,z)$$

- by the symbolic method:

$$A(y,z) = \frac{1}{1 - C(y,z)} P(y,z)$$

- $P(1,z)$ is analytic in $|z| < 1 + \varepsilon$
- $P(1,1) \neq 0$
- $1 - C(1,z) = (1-z)g(z)$ with $g(1) \neq 0$

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Idea of the Proof of the Theorem

- Singularity Analysis $\rightsquigarrow$

$$\mathbb{V}(\mathsf{Output}(X_n)) = P(1,1)g(1)^{-2}C_{yy}(1,1)n + \mathcal{O}(1)$$

- thus,

$$\begin{aligned} & \mathbb{V}(\mathsf{Output}(X_n)) = \mathcal{O}(1) \\ \iff \quad & C_{yy}(1,1) = 0 \\ \iff \quad & \sum_{C \in \mathcal{C}} \mathsf{Output}(C)^2 K^{-\mathsf{Length}(C)} = 0 \\ \iff \quad & \forall C \in \mathcal{C}: \ \mathsf{Output}(C) = 0 \end{aligned}$$

□

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Singular Variance-Covariance Matrix

Consider m different outputs $k_1, \ldots, k_m$ of a transducer instead of Output.
Using a multi-dimensional Quasi-Power-Theorem:

Theorem (K. 2015+)

The m output sums are asymptotically jointly normally distributed, if and only if:

$$a_0 Length(C) + a_1 k_1(C) + \cdots + a_m k_m(C) = 0$$

holding for all cycles C implies that $a_0 = \cdots = a_m = 0$.

Bounded Covariance

- random variable $(\text{Input}(X_n), \text{Output}(X_n))$
- 2-dimensional version of the Quasi-Power-Theorem
- $\rightsquigarrow$ asymptotic normal distribution

Bounded Covariance

- random variable $(\text{Input}(X_n), \text{Output}(X_n))$
- 2-dimensional version of the Quasi-Power-Theorem
- $\rightsquigarrow$ asymptotic normal distribution
- When is the covariance bounded?
- covariance bounded $\leftrightarrow$ components of the asymptotic random variable are independent

Definition

An independent transducer is a transducer which has a bounded covariance of $(\text{Input}(X_n), \text{Output}(X_n))$.

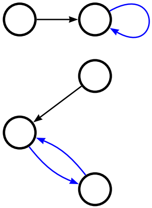
ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

Functional Digraph

Definition (Functional Digraph)

A functional digraph is a directed graph where every vertex has out-degree 1.

This is a map from a finite set into itself.



Functional Digraph

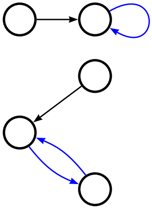
Definition (Functional Digraph)

A functional digraph is a directed graph where every vertex has out-degree 1.

This is a map from a finite set into itself.

Definition

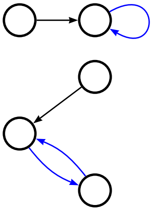
$\mathcal{D}_1$ and $\mathcal{D}_2$ are the sets of functional digraphs with one respectively two components which are subgraphs of the given transducer.



Bounded Covariance

$$\text{InputOutput}(\mathcal{D}_1) = \sum_{D \in \mathcal{D}_1} \text{Input(cycle)Output(cycle)},$$

$$\text{InputOutput}(\mathcal{D}_2) = \sum_{D \in \mathcal{D}_2} \text{Input(one cycle)Output(other cycle)}$$



Bounded Covariance

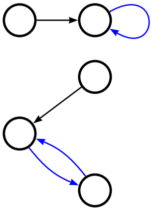
$$\text{InputOutput}(\mathcal{D}_1) = \sum_{D \in \mathcal{D}_1} \text{Input(cycle)Output(cycle)},$$

$$\text{InputOutput}(\mathcal{D}_2) = \sum_{D \in \mathcal{D}_2} \text{Input(one cycle)Output(other cycle)}$$

Theorem (Heuberger–K.–Wagner 2015)

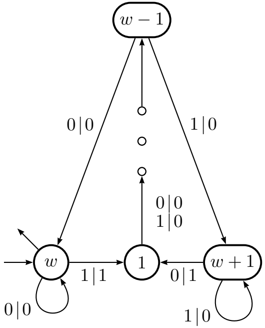
Suppose the asymptotic expected value of $(\text{Input}(X_n), \text{Output}(X_n))$ *is* $(0, 0)$. *Then the transducer is independent if and only if*

$$\text{InputOutput}(\mathcal{D}_2) = \text{InputOutput}(\mathcal{D}_1).$$



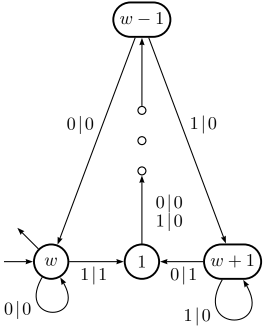
Also possible: 2 outputs, Markov chain

Width-w Non-Adjacent Form



- asymptotic covariance $= 0$
- arbitrarily large independent transducers
- Hamming weight of binary expansion and Hamming weight of w-NAF are independent
- $w = 2$: NAF (Heuberger–Prodinger 2007)

Width-w Non-Adjacent Form



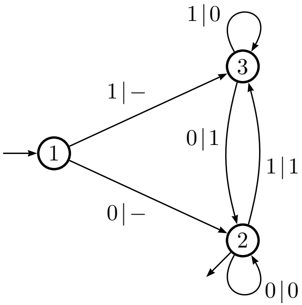
$2 \leq w_1 < w_2$ with $w_1 \neq w_2 - 1$:

- closed walk with input 0
- closed walk with input 10^{w_2-1}
- closed walk with input $10^{w_1-1}10^{w_1-1}0\cdots0$

$$\Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ * & 1 & 1 \\ * & 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = 0$$

$\rightsquigarrow$ asymptotic normal distribution

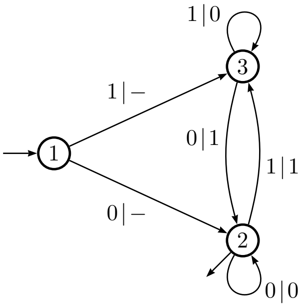
Gray Code



First values:

0	0	6	101
1	1	7	100
2	11	8	1100
3	10	9	1101
4	110	10	1111
5	111	11	1110

Gray Code

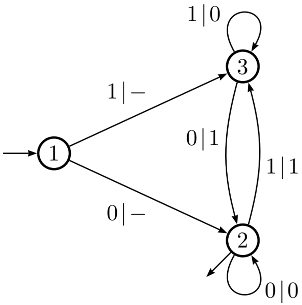


First values:

0	0	6	101
1	1	7	100
2	11	8	1100
3	10	9	1101
4	110	10	1111
5	111	11	1110

- starting transitions unimportant

Gray Code



First values:

0	0	6	101
1	1	7	100
2	11	8	1100
3	10	9	1101
4	110	10	1111
5	111	11	1110

- starting transitions unimportant
- asymptotic covariance $= 0$
- independent transducer
- Hamming weight of binary expansion and Hamming weight of Gray code are independent

Conclusion

- combinatorial description for transducers with
  - bounded variance
  - singular variance-covariance matrix
  - bounded covariance
- $\rightsquigarrow$ asymptotically normally distributed
- can be checked
  - without long computations
  - in general settings