Некоторые направления исследований лаборатории Автоматизации программирования ИПС РАН

(примеры и пояснения)

Андрей П. Немытых
Институт программных систем РАН
г. Переславль-Залесский

Онлайн-встреча – обсуждение технологий
анализа и верификации программных моделей вычислительных систем с экспертами
Санкт-Петербургского Филиала Российского Исследовательского Института (RRI)
и лаборатории Автоматизации программирования ИПС РАН
22 января 2024 г.

Краткий обзор направлений исследований:

- функциональные языки программирования;
- технология суперкомпиляции;
- доказательное (верифицируемое) программирование;
- верификация программных моделей вычислительных систем;
- задачи (не)выполнимости формул в формальных теориях (satisfiability modulo theories, SMT);
- .....

Сергей Давидович Мешвелиани
старший научный сотрудник ИПС РАН

- Разработка программ вычислительной алгебры, снабжаемых формальными машинно-проверяемыми доказательствами
  - (язык программирования Agda, некоторое подобие системы Coq).
- Опыт в создании автоматического доказывателя (прувера) для алгебры и для программ, основанного на аппарате переписывания термов.

Работы можно найти в Интернете по ключам:
«С.Д. Мешвелиани», «S.D. Meshveliani», «S.D. Mechveliani».

Может помочь советами.

Задачи невыполнимости формул в формальных теориях и бесконечные циклы

Пример №1. Формальная теория над множеством строк конечной длины (слов), описанная на языке SMT2-LIB:

```
(declare-fun i () String)
(declare-fun x () String)
(assert (not (str.contains x "A")))
(assert (= i (str.replace_all x "A" "B")))
(assert (not (= i x)))
```

– Длины строк конечные, но неограниченные.

Результат преобразований:
http://refal.botik.ru/mscp/test_web_eng/_test_smt1.ref
не содержит путей вычислений, приводящих к значению True.

Задачи невыполнимости формул в формальных теориях и бесконечные циклы - проверка теорий на пустоту

Пример №2. Формальная теория над множеством строк конечной длины (слов), описанная на языке SMT2-LIB:

```
(declare-fun x () String)
(assert (not (str.contains (str.replace_all x "bbb" "cb") "a")))
(assert (str.contains x "ca"))
```

– Длины строк конечные, но неограниченные.

Результат преобразований:

http://refal.botik.ru/mscp/test_web_eng/_test_smt3.ref

не содержит путей вычислений, приводящих к значению True.

Сравнение с широко известным зарубежными SMT-решателями Z3str3, CVC5

- Другие примеры можно найти на странице:
  - http://refal.botik.ru/mscp/MSCP-A_examples.html
- Автоматическая компиляция SMT2-LIB описаний формальных теорий в программные модели на функциональном языке переписывания термов.
- Преобразование (оптимизация) программных моделей инструментами суперкомпиляции.

Сравнение с широко известным зарубежными SMT-решателями Z3str3, CVC5

- Тестовый набор состоял из 47 контрольных тестов. Все эти тесты не поддаются анализу SMT-решателями Z3str3 и CVC5.
  - Автоматически были построены случайные строковые модели в языке SMT2-LIB.
  - Все 47 тестовых противоречивых моделей были построены последовательным выбрасыванием аксиом из этих случайных моделей.
- Анализ результатов тестирования:
  - http://refal.botik.ru/mscp/smt_bench/SMT-results.pdf

Системы уравнений в свободном конечнопорожденном моноиде

Определение. **Пусть даны конечный алфавит констант Σ и алфавит переменных $\mathcal{V}$. Уравнением в словах из Σ^* называется выражение вида:**

$$\Phi = \Psi, \text{ где } \Phi, \Psi \in \{\Sigma \cup \mathcal{V}\}^*.$$

Если $|\Phi\,\Psi|_\Sigma = 0$, то уравнение называется бескоэффициентным.

Примеры: **Пусть** $\mathrm{A} \in \Sigma, \quad \mathrm{x}, \mathrm{y}, \mathrm{z}, \mathrm{v} \in \mathcal{V}$.

$\mathrm{x}\,\mathrm{A} = \mathrm{A}\,\mathrm{x}$,
$\mathrm{x}\,\mathrm{y}\,\mathrm{z} = \mathrm{z}\,\mathrm{v}\,\mathrm{x}$ **– бескоэффициентное уравнение Хмелевского.**

О постановке задачи решения уравнений в словах

- **Решить уравнение в словах.**
  - **Что это означает, если множество решений уравнения бесконечно?**
  - **Вопрос о структуре этого множества решений нетривиален.**
- **Алгоритм Маканина перечисляет множество решений, если оно не пусто.**
  - **См.: Г.С. Маканин. Проблема разрешимости уравнений в свободной полугруппе, 1977,** https://www.mathnet.ru/rus/sm2805

Почему задача решения уравнений в словах сложная?

Пусть алфавит $\Sigma = \{\mathrm{A}\}$. Уравнения в словах

– $\Phi = \Psi$ **кодируют диофантовые уравнения:** $|\Phi| = |\Psi|$.

– $\mathrm{x}\,\Phi = \Psi$ **кодируют диофантовые неравенства:** $|\Phi| \leq |\Psi|$.

Простые алгоритмы для некоторых классов уравнений

– **Уравнения с постоянными правыми частями:**

  – $\Phi = \mathrm{C_0}$, **где** $\mathrm{C_0} \in \Sigma^*$ **константа.**

– **Уравнения с одной неизвестной.**

– **Квадратичные уравнения.**

Определение. **Уравнение в словах** $\Phi = \Psi$ **называется квадратичным, если каждая переменная входит в уравнение не более двух раз: для любой** $\mathrm{x} \in \mathcal{V}$ $|\Phi\,\Psi|_{\mathrm{x}} \leq 2$.

Простейший пример

Квадратичное уравнение с одной переменной. x **— переменная,** A **— коэффициент.**

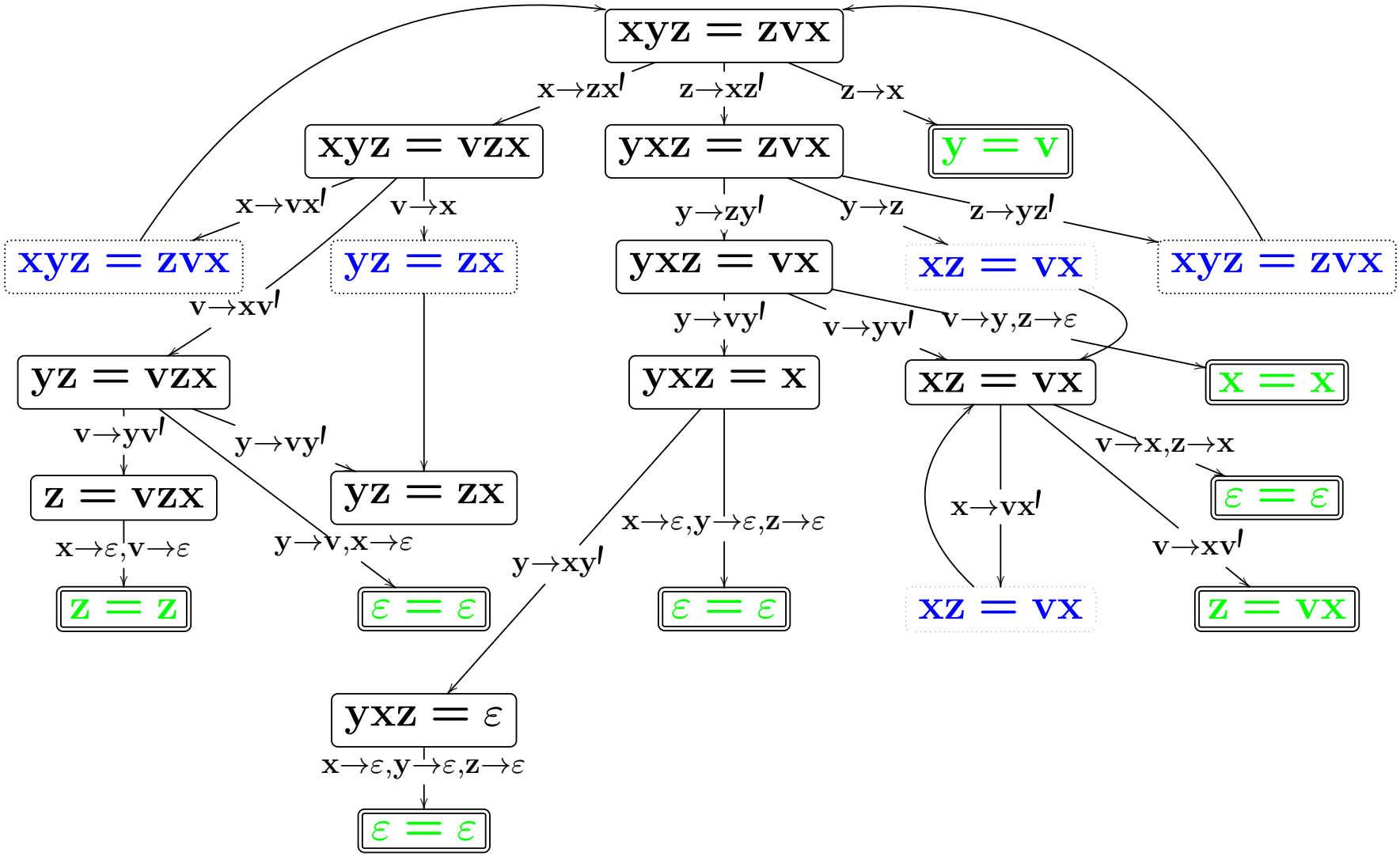
$$\mathrm{x}\,\mathrm{A} = \mathrm{A}\,\mathrm{x}$$

Множество решений: A^***.**

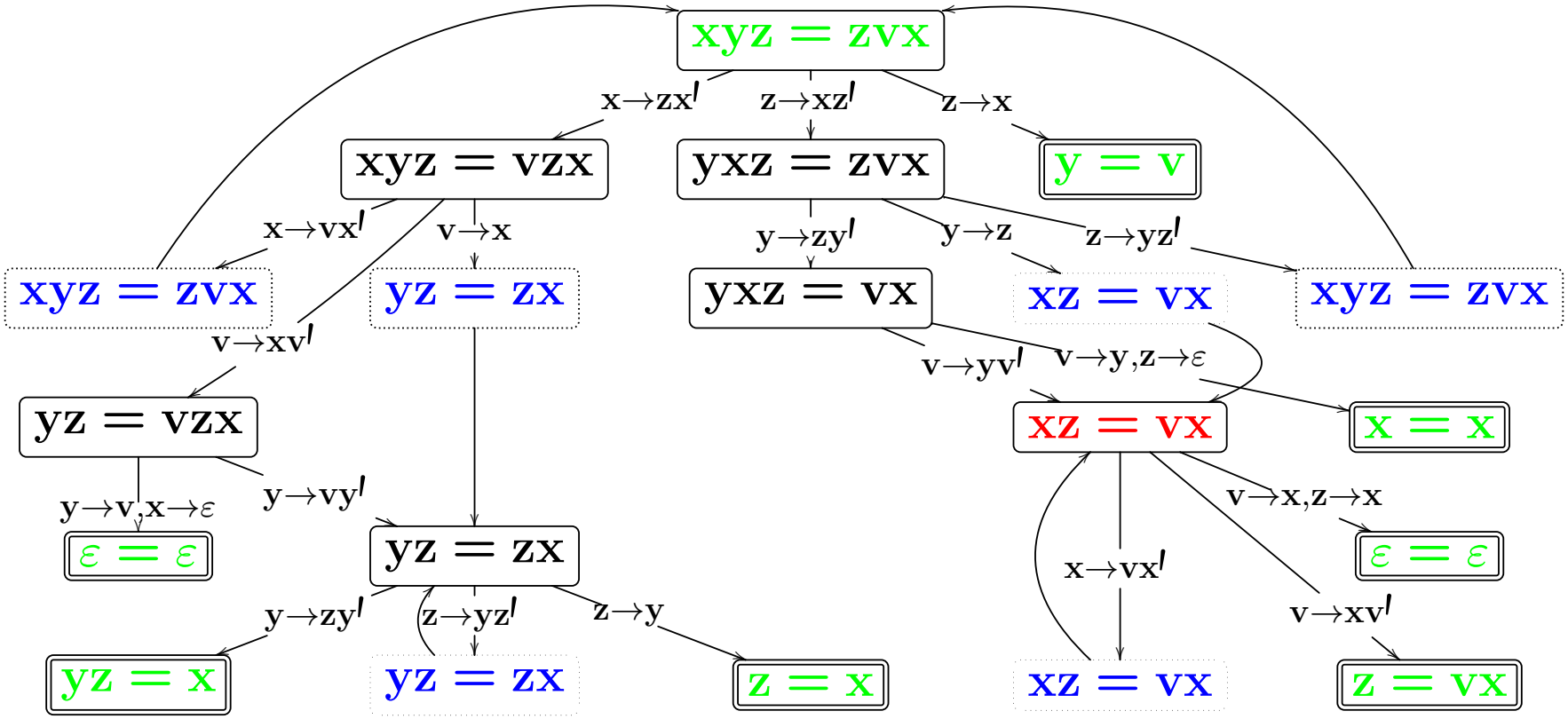
Квадратичные уравнения

Пример решения квадратичного уравнения.
Уравнение Хмелевского, $\mathrm{x}, \mathrm{y}, \mathrm{z}, \mathrm{v}$ **— переменные.**

$$\mathrm{x}\,\mathrm{y}\,\mathrm{z} = \mathrm{z}\,\mathrm{v}\,\mathrm{x}$$



Здесь любая $\mathrm{u} \in \mathcal{V}$, вхождение которой в сужение на ребре E имеет вид u', пробегает (под)множество Σ^+. Во входной вершине ребра E штрих над переменными не ставится.



$z \to (sr)^n srs,\ y \to rsp,\ v \to psr,\ x \to (sr)^n s,\ v \to sr,\ z \to rs$;

Множество решений уравнения сопряжения $xz = vx$:

$x = (sr)^n s,\ z = rs,\ v = sr$, **где** $\mathbb{N} \ni n \geq 0,\ r, s$ **– словарные параметры.**

Пусть дан алфавит $\mathcal{A}$.

Определим множество $\mathcal{P}$ параметрических слов: (1) $\mathcal{A}^* \subset \mathcal{P}$;
(2) если $\phi \in \mathcal{P}$, n – натуральный параметр, тогда $\phi^{\mathrm{n}} \in \mathcal{P}$;
(3) если $\phi_1, \phi_2 \in \mathcal{P}$, тогда $\phi_1\phi_2 \in \mathcal{P}$.

Нас будут интересовать параметрические слова в алфавите $\Sigma \cup \mathcal{V}$.
Термы $\mathrm{p} \in \mathcal{V}$ будем называть словарными параметрами.

Пример:
Множество решений уравнения сопряжения $\mathrm{x\,z} = \mathrm{v\,x}$
параметризуемо: $\mathrm{x} = (\mathrm{s\,r})^{\mathrm{n}}\mathrm{s},\ \mathrm{z} = \mathrm{r\,s},\ \mathrm{v} = \mathrm{s\,r}$,
где $\mathbb{N} \ni \mathrm{n} \geq 0,\ \mathrm{r}, \mathrm{s}$ – параметры, принимающие значения в Σ^*.

Две теоремы Ю.И. Хмелевского

Теорема 1. (Хмелевский)
Множество решений любого бескоэффициентного уравнения с тремя переменными параметризуемо.

Теорема 2. (Хмелевский)
Множество решений бескоэффициентного уравнения $\mathrm{x\,y\,z = z\,v\,x}$ в свободном моноиде с не менее чем двумя образующими непараметризуемо.

Язык решений уравнений в словах $\mathcal{W}$ vs. язык регулярных выражений $\mathcal{R}$

Языки $\mathcal{W}$, $\mathcal{R}$ пересекаются, но не являются подмножествами друг друга:

– $\mathcal{W} \cap \mathcal{R} \neq \emptyset$

– $\mathcal{R} \not\subseteq \mathcal{W}$

– $\mathcal{W} \not\subseteq \mathcal{R}$

Оба формальных языка $\mathcal{W}$, $\mathcal{R}$ являются «предельными» для описания алгоритмически разрешимых множеств.

Специализация интерпретаторов как оптимизация абстрактной интерпретации

Пример №3. **Формальная спецификация протокола когерентности кэша MOESI (недетерминированная система переписывания):**

$$
\begin{array}{ll}
(\mathbf{rh}) & \mathbf{modified} + \mathbf{owned} + \mathbf{shared} + \mathbf{exclusive} \geq 1 \to . \\
(\mathbf{rm}) & \mathbf{invalid} \geq 1 \to \\
& \quad \mathbf{invalid}' = \mathbf{invalid} - 1, \mathbf{exclusive}' = 0, \mathbf{modified}' = 0, \\
& \quad \mathbf{shared}' = \mathbf{shared} + \mathbf{exclusive} + 1, \mathbf{owned}' = \mathbf{owned} + \mathbf{modified}. \\
(\mathbf{wh1}) & \mathbf{modified} \geq 1 \to . \\
(\mathbf{wh2}) & \mathbf{exclusive} \geq 1 \to \mathbf{exclusive}' = \mathbf{exclusive} - 1, \mathbf{modified}' = \mathbf{modified} + 1. \\
(\mathbf{wh3}) & \mathbf{shared} + \mathbf{owned} \geq 1 \to \\
& \quad \mathbf{shared}' = 0, \mathbf{exclusive}' = 1, \mathbf{modified}' = 0, \mathbf{owned}' = 0, \\
& \quad \mathbf{invalid}' = \mathbf{invalid} + \mathbf{modified} + \mathbf{exclusive} + \mathbf{shared} + \mathbf{owned} - 1. \\
(\mathbf{wm}) & \mathbf{invalid} \geq 1 \to \\
& \quad \mathbf{shared}' = 0, \mathbf{exclusive}' = 1, \mathbf{modified}' = 0, \mathbf{owned}' = 0, \\
& \quad \mathbf{invalid}' = \mathbf{invalid} + \mathbf{modified} + \mathbf{exclusive} + \mathbf{shared} + \mathbf{owned} - 1.
\end{array}
$$

Специализация интерпретаторов как оптимизация абстрактной интерпретации

Протокол когерентности кэша MOESI

Параметризованное начальное состояние:

$$\mathbf{invalid} \geq 1, \mathbf{exclusive} = 0, \mathbf{shared} = 0, \mathbf{modified} = 0, \mathbf{owned} = 0$$

Потенциально ненадёжные состояния:

(1) $\mathbf{exclusive} + \mathbf{shared} + \mathbf{owned} \geq 1, \mathbf{modified} \geq 1$
(2) $\mathbf{exclusive} \geq 1, \mathbf{shared} + \mathbf{owned} \geq 1$
(3) $\mathbf{modified} \geq 2$
(4) $\mathbf{exclusive} \geq 2$

– Длины путей эволюции конечные, но неограниченные.

Результат преобразований (анализа) см. на странице:
http://refal.botik.ru/protocols/#MOESI

Абстрактный интерпретатор f

Дано: вычислительная система S и её свойство надёжности $\psi(\cdot)$.

▼ $\mathrm{f}(\mathrm{n}, \tilde{\mathrm{x}})$ вычисляет n-ое состояние системы S, σ есть некоторое состояние системы S.

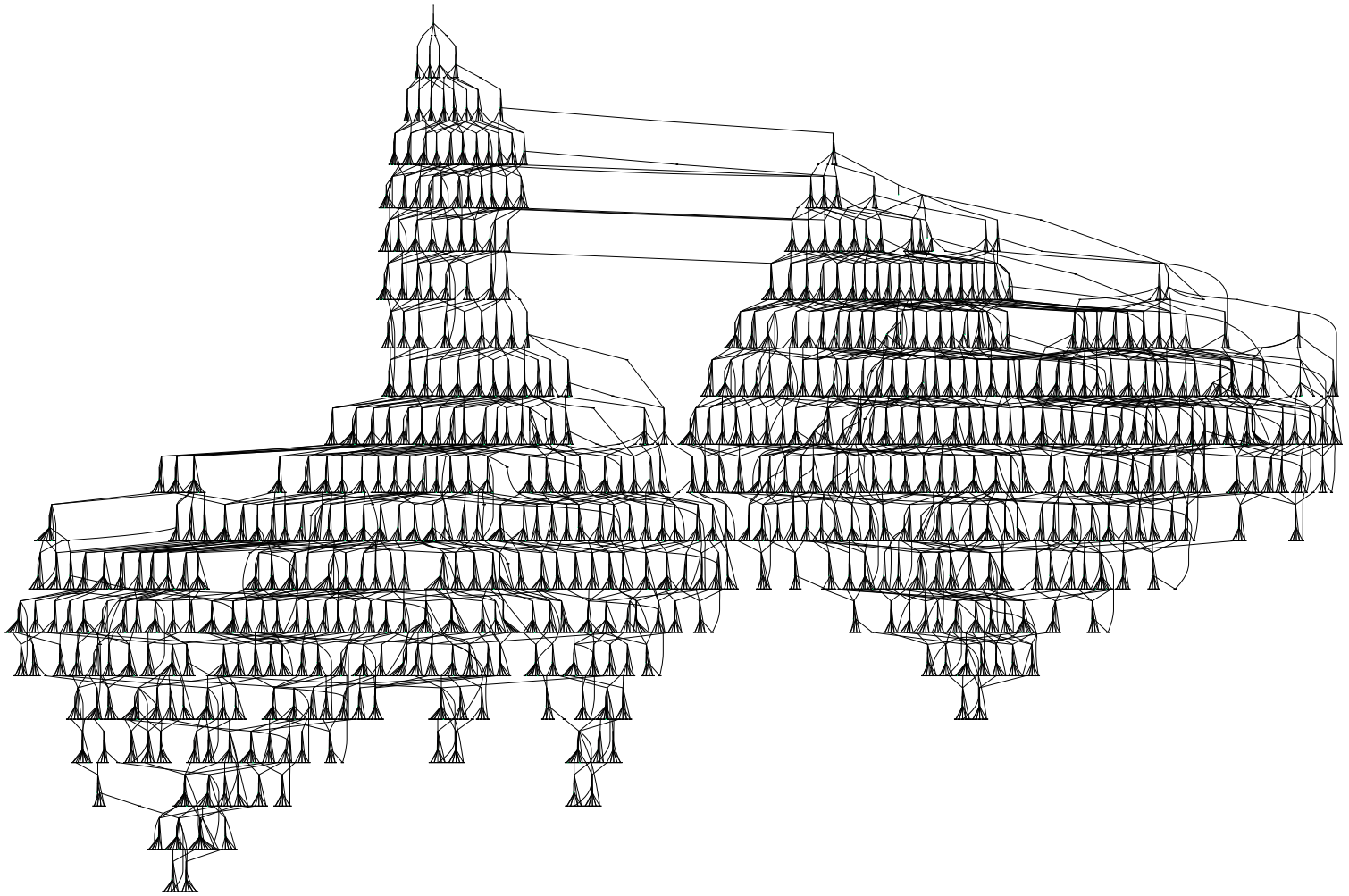
$$\begin{cases} \mathrm{p}(\sigma) \text{ возвращает } \mathrm{False} & \text{, если свойство } \psi(\sigma) \text{ неверно;} \\ \mathrm{p}(\sigma) = \mathrm{True} & \text{иначе.} \end{cases}$$

▲ Пусть $\mathrm{p}(\cdot)$ и $\mathrm{f}(\cdot,\cdot)$ не циклятся.
Тогда если $\forall \mathrm{n} \,.\, (\mathbb{N} \ni \mathrm{n} > 0)\ \forall \tilde{\mathrm{x}} \in \mathrm{Init} \Longrightarrow \mathrm{p}(\mathrm{f}(\mathrm{n}, \tilde{\mathrm{x}}))$ не возвращает False, то S надёжна, если она стартует из множества начальных состояний Init.

– «Время» n дискретное - конечная, но неограниченная последовательность действий недетерминированной системы S.

Структура доказательства посредством суперкомпилятора SCP4 свойства надёжности (safety property) протокола

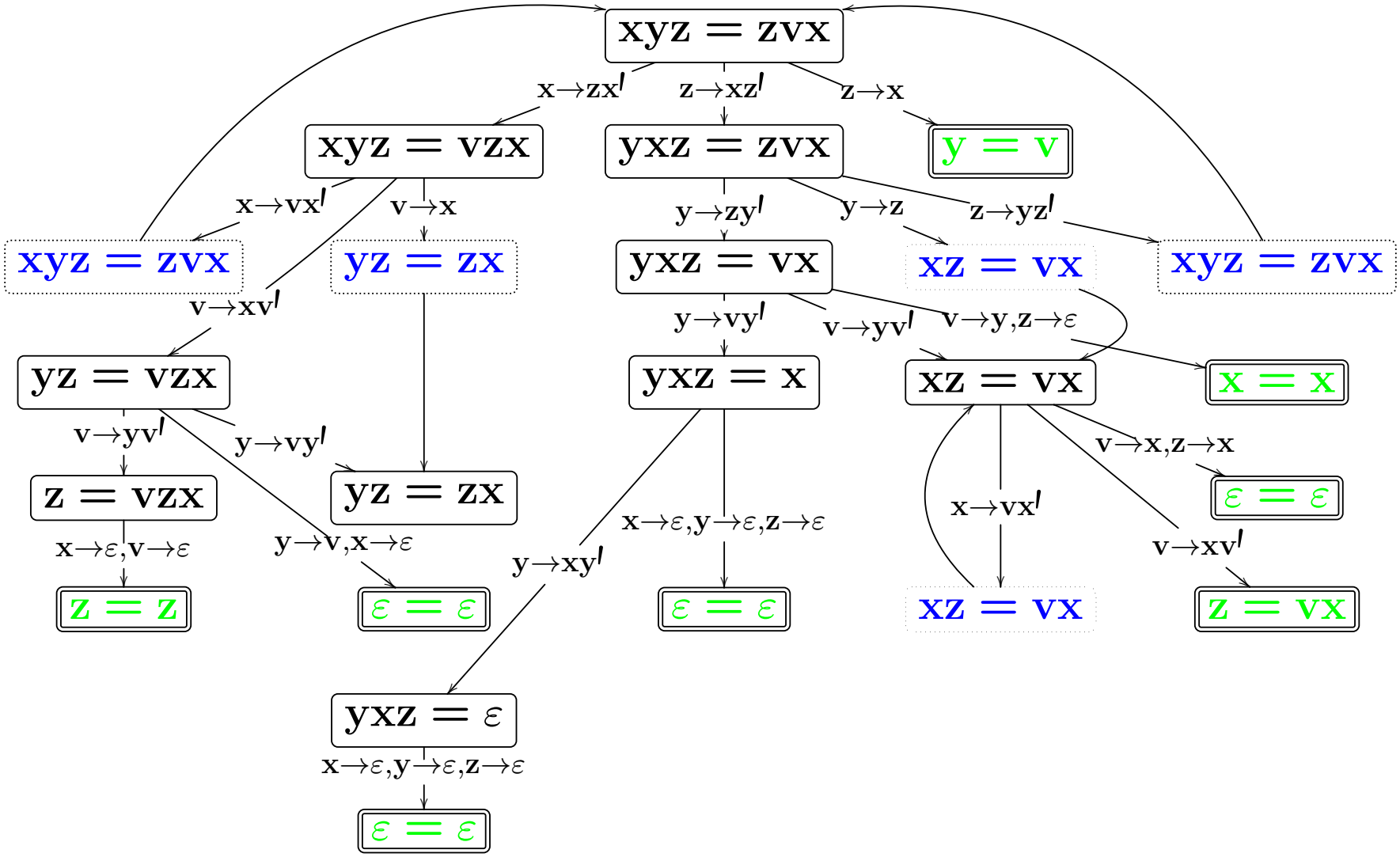
Two Consumers - Two Producers protocol / Abstract Multithreaded Java Program



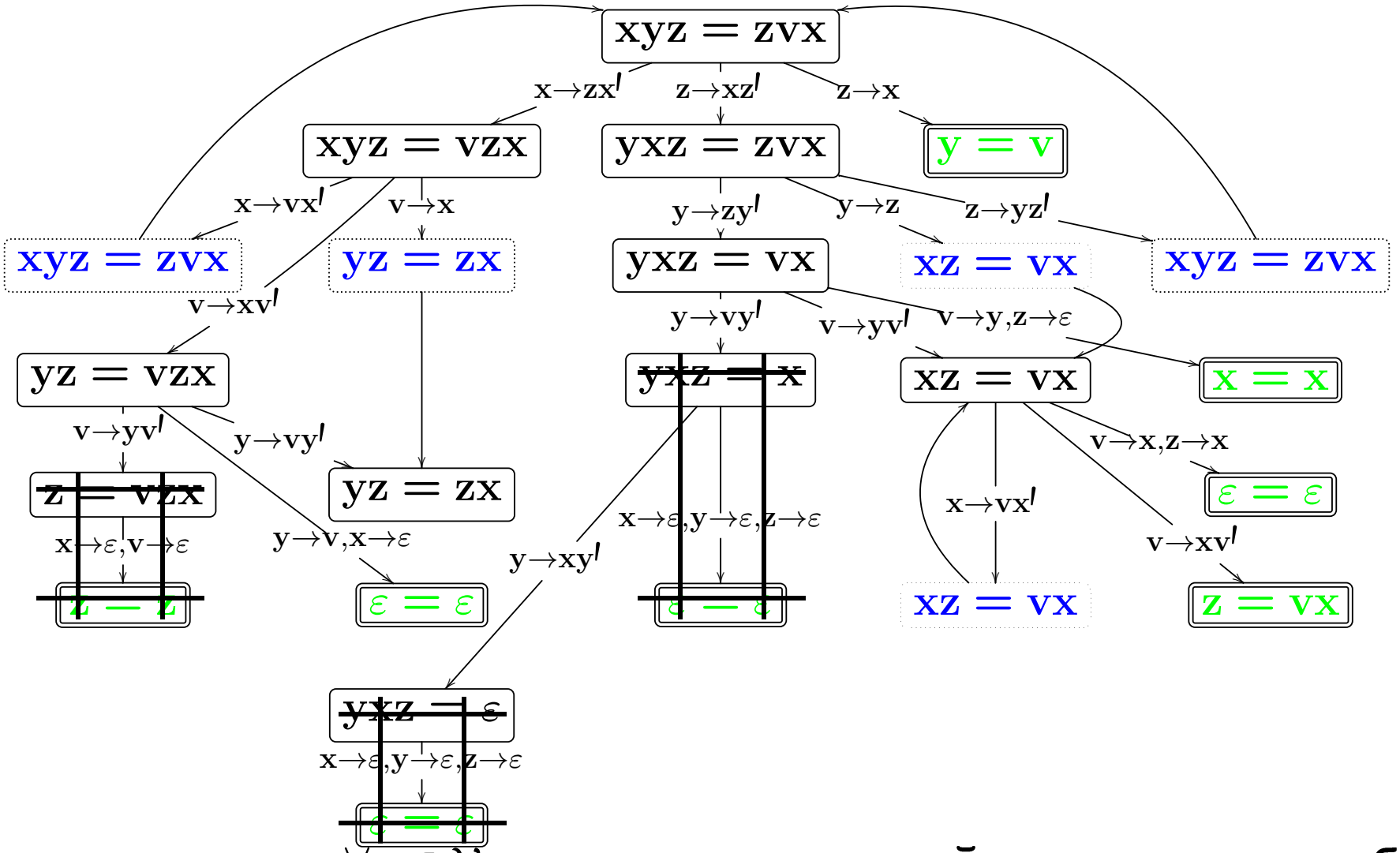
Подробности см. на странице: http://refal.botik.ru/protocols/#2P_2C

**См. продолжение данной презентации,
подготовленное и доложенное
Антониной Н. Непейвода:**

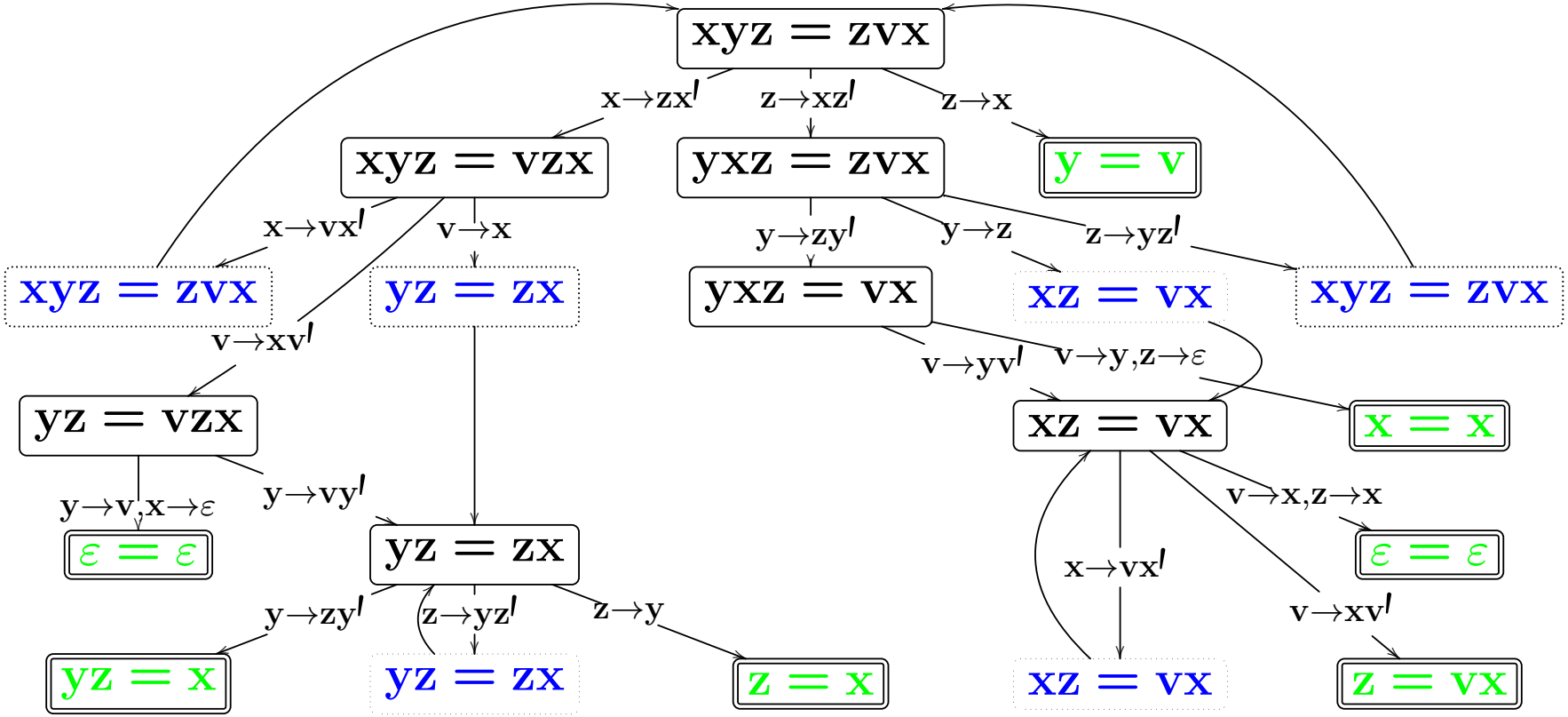
http://refal.botik.ru/events/Antonia-Nepeivoda_online-meeting_Russian-Research-Institute_and_AP-Lab_ISP-RAS_part-2.pdf



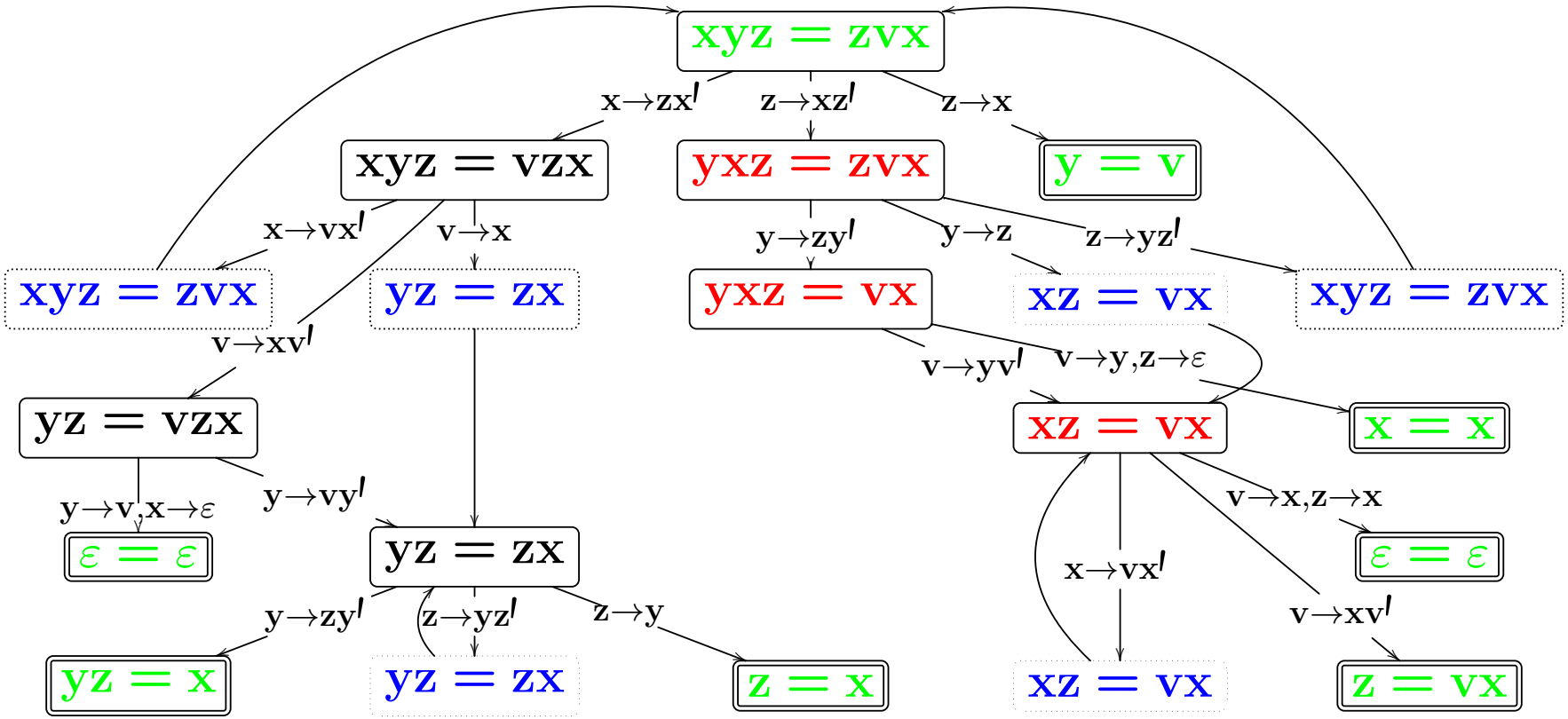
Здесь любая $\mathrm{u} \in \mathcal{V}$, вхождение которой в сужение на ребре E имеет вид u', пробегает (под)множество Σ^+. Во входной вершине ребра E штрих над переменными не ставится.



Здесь значение u_0 $\forall u \in \mathcal{V}$, вхождение которой в сужение на ребре E имеет вид u', такое, что $|u_0| > 0$. В вершинах E штрих не ставится.



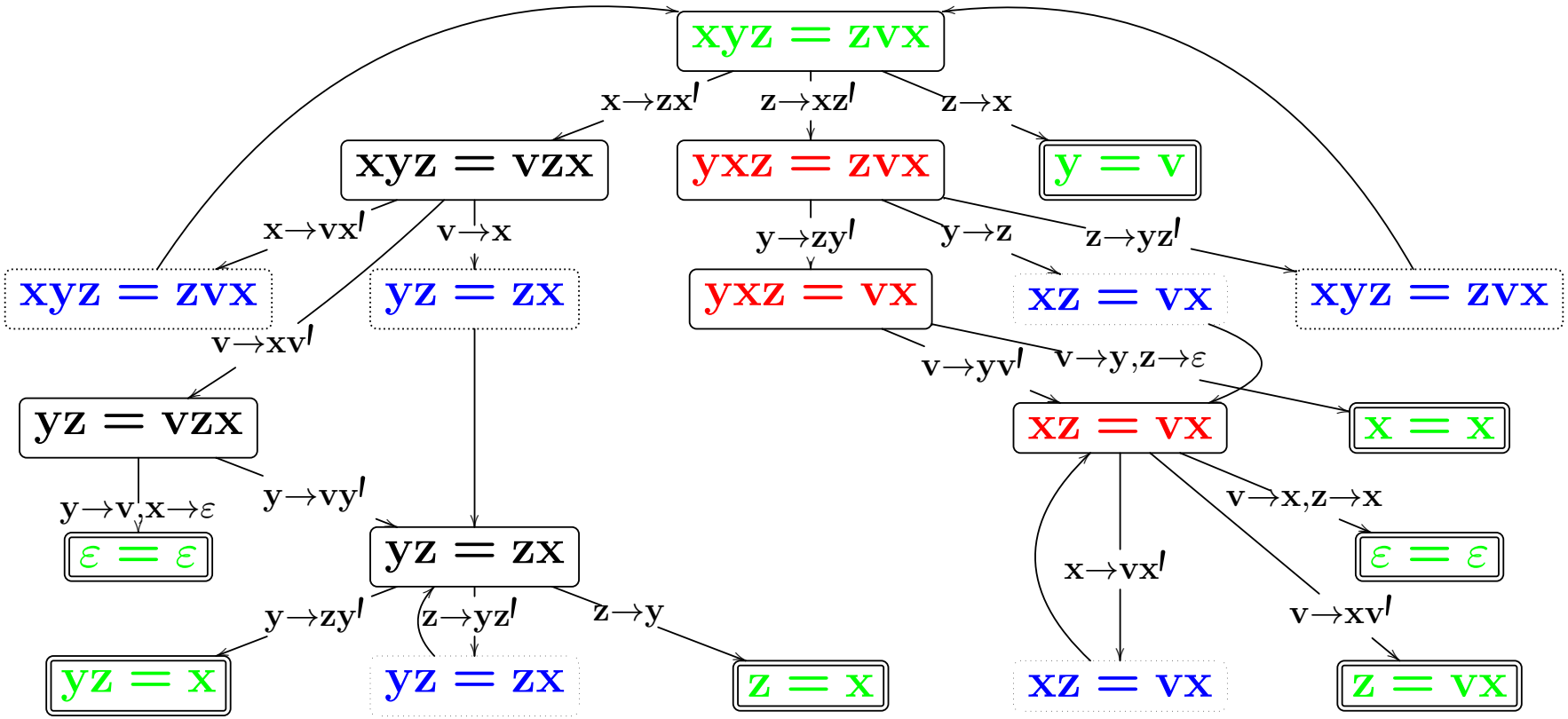
Здесь значение $\mathrm{u_0}$ $\forall \mathrm{u} \in \mathcal{V}$, вхождение которой в сужение на ребре E имеет вид u', такое, что $|\mathrm{u_0}| > 0$. В вершинах E штрих не ставится.



$z \to xz^{\prime}$, $y \to zy^{\prime}$, $v \to yv^{\prime}$, $(x \to vx^{\prime})^{n}$, $v \to xv^{\prime}$, $z \to vx$, **где** $n \geq 0$.

$z \to xz^{\prime}$, $y \to zp$, $v \to pv^{\prime}$, $x \to v^{n}s$, $v \to sr$, $z \to rs$, **где** p, r, s **параметры.**

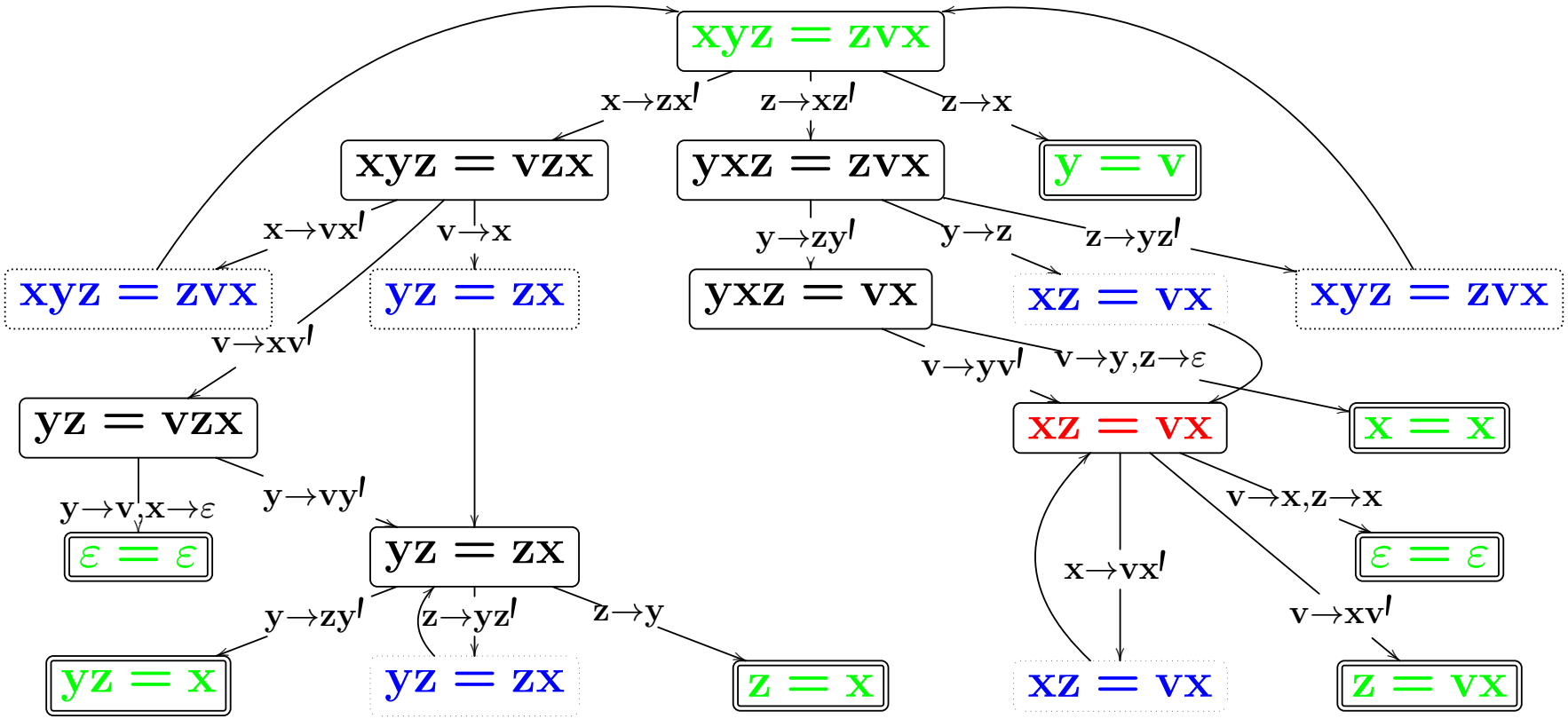
$z \to (sr)^{n}srs$, $y \to rsp$, $v \to psr$, $x \to (sr)^{n}s$, $v \to sr$, $z \to rs$;



$\mathbf{z \to (sr)^n srs,\ y \to rsp,\ v \to psr,\ x \to (sr)^n s,\ v \to sr,\ z \to rs}$;

Пример семейства решений: $\mathbf{x = (sr)^n s,\ y = rsp,\ z = (sr)^n srs,\ v = psr,}$

где $\mathbb{N} \ni \mathbf{n \geq 0},\ \mathbf{p, r, s}$ **– параметры, принимающие значения в** Σ^*.



$\mathbf{z \to (sr)^n srs,\ y \to rsp,\ v \to psr,\ x \to (sr)^n s,\ v \to sr,\ z \to rs}$;

Множество решений уравнения сопряжения $\mathbf{xz = vx}$ **:**

$\mathbf{x = (sr)^n s,\ z = rs,\ v = sr}$, **где** $\mathbb{N} \ni \mathbf{n \geq 0},\ \mathbf{r, s}$ **– словарные параметры.**

Определение. **Пусть $\mathrm{z}, \mathrm{v} \in \mathcal{V}^*$. Слово v называется сопряженным слову z, если существует $\mathrm{x} \in \mathcal{V}^*$ такое, что $\mathrm{x}\,\mathrm{z} = \mathrm{v}\,\mathrm{x}$.**

Если моноид M одновременно является группой относительно операции в M, тогда отношение сопряженности в моноиде M есть групповое отношение сопряженности:

$$\mathrm{x}\,\mathrm{z} = \mathrm{v}\,\mathrm{x} \Leftrightarrow \mathrm{v} = \mathrm{x}\,\mathrm{z}\,\mathrm{x}^{-1}.$$

Множество решений уравнения сопряжения $\mathrm{x\,z = v\,x}$:
$\mathrm{x = (sr)^n s},\ \mathrm{z = rs},\ \mathrm{v = sr}$, **где** $\mathbb{N} \ni \mathrm{n} \geq 0,\ \mathrm{r, s}$ **– словарные параметры.**

Следствие 1: **Слово** v **сопряжено слову** z **тогда и только тогда, когда существует циклическая перестановка** σ **букв в слове** z **такая, что** $\sigma(\mathrm{z}) = \mathrm{v}$.

Следствие 2: **Отношение сопряженности в свободном моноиде является отношением эквивалентности.**

Пусть дан алфавит $\mathcal{A}$. Определим множество $\mathcal{P}$ параметрических слов: (1) $\mathcal{A}^* \subset \mathcal{P}$;
(2) если $\phi \in \mathcal{P}$, $\mathbf{n}$ – натуральный параметр, тогда $\phi^{\mathbf{n}} \in \mathcal{P}$;
(3) если $\phi_1, \phi_2 \in \mathcal{P}$, тогда $\phi_1\phi_2 \in \mathcal{P}$.

Нас будут интересовать параметрические слова в алфавите $\Sigma \cup \mathcal{V}$.
Термы $\mathbf{p} \in \mathcal{V}$ будем называть словарными параметрами.

Пример параметрического слова:
$(\mathbf{s}\,\mathbf{A}\,\mathbf{r})^{\mathbf{n}}\,\mathbf{s})^{\mathbf{m}}\mathbf{s}^{\mathbf{n}}$,
где $\mathbb{N} \ni \mathbf{n}, \mathbf{m} \geq \mathbf{0},\ \mathbf{r}, \mathbf{s}$ – параметры, принимающие значения в Σ^*.

Пусть дано уравнение в словах $\mathcal{E}$ от n переменных $\mathrm{x_1}, \ldots, \mathrm{x_n}$.

Определение. **n-ка парам. слов $(\phi_1, \ldots, \phi_\mathrm{n})$ называется параметрическим подмножеством решений уравнения $\mathcal{E}$, если при любой подстановке σ конкретных значений параметров n-ка**

$$(\sigma(\phi_1), \ldots, \sigma(\phi_\mathrm{n}))$$

есть решение уравнения $\mathcal{E}$.

Определение. **Множество решений уравнения $\mathcal{E}$ параметризуемо, если существует конечное множество $\mathcal{M}$ параметрических подмножеств решений уравнения $\mathcal{E}$ такое, что для каждого решения $(\overline{\mathrm{x_1}}, \ldots, \overline{\mathrm{x_n}})$ уравнения $\mathcal{E}$ $\exists \mathrm{M} \in \mathcal{M}$ и подстановка σ конкретных значений параметров в M такая, что**

$$\sigma(\mathrm{M}) = (\overline{\mathrm{x_1}}, \ldots, \overline{\mathrm{x_n}}).$$

Пусть дан алфавит $\mathcal{A}$.

Определим множество $\mathcal{P}$ параметрических слов: (1) $\mathcal{A}^* \subset \mathcal{P}$;
(2) если $\phi \in \mathcal{P}$, n – натуральный параметр, тогда $\phi^{\mathrm{n}} \in \mathcal{P}$;
(3) если $\phi_1, \phi_2 \in \mathcal{P}$, тогда $\phi_1\phi_2 \in \mathcal{P}$.

Нас будут интересовать параметрические слова в алфавите $\Sigma \cup \mathcal{V}$.
Термы $\mathrm{p} \in \mathcal{V}$ будем называть словарными параметрами.

Пример:
Множество решений уравнения сопряжения $\mathrm{x\,z} = \mathrm{v\,x}$
параметризуемо: $\mathrm{x} = (\mathrm{s\,r})^{\mathrm{n}}\mathrm{s},\ \mathrm{z} = \mathrm{r\,s},\ \mathrm{v} = \mathrm{s\,r}$,
где $\mathbb{N} \ni \mathrm{n} \geq 0,\ \mathrm{r}, \mathrm{s}$ – параметры, принимающие значения в Σ^*.

Лемма ℵ. Для всякого бескоэффициентного уравнения в словах $\Phi = \Psi$ в свободном моноиде с не менее чем двумя образующими и любого его параметрического решения ξ результат подстановки σ натуральным параметрам любых конкретных значений из $\mathbb{N}_0$ – $\sigma(\xi)$ – является решением уравнения $\Phi = \Psi$.

Доказательство:

Почти очевидно. □

Две теоремы Ю.И. Хмелевского

Теорема 1. (Хмелевский)
Множество решений любого бескоэффициентного уравнения с тремя переменными параметризуемо.

Теорема 2. (Хмелевский)
Множество решений бескоэффициентного уравнения $\mathrm{x\,y\,z = z\,v\,x}$ в свободном моноиде с не менее чем двумя образующими непараметризуемо.

1: С.Д. Мешвелиани. О машинном доказательстве для арифметики дробей над кольцом с НОД. 2020
https://elibrary.ru/item.asp?id=42339763

2: С.Д. Мешвелиани. Доказательная программа для способа Карацубы умножения многочленов. 2022
https://elibrary.ru/item.asp?id=47489107

3: С.Д. Мешвелиани. Допустимый порядок на мономах вполне задан. Конструктивное доказательство. 2023
https://journals.rcsi.science/0132-3474/article/view/137636

4: С.Д. Мешвелиани. DoCon-A-3.2-rc3. Библиотека доказательных программ компьютерной алгебры. Предварительный выпуск. 2023.
http://www.botik.ru/pub/local/Mechveliani/docon-A/docon-A-3.2-rc3.zip

5: А.П. Немытых. О суперкомпиляции (к 80-тилетию со дня рождения В. Ф. Турчина). 2011
http://conf.nsc.ru/files/conferences/Lyap-100/fulltext/69293/69928/nemytykh_supercom
Lyapunov100.pdf

6: А.П. Немытых. О некоторых понятиях суперкомпиляции – метода специализации программ. 2015
http://refal.botik.ru/library/refal2015_issue-II.pdf

7: Антонина Н. Непейвода. Модельный суперкомпилятор MSCP-A.
http://refal.botik.ru/mscp/mscp-a_eng.html

8: Антонина Н. Непейвода. Примеры суперкомпиляции.
http://refal.botik.ru/mscp/MSCP-A_examples.html

9: A.P. Lisitsa, A.P. Nemytykh. Verification as a Parameterized Testing (Experiments with the SCP4 Supercompiler). 2007
https://link.springer.com/article/10.1134/S0361768807010033

10: A.P. Lisitsa, A.P. Nemytykh. A Note on Specialization of Interpreters. 2007
https://link.springer.com/chapter/10.1007/978-3-540-74510-5_25

11: SCP 4 : Verification of Protocols. http://refal.botik.ru/protocols/

12: A.P. Lisitsa, A.P. Nemytykh. Finite Countermodel Based Verification for Program Transformation (A Case Study). 2015 https://arxiv.org/pdf/1512.03859

13: A.P. Lisitsa, A.P. Nemytykh. Verifying Programs via Intermediate Interpretation. 2017
http://refal.botik.ru/vpt/vpt2017/VPT2017-Lisitsa_Nemytykh_presentation.pdf

14: A. Ahmed, A.P. Lisitsa, A.P. Nemytykh. Cryptographic Protocol Verification via Supercompilation (A Case Study). 2013
http://refal.botik.ru/vpt/Ahmed_Lisitsa_Nemytykh_VPT-2013_talk.pdf