Ancestral lines under selection and recombination

Ellen Baake

Bielefeld University

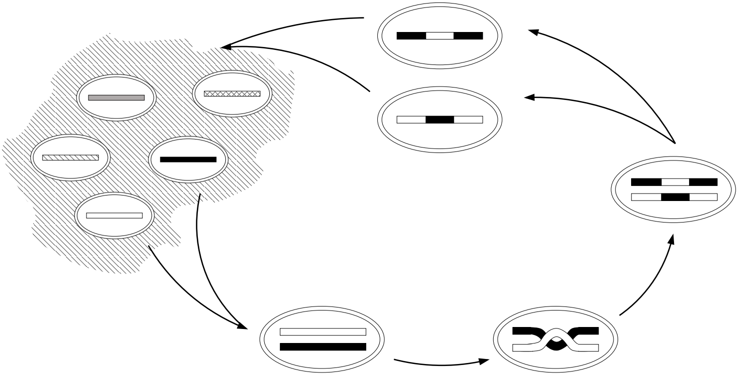
joint work with Frederic Alberti

1. The selection-recombination equation (forward in time)
2. Its solution via genealogical thinking
3. The dual process

PROBABILISTIC STRUCTURES
IN EVOLUTION

DFG SPP 1590

Recombination



Sequences and selection

individual: sequence of n sites, $S = \{1, \dots, n\}$

types: $x := (x_1, \dots, x_n) \in \bigtimes_{i \in S} X_i =: X$, $X_i = \{0, 1\}$

marginal types: $x_U := (x_i)_{i \in U}$, $U \subseteq S$

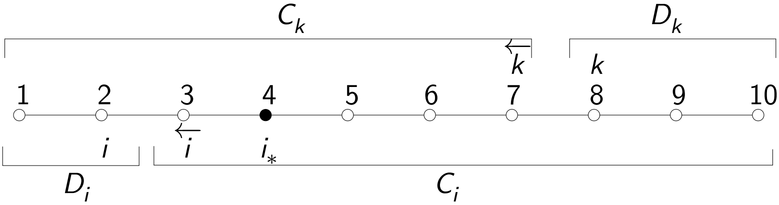
reproduction and selection:

- single selected site: $i_* \in S$ (all others 'neutral')
- types x with $x_{i_*} = 1$: reproduce at rate 1 ('bad, less fit')
- types x with $x_{i_*} = 0$: reproduce at rate $1 + s$ ('good, fit')

Ordering the sites

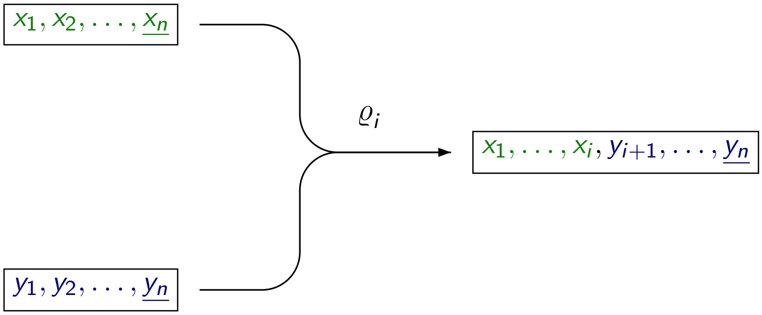
partial order on S:

- $i \preccurlyeq j$ if either $i_* \leqslant i \leqslant j$ or $i_* \geqslant i \geqslant j$
- $i \prec j$ if $i \preccurlyeq j$ and $i \neq j$
- $\overleftarrow{i}$ maximal element that precedes i (predecessor)
- $D_i := \{j \in S \mid i \preccurlyeq j\}$ (i-tail), $C_i := S \setminus D_i$ (i-head), $i \in S \setminus i_*$

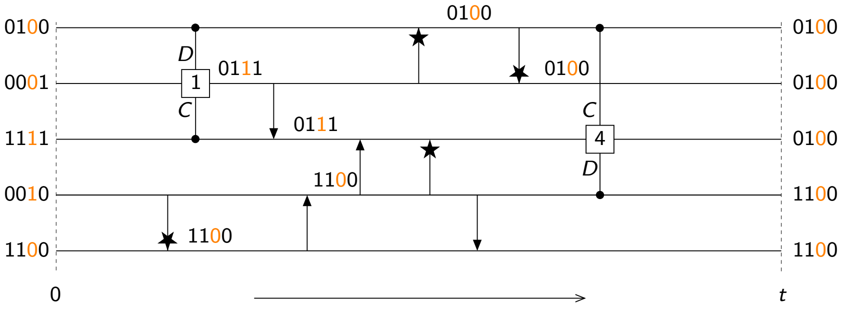


- $\rightsquigarrow i_* \in C_i$
- (single-crossover) recombination between C_i and D_i, $i \in S \setminus i_*$ (partitions S into C_i and D_i) at rate ϱ_i

Sequences and recombination



Moran model with selection and recombination



$N (= 5)$ individuals, $n = 4$ sites, selected site $i_* = 3$

→	neutral reproduction, rate 1	(for all individuals)
→★	selective reproduction, rate s	(for $(**0*)$ ind.'s)
C [i] D	reco at site i, rate ϱ_i, $\quad i \in S \setminus i_*$	(for all individuals)

Selection-recombination equation

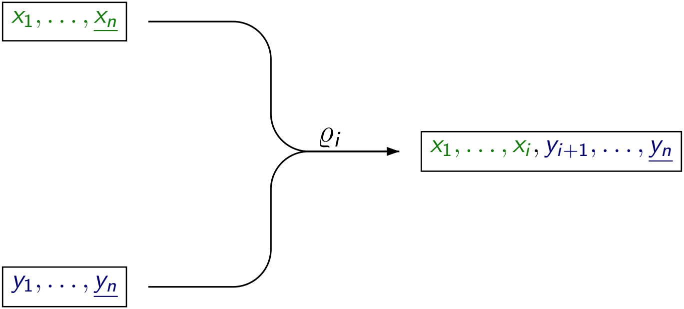
- $Z_t^N = \big(Z_t^N(x)\big)_{x \in X}$ counting measure on X
- $Z_t^N(x)$ $\sharp$ individuals of type x at time t, $x \in X$
- deterministic limit (dynamical law of large numbers):
  $N \to \infty$ w/o rescaling of parameters or time $\rightsquigarrow$

$$\frac{1}{N}(Z_t^N)_{t \geqslant 0} \longrightarrow (\omega_t)_{t \geqslant 0}$$

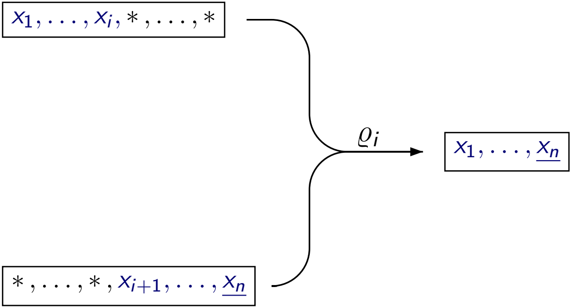
- $\omega_t = \big(\omega_t(x)\big)_{x \in X}$ probability measure on X
  solves selection-recombination equation (SRE)

$$\dot{\omega}_t(x) = s\big[(1 - x_{i_*}) - \omega_t(*, 0, *)\big]\,\omega_t(x) + \sum_{i \in S \setminus i_*} \varrho_i \big[\omega_t(x_{C_i}, *)\,\omega_t(*, x_{D_i}) - \omega_t(x)\big], \quad x \in X$$

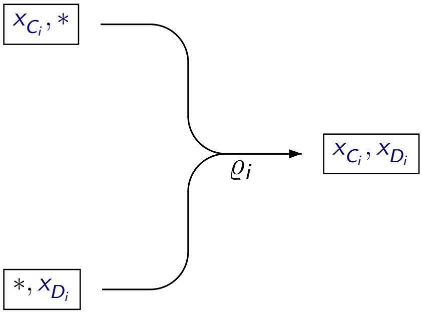
Recombination



Recombination



Recombination



Selection-recombination equation

- $Z_t^N = \left(Z_t^N(x)\right)_{x \in X}$ counting measure on X
- $Z_t^N(x)$ $\sharp$ individuals of type x at time t, $x \in X$
- deterministic limit:

  $N \to \infty$ w/o rescaling of parameters or time $\rightsquigarrow$

$$\frac{1}{N}(Z_t^N)_{t \geqslant 0} \longrightarrow (\omega_t)_{t \geqslant 0}$$

- $\omega_t = \left(\omega_t(x)\right)_{x \in X}$ probability measure on X

  solves selection-recombination equation (SRE)

$$\dot{\omega}_t(x) = s\left[(1 - x_{i_*}) - \omega_t(*, 0, *)\right]\omega_t(x) + \sum_{i \in S \setminus i_*} \varrho_i \left[\omega_t(x_{C_i}, *)\,\omega_t(*, x_{D_i}) - \omega_t(x)\right], \quad x \in X$$

Recombinators

- canonical projection: for $\varnothing \neq U \subseteq S$,

$$\pi_U : X \to \bigtimes_{i \in U} X_i = X_U, \quad \pi_U(x) = (x_i)_{i \in U} = x_U$$

- marginal measure wrt sites in U: for $\nu \in \mathcal{P}(X)$,

$$\pi_U.\nu = \nu \circ \pi_U^{-1} =: \nu^U$$

type distribution of sites in U
for $x_U \in X_U$: $\quad \nu^U(x_U) = \nu(x_U, *)$

- recombinator: for $i \in S \setminus i_*$,

$$\mathcal{P}(X) \longrightarrow \mathcal{P}(X)$$

$$R_i(\nu) := \nu^{C_i} \otimes \nu^{D_i}$$

$$\big(R_i(\nu)\big)(x) = \nu(x_{C_i}, *) \cdot \nu(*, x_{D_i})$$

distribution of sequences randomly pieced together from i-heads and i-tails

Selection-recombination equation

$$\dot{\omega}_t(x) = s\big[(1 - x_{i_*}) - \omega_t^{\{i_*\}}(0)\big]\,\omega_t(x) + \sum_{i \in S \setminus i_*} \varrho_i \big[\omega_t(x_{C_i}, *)\,\omega_t(*, x_{D_i}) - \omega_t(x)\big], \quad x \in X$$

or

$$\dot{\omega}_t = s\big(F - f(\omega_t)\big)\omega_t + \sum_{i \in S \setminus i_*} \varrho_i (R_i - \mathbb{1})\omega_t$$

with

$$F\omega(x) = (1 - x_{i_*})\omega(x)$$

and

$$f(\omega) := \omega^{\{i_*\}}(0)$$

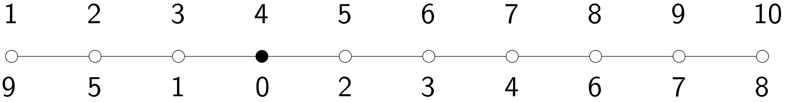
(proportion of fit individuals = proportion of individuals with 0 at selected site)

Research context

- hitchhiking: neutral x_i ($i \neq i_*$) selectively favoured if linked with $x_{i_*} = 0$
- recombination and selection on same time scale (unlike Cuthbertson, Etheridge, Yu 2012; Pfaffelhuber, Studeny 2007; Bossert, Pfaffelhuber 2017)
- Stephan, Song, Langley 2006: approximate solution forward in time ($n = 3$, wild approximations, technical result, not generalisable)

Recursive solution

- $(i_k)_{0 \leqslant k < n}$ non-decreasing permutation of S (sensu $\preccurlyeq$)
- $i_0 = i_*$
- $C^{(k)} = C_{i_k}, D^{(k)} = D_{i_k}, \varrho^{(k)} := \varrho_{i_k}$



hierarchy of solutions:

- $\left(\omega_t^{(0)}\right)_{t \geqslant 0}$ solution of SRE with $\varrho^{(\ell)} = 0$ for all ℓ (pure selection equation)
- $\left(\omega_t^{(k)}\right)_{t \geqslant 0}$ solution of SRE truncated at k (with $\varrho^{(\ell)} = 0$ for all $\ell > k$), $\quad 1 \leqslant k < n$

Recursive solution

Theorem

The solution of the pure selection equation is given by

$$\omega_t^{(0)} = \frac{\mathrm{e}^{stF}\omega_0}{\mathrm{e}^{st} f(\omega_0) + 1 - f(\omega_0)},$$

and the solutions $\omega_t^{(k)}$ *can be computed via the recursion*

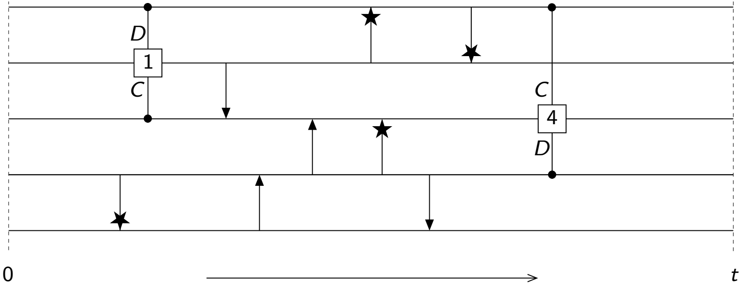
$$\omega_t^{(k)} = \mathrm{e}^{-\varrho^{(k)} t}\omega_t^{(k-1)} + \pi_{C^{(k)}}.\omega_t^{(k-1)} \otimes \int_0^t \varrho^{(k)} \mathrm{e}^{-\varrho^{(k)}\tau} \pi_{D^{(k)}}.\omega_\tau^{(k-1)} \, d\tau.$$

analytical proof ✓

genealogical content?

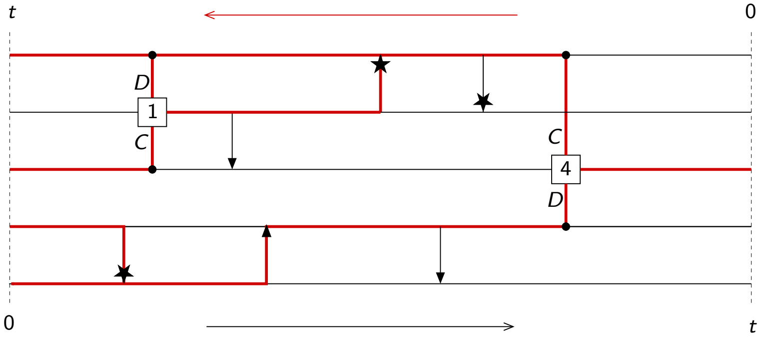
Genealogical approach

Moran model:



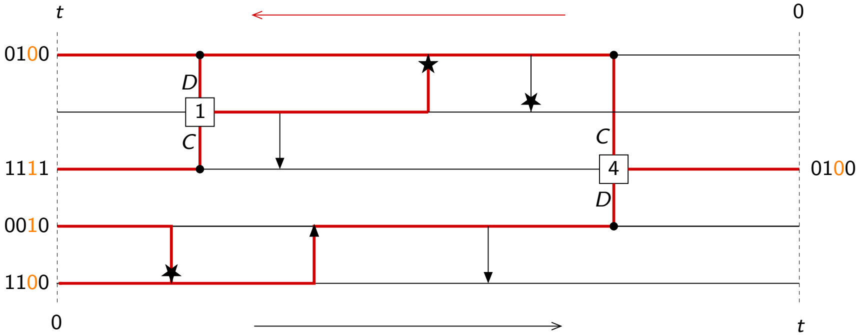
Genealogical approach

Ancestral selection-recombination graph (ASRG):

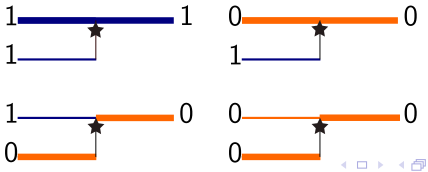


Genealogical approach

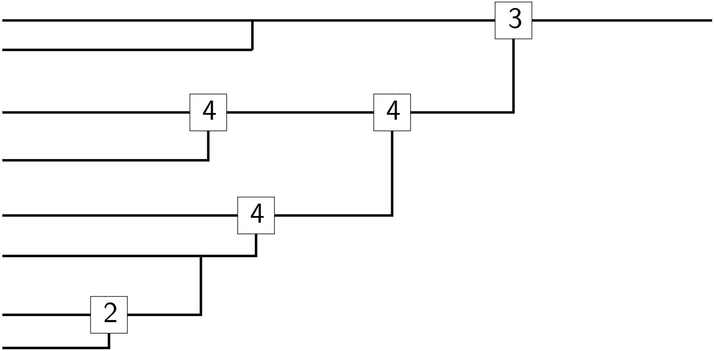
Ancestral selection-recombination graph (ASRG):



pecking order:

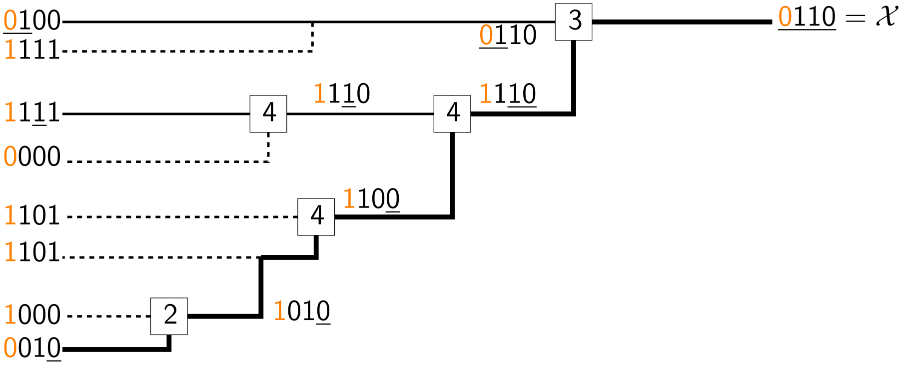


$N \to \infty$ limit of ASRG



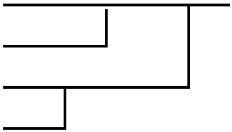
- no coalescence $\rightsquigarrow$ branches conditionally independent
- every line branches at rate s, splits at rates ϱ_i, $i \in S \setminus i_*$ (here $i_* = 1$)
- pecking order and 'heads up' order

$N \to \infty$ limit of ASRG



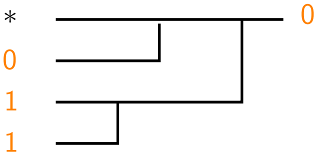
- run ASRG backward
- sample types according to ω_0 (iid)
- propagate forward by pecking and glueing
- $\mathcal{X} \sim \omega_t$
- ancestral lines

Selection only: $\omega_t^{(0)}$



- $\varrho^{(\ell)} = 0$, for all $\ell \rightsquigarrow$ pure branching
  $\rightsquigarrow$ Yule process at rate s ('ASG')
- $(K_t)_{t \geqslant 0}$ line-counting process
- $\mathbb{P}(K_t = k) = \mathrm{e}^{-st}(1 - \mathrm{e}^{-st})^{k-1}$
- $\mathbb{E}(K_t) = \mathrm{e}^{st}$
- $f(\omega_t) = 1 - \mathbb{E}\Big[\big(1 - f(\omega_0)\big)^{K_t}\Big] = \dfrac{\mathrm{e}^{st} f(\omega_0)}{\mathrm{e}^{st} f(\omega_0) + 1 - f(\omega_0)}$

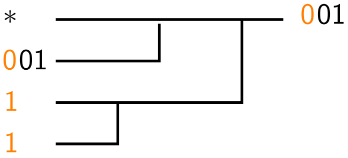
Selection only: $\omega_t^{(0)}$



$$\omega_t^{(0)} = \mathcal{L}(\mathcal{X}) = \mathbb{E}\Big[\big(1 - f(\omega_0)\big)^{K_t}\Big]\, \omega_0\big(\,.\mid \widetilde{\mathcal{X}}_{i_*} = 1\big) + \Big(1 - \mathbb{E}\big[(1 - f(\omega_0))^{K_t}\big]\Big)\, \omega_0\big(\,.\mid \widetilde{\mathcal{X}}_{i_*} = 0\big)\Big]$$

$$= \frac{\mathrm{e}^{stF}\omega_0}{\mathrm{e}^{st} f(\omega_0) + 1 - f(\omega_0)}$$

Selection only: $\omega_t^{(0)}$

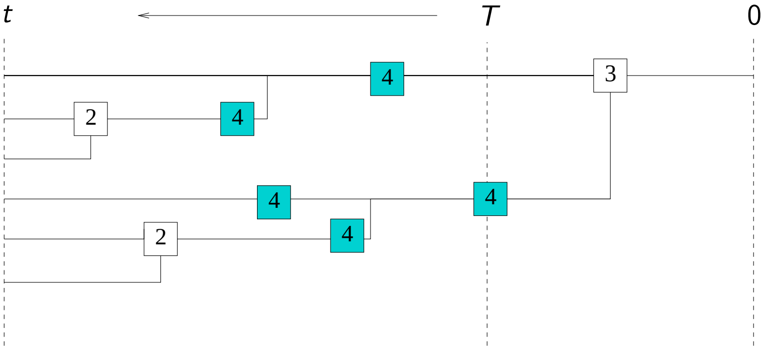


$$\omega_t^{(0)} = \mathcal{L}(\mathcal{X}) = \mathbb{E}\Big[\big(1 - f(\omega_0)\big)^{K_t}\Big]\,\omega_0\big(\,.\,\mid \widetilde{\mathcal{X}}_{i_*} = 1\big) + \Big(1 - \mathbb{E}\big[(1 - f(\omega_0))^{K_t}\big]\Big)\,\omega_0\big(\,.\,\mid \widetilde{\mathcal{X}}_{i_*} = 0\big)\Big]$$

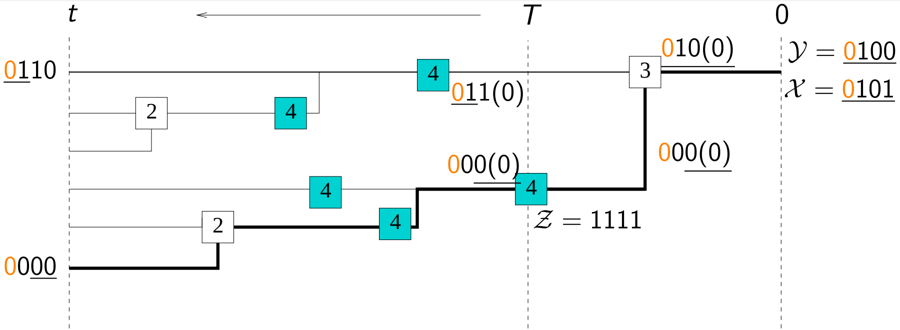
$$= \frac{\mathrm{e}^{stF}\omega_0}{\mathrm{e}^{st} f(\omega_0) + 1 - f(\omega_0)}$$

Truncated and collapsed ASRG

- $\text{ASRG}^{(k)}$: ASRG truncated at k ($\varrho^{(\ell)} = 0$ for $\ell > k$)
- $\text{cASRG}^{(k)}$: collapsed $\text{ASRG}^{(k)}$ ($\text{ASRG}^{(k-1)}$ decorated with i_k-squares at rate $\varrho^{(k)}$ on every line)
- $\rightsquigarrow$ $\text{ASRG}_t^{(k)} = \text{cASRG}_t^{(k)}$ with independent $\text{ASRG}^{(k)}$ attached to every i_k-square in a $\text{cASRG}_t^{(k)}$ for the remaining time

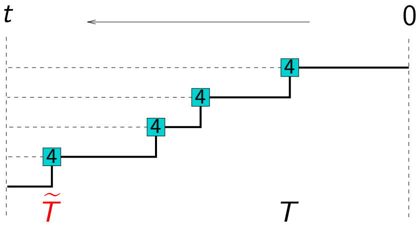


Induction step: $\omega_t^{(k-1)} \to \omega_t^{(k)}$



- $\mathcal{Y}$ type at root of $\text{ASRG}_t^{(k-1)}$ $(\sim \omega_t^{(k-1)})$
  $=$ type at root of $\text{cASRG}_t^{(k)}$ when ignoring the i_k-squares
- T time of rightmost i_k-square on line ancestral to $D^{(k)}$ in $\text{ASRG}_t^{(k-1)}$; if no such square, $T := \Delta$
- $\mathcal{Z}$ type at root of $\text{ASRG}_{t-T}^{(k)}$ $(\sim \omega_{t-T}^{(k)})$
- $\mathcal{X} = \mathbb{1}_{T=\Delta}\mathcal{Y} + \mathbb{1}_{T\neq\Delta}\big(\pi_{C^{(k)}}(\mathcal{Y}), \pi_{D^{(k)}}(\mathcal{Z})\big)$
  type at root of $\text{ASRG}_t^{(k)}$ $(\sim \omega_t^{(k)})$

Induction step: $\omega_t^{(k-1)} \to \omega_t^{(k)}$



$$\mathcal{X} = \mathbb{1}_{T=\Delta}\mathcal{Y} + \mathbb{1}_{T\neq\Delta}\big(\pi_{C^{(k)}}(\mathcal{Y}), \pi_{D^{(k)}}(\mathcal{Z})\big)$$

$$\rightsquigarrow \ldots \rightsquigarrow \mathcal{X} = \mathbb{1}_{\widetilde{T}=\Delta}\mathcal{Y} + \mathbb{1}_{\widetilde{T}\neq\Delta}\big(\pi_{C^{(k)}}(\mathcal{Y}), \pi_{D^{(k)}}(\widetilde{\mathcal{Z}})\big)$$

$\widetilde{\mathcal{Z}}$ type at root of ASRG$_{t-\widetilde{T}}^{(k-1)}$ $(\sim \omega_{t-\widetilde{T}}^{(k-1)})$

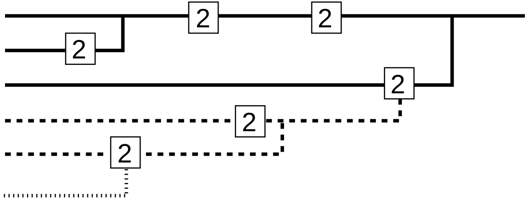
$$\omega_t^{(k)} = \mathcal{L}(\mathcal{X}) = \mathbb{E}\big[\mathcal{L}(\mathcal{X} \mid \widetilde{T})\big]$$

$$= \mathrm{e}^{-\varrho^{(k)}t}\omega_t^{(k-1)} + \pi_{C^{(k)}}.\omega_t^{(k-1)} \otimes \pi_{D^{(k)}}.\int_0^t \varrho^{(k)}\,\mathrm{e}^{-\varrho^{(k)}\tau}\omega_\tau^{(k-1)}\,\mathrm{d}\tau$$

(via symmetry of Poisson process)

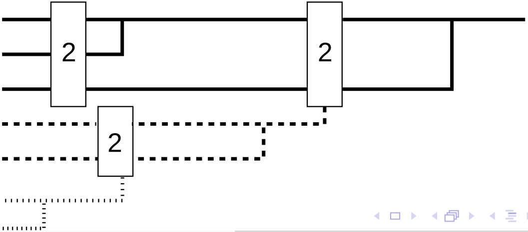
Synchronising the ASRG

- expand cASRG$^{(1)}$ along ancestral line

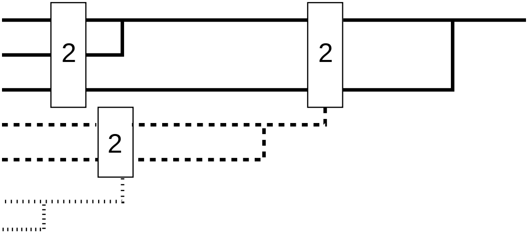


until new copy has no i_1-square on ancestral line

- only i_1-squares on the ancestral line of $D^{(1)}$ relevant
  $\rightsquigarrow$ synchronise i_1-squares across lines:



Synchronised ASRG



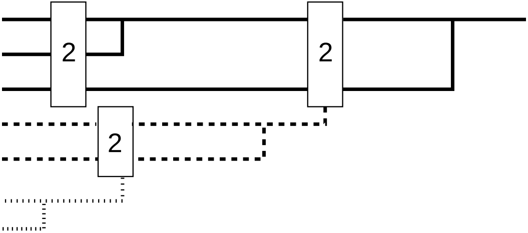
only i_1-squares on the ancestral line of $D^{(1)}$ relevant

$\rightsquigarrow$ identical distribution across lines matters,
but independence does not

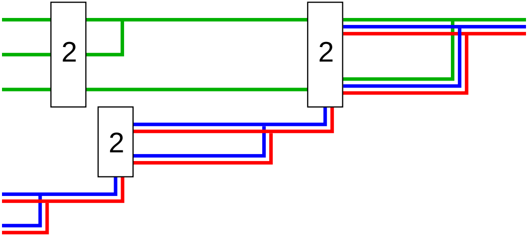
$\rightsquigarrow$ ASG decorated by i_1-bars laid down at rate $\varrho^{(1)}$

$\rightsquigarrow$ new tail attached to whichever 'head-line' is ancestral

Labelling and pruning

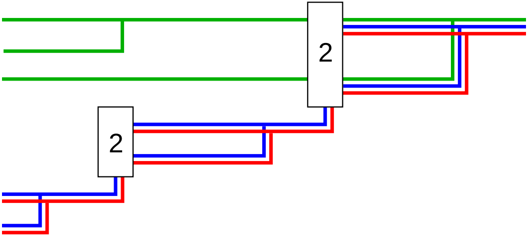


Labelling and pruning



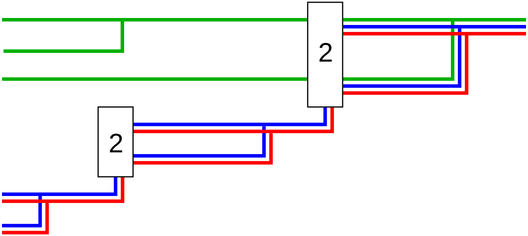
- label every line with the set A of sites to which it is potentially ancestral $\rightsquigarrow$ A splits into $A \cap C_{i_1}$ and $A \cap D_{i_1}$ at i_1-bar

Labelling and pruning

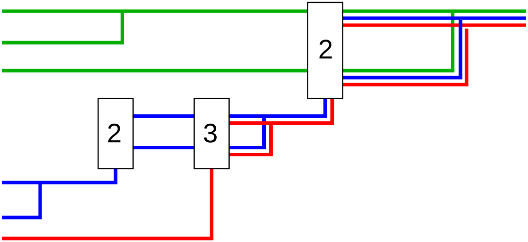


- label every line with the set A of sites to which it is potentially ancestral $\rightsquigarrow$ A splits into $A \cap C_{i_1}$ and $A \cap D_{i_1}$ at i_1-bar
- delete i_1-bar when $A \subseteq C_{i_1}$

Iteration

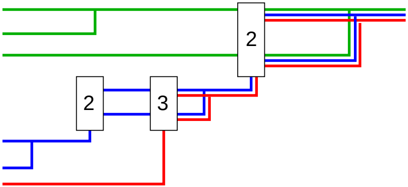


Iteration



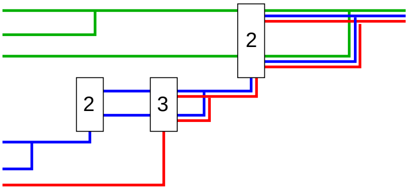
add i_2-bars at rate ϱ_{i_2} on every ASG, label, prune . . .
. . . and so forth

Dual process: Weighted partitioning process



- $(\Sigma_t, V_t)_{t \geqslant 0}$ weighted partitioning process
- $(\Sigma_t)_{t \geqslant 0}$ partitioning process ('ARG')
  on set of interval (or ordered) partitions of S
- V_t assigns an integer to every part of Σ_t
  (the number of lines in its ASG)

Dual process: Weighted partitioning process



if $(\Sigma_t, V_t) = (\mathcal{A}, v)$,

$$\mathcal{A} = \{A_1, \dots, A_m\},\ v = (v_A)_{A \in \mathcal{A}}:$$

- $\mathcal{A} \to \{A_1, \dots, A_j \cap C_i, A_j \cap D_i, \dots, A_m\}$,
  $v \to (v_{A_1}, \dots, v_{A_j}, 1, v_{A_{j+1}}, \dots, v_{A_m})$, at rate ϱ_i, $i \in S \setminus i_*$
- independently of Σ_t:
  - $v_A \to v_A + 1$ at rate $s v_A$ for every $A \in \mathcal{A}$ ('Yule')
  - $v_A \to 1$ at rate $\sum_{\ell \preccurlyeq \min(A)} \varrho_\ell$ for every $A \in \mathcal{A}$
    ('reset' at any split between A and i_*)

Duality

duality function:

$$H\big((\mathcal{A}, v), \omega\big) := \bigotimes_{A \in \mathcal{A}} \Big[\big(1 - (1 - f(\omega))^{v_A}\big) \omega^A(\,.\, | \, \widetilde{\mathcal{X}}_{i_*} = 0) + (1 - f(\omega))^{v_A} \omega^A(\,.\, | \, \widetilde{\mathcal{X}}_{i_*} = 1) \Big]$$

Theorem

$$\mathbb{E}_{\omega_0} \big[H\big((\mathcal{A}, v), \omega_t\big) \big] = \mathbb{E}_{(\mathcal{A},v)} \big[H\big((\Sigma_t, V_t), \omega_0\big) \big]$$

in particular,

$$\omega_t = \mathbb{E}_{(\{S\},1)} \big[H\big((\Sigma_t, V_t), \omega_0\big) \big]$$