# Structure Theory of Petri Nets

Eike Best[1] and Harro Wimmel[2]

[1] Parallel Systems, Department of Computing Science
Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany
eike.best@informatik.uni-oldenburg.de
[2] Theory of Programming Languages and Programming,
Institut für Informatik
Universität Rostock, Albert-Einstein-Straße 22, D-18059 Rostock, Germany
harro.wimmel@uni-rostock.de

**Abstract.** The aim of this tutorial is to give a concise, but nonetheless not too narrow, overview of definitions and results pertaining centrally to Petri net structure theory. The Petri net model considered in these notes are the classical place/transition nets, as they have been defined in the First Advanced Course held in 1979 and originate back to the late Sixties. Structure theory asks what behavioural properties of a Petri net can be derived from its structural properties. Other aspects of Petri nets are neglected to a large extent in the present notes, such as various extensions and generalisations of central notions and results, as well as almost all algorithmic and complexity-theoretic consequences that accompany the structure-theoretic results. Because full proofs can easily be retrieved from the literature, they are not given, unless they are small and perhaps somewhat characteristic for Petri net oriented reasoning. Proof ideas are often sketched, however, and the sharpness of various results is accentuated by means of examples and counterexamples. A list for further reading is also provided.

## 1 First Steps in Petri Nets

We all may remember a lecture in Theoretical Computer Science in which Finite Automata were introduced. Finite Automata may be used to represent regular languages. Other classes of languages were also introduced and analysed. This was done for good reasons. For instance, compiler construction is grounded on a variety of language types.

However, we may also view formal languages from a more system-oriented perspective. If we interpret every letter as an atomic activity, then the words of a language describe the sequences of actions that are feasible. For instance, the evolutions permitted in some industrial production process could be described in this way. The atomic actions could perhaps be ascribed to the activities of various machines involved in the process. This idea can be exploited both for the simulation and for the validation of a production process in its planning stage, before it is actually implemented. Much money can be saved if design errors are detected and corrected in this way, well before the physical realisation

K. Jensen et al. (Eds.): ToPNoC VII, LNCS 7480, pp. 162–224, 2013.

of a production process. However, it must be realised that in general, and in particular in such processes, more than just one activity can be executed in parallel. Such parallelism (or concurrency) cannot be described either by a formal language or by a finite automaton, at least not in the form in which they are usually taught in a beginners' course. Problems concerning parallelism cannot be handled just by using words built as sequences of the letters of an alphabet.

In the beginning of the Sixties of the last century, Carl Adam Petri became aware of this lack of descriptive power of the traditional finite automata model, and of its bias towards sequential rather than concurrent execution. His dissertation, which he completed in the year 1962, was fundamentally concerned with redressing this balance. The idea behind Petri nets (as they were called some years later) is to modify some of the concepts behind finite automata. Most fundamentally, in his view, states are thought to be structured and may consist of smaller parts (called local states). Transitions may affect certain local states but may leave other local states unchanged or unaffected. Local states are represented in Petri nets by means of *places* while state transitions are again simply called *transitions*. It is this principle of locality, together with the duality between states and transitions, which underlies the very definition of a Petri net.

## 1.1 Basic Definitions

We shall use standard, though not always unique, mathematical notation. For instance, $f\colon X \to Y$ and $f \in Y^X$ both denote the fact that $f$ is a function from $X$ to $Y$.

**Definition 1.** *Petri net*
A Petri net is a triple $(S, T, F)$ consisting of

- a countable set $S$ of *places* and a countable set $T$ of *transitions* with $S \cap T = \emptyset$,
- and a mapping $F\colon (S \times T) \cup (T \times S) \to \mathbb{N}$ which defines *arcs* (also called *arrows*, or *edges*) between places and transitions. $F(s,t)$ defines the number of arcs from $s$ to $t$. Analogously, $F(t,s)$ defines the number of arcs from $t$ to $s$. ■

In the following, we will almost exclusively consider *finite* Petri nets, that is, Petri nets in which both the set of places and the set of transitions are finite. For such nets, we often use finite sets of indices as follows: $S = \{s_1, \ldots, s_{|S|}\}$ and $T = \{t_1, \ldots, t_{|T|}\}$. Sometimes transitions are simply denoted $\{a, b, c, \ldots\}$. This is to be understood such that $a$ is $t_1$, $b$ is $t_2$, etc.

In the graphical representation of a Petri net, we draw every place as a circle and every transition as a box (normally square, and in general, rectangular). Furthermore, we draw exactly $F(s_i, t_j)$ arcs from the $i$th place $s_i$ to the $j$th transition $t_j$, and $F(t_j, s_i)$ arcs from the $j$th transition $t_j$ to the $i$th place $s_i$.

Places and transitions are enumerated for reasons of convenience, but in general, any naming is allowed. Enumerations are helpful for an alternative representation of the arcs in the calculus of matrices. In this representation, we replace the mapping $F$ by two $|S| \times |T|$-matrices $\mathbb{F}, \mathbb{B} \in \mathbb{N}^{S \times T}$. The value $\mathbb{B}_{i,j}$ in the $i$th row and $j$th column is, by definition, the number of arcs from $s_i$ to $t_j$, and the value $\mathbb{F}_{i,j}$ is, by definition, the number of arcs from $t_j$ to $s_i$. We will also occasionally write $\mathbb{F}(s)$ or $\mathbb{F}(s,.)$ for the '$s$th row' in $\mathbb{F}$, given $s \in S$, $\mathbb{F}(t)$ or $\mathbb{F}(.,t)$ for the '$t$th column', given $t \in T$, and $\mathbb{F}(s,t)$ for the entry in row $s$ and column $t$ of $\mathbb{F}$. $\mathbb{B}$ and $\mathbb{F}$ are called *backward matrix* and *forward matrix*, respectively. This is to be understood from the point of view of transitions: arcs emanating from a transition (i.e. 'forward arcs', as seen from this transition) are described by the forward matrix. Looking backward from a transition, one encounters its incoming arcs which are described in the backward matrix. As we will see later, these matrices allow linear algebra to be applied.

We introduce a number of elementary concepts.

**Definition 2.** *Preset, postset, and related notions*
Let $(S,T,F)$ be a Petri net. For $x \in S \cup T$, we call ${}^\bullet x = \{y \in S \cup T \mid F(y,x) \geq 1\}$ and $x^\bullet = \{y \in S \cup T \mid F(x,y) \geq 1\}$ the *preset* (*postset*, respectively) of $x$. Generalising this, we define ${}^\bullet X = \bigcup_{x \in X} {}^\bullet x$ and $X^\bullet = \bigcup_{x \in X} x^\bullet$, for $X \subseteq S \cup T$. An element $x \in S \cup T$ satisfying ${}^\bullet x \cup x^\bullet = \emptyset$ is called *isolated*. If there are arcs in both directions between a place $s$ and a transition $t$, i.e. if $F(s,t) \geq 1 \leq F(t,s)$, then this situation is called a *loop* or a *self-loop*. A loop is called *simple* if $F(s,t) = 1 = F(t,s)$. A net is *pure* if there are no self-loops, and *plain* if there are no multiple arcs, that is, if the function $F$ returns 0 or 1, but no number greater than 1. ■

**Definition 3.** *States and markings*
Let $N = (S,T,F)$ be a Petri net. The *state set* of $N$ is defined to be $\mathbb{N}^S$, that is, the set of all functions from $S$ to $\mathbb{N}$. This is to be understood as the set of all *potential* states of $N$, of which the actually possible states – to be defined later – are a subset. A function $M \colon S \to \mathbb{N}$ is called a *state*, or a *marking*, of $N$. If $M(s_i) = m$ then we say that 'place $s_i$ carries $m$ tokens (in the state $M$)'. We often write states as column vectors, indexed by places. ■

Graphically, we represent tokens as solid dots within a place.

Using the matrix representation of arcs, the Petri net shown in Figure 1 can be described by the quadruple $(S,T,\mathbb{B},\mathbb{F})$ and the state $M$ with

$$S = \{s_1, s_2, s_3\}, \quad T = \{t_1, t_2, t_3\}, \quad \mathbb{B} = \begin{pmatrix} t_1 & t_2 & t_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{matrix} \\ s_1 \\ s_2 \\ s_3 \end{matrix}, \quad \mathbb{F} = \begin{pmatrix} t_1 & t_2 & t_3 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 0 \end{pmatrix} \begin{matrix} \\ s_1 \\ s_2 \\ s_3 \end{matrix}, \quad M = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \begin{matrix} s_1 \\ s_2 \\ s_3 \end{matrix}$$

We have ${}^\bullet s_1 = \{t_1, t_3\}$, $s_1{}^\bullet = \{t_1\}$, ${}^\bullet s_2 = \emptyset$, $s_2{}^\bullet = \{t_2\}$, ${}^\bullet s_3 = \{t_1, t_2\}$, $s_3{}^\bullet = \{t_3\}$, ${}^\bullet t_1 = \{s_1\}$, $t_1{}^\bullet = \{s_1, s_3\}$, ${}^\bullet t_2 = \{s_2\}$, $t_2{}^\bullet = \{s_3\}$, ${}^\bullet t_3 = \{s_3\}$, $t_3{}^\bullet = \{s_1\}$. The function $F$ can also be written down, but this would be rather circumstantial, compared with the graphical representation.

This Petri net has a loop between $s_1$ and $t_1$, as well as a *multiple arc* leading from $t_2$ to $s_3$; in other words, we have $F(t_2, s_3) > 1$. A multiple arc will often be

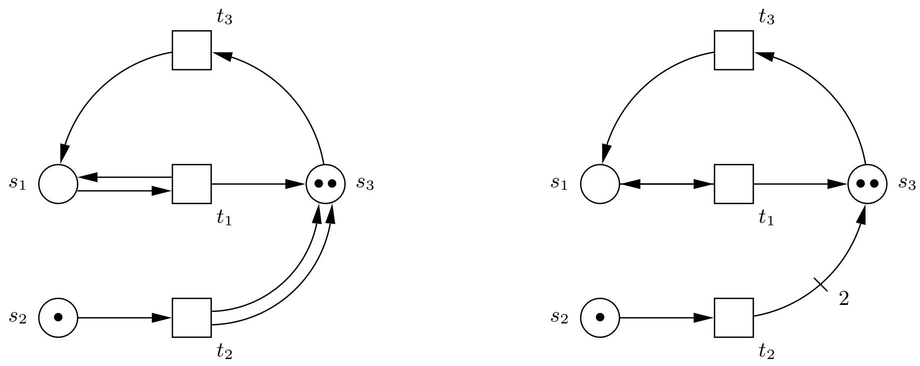



**Fig. 1.** Two representations of the same Petri net

represented as a plain arc together with some natural number which is inscribed at it. This number denotes the multiplicity of the arc. If the multiplicity is 1, we simply omit the number. A loop may also be represented by a simple arc with two arrow heads leading in both directions. (However, this representation can be problematic in small figures where loops could easily be mistaken for plain arcs.)

**Definition 4.** *Petri net with an initial marking*
An *initial Petri net* $N$ is defined to be a tuple $N = (S, T, \mathbb{B}, \mathbb{F}, M_0)$, or $N = (S, T, F, M_0)$, where

- $(S, T, \mathbb{B}, \mathbb{F})$ (respectively, $(S, T, F)$) is a Petri net;
- $M_0 \in \mathbb{N}^S$ is an *initial state*.

We also denote a Petri net $N = (S, T, F)$ with initial state $M_0$ by $(N, M_0)$. The initial state is also called *initial marking* or *starting state*. ■

In the following, we will call an initial Petri net simply a Petri net, if it is clear from the context that we are talking about an initially marked net. Also, we often use the word *system* in an informal way when talking about an initially marked net. Typically, the starting state is included in the graphical representation of a net by means of tokens (solid dots) on the places. In Figure 1, for example, $M = (0 \;\; 1 \;\; 2)^{\mathsf{T}}$ is the starting state. The notation $^{\mathsf{T}}$ means 'transposed' and shows, in this case, that the marking is viewed as a column vector.

Sometimes it is desirable to ignore certain parts of a Petri net, for instance a set of transitions and/or places, together with all arcs connected to them. This leads to the notion of a *subnet*.

**Definition 5.** *Subnet*
Let $N = (S, T, F, M_0)$ be a Petri net and let $S' \subseteq S$ as well as $T' \subseteq T$. The *subnet induced by* $S'$ *and* $T'$ is denoted by $N(S', T')$ and defined by

$$N(S', T') = (S', T', F|_{(S' \times T') \cup (T' \times S')}, M_0|_{S'}).$$

Generally, the vertical bar denotes the restriction of the domain of a set or function to the set given in its index. In this particular case, we have $F|_{(S' \times T') \cup (T' \times S')} = F \cap (((S' \times T') \cup (T' \times S')) \times \mathbb{N})$. ■

The idea is that in $N(S', T')$, the part of the net defined by $S'$ and $T'$ is left intact, while all elements of $S \setminus S'$ and $T \setminus T'$ are neglected. The definition of $F'$ means that all arcs between elements of $S'$ and $T'$ (including their multiplicities) are just inherited from $N$ to $N(S', T')$. All other arcs, that is, arcs with at least one endpoint in $S \setminus S'$ or in $T \setminus T'$, are ignored. The definition of $M_0|_{S'}$ means that places in $S'$ carry exactly as many tokens in $N(S', T')$ as they do in $N$. Places in $S \setminus S'$ and all tokens on them are ignored.

In Figure 1, the subnet induced by $S' = \{s_1\}$ and $T' = \{t_1, t_2, t_3\}$ consists of one place, $s_1$, with zero tokens, three arcs (two of which form a loop), and three transitions (one of which, viz. $t_2$, is isolated).

### 1.2 Firing Transitions

We now describe the dynamics, i.e. the behaviour, of initially marked nets. This will be defined in analogy to the successive execution of state transitions in a finite automaton. In Petri nets, this process is called *firing* or *executing* transitions.

**Definition 6.** *The transition rule of Petri nets*
Let $N = (S, T, \mathbb{B}, \mathbb{F})$ be a Petri net, let $M \in \mathbb{N}^S$ be a state of $N$, and let $t \in T$ be a transition of $N$. We call $t$ *$M$-activated* (or *enabled*, *firable*, *executable* in state $M$), if $M \geq \mathbb{B}(t)$ (that is, $\forall s \in S\colon M(s) \geq \mathbb{B}_{s,t} = F(s, t)$).
A transition $t$ *fires in state $M$ to state $M'$* (or *is executed* in state $M$, leading to state $M'$, or simply *leads from $M$* to $M'$) if:

- $M \geq \mathbb{B}(t)$ (that is, $t$ is activated in $M$), and
- $M' = M - \mathbb{B}(t) + \mathbb{F}(t)$.

This rule is called the *transition (or firing) rule*, and it is the basic behavioural (state change) rule for Petri nets. Formally, the fact that $t$ is firable in $M$ and leads from $M$ to $M'$ is denoted by $M\,[t\rangle M'$ or by $M \xrightarrow{t} M'$. ■

Informally, when a transition fires, it consumes tokens from every place of its preset (whence there needs to be at least one token on every such place just prior to firing) and produces tokens on every place of its postset. The number of tokens consumed and produced are calculated according to the multiplicity of arcs around the transition. More precisely, if a place $s$ has an outgoing arc of multiplicity $k$ towards $t$, then every single firing of $t$ needs at least $k$ tokens on $s$ and consumes exactly $k$ tokens from $s$. Similarly, if $t$ is connected to a place $s'$ of its postset by an arc of multiplicity $k$, then every single firing of $t$ produces exactly $k$ tokens on $s'$, which are added to the already existing ones. In the case of self-loops, tokens are first taken from a place and later reproduced. That is, if $F(s'', t) = k > 0$ and $F(t, s'') = m$, $k$ tokens on $s''$ are necessary for firing $t$. When firing $t$, we might think of it as the $k$ tokens being removed from $s''$ first, and then, in a second step, $m$ tokens being added to $s''$ again. In the special case of a simple loop ($k = m = 1$), the effect of firing is that the number of tokens on such a place is neither decreased nor increased, because the single token that is taken away by firing, is put back by the same firing.

As an example, let us reconsider the Petri net from Figure 1. We have, on the one hand, that

$$(0 \;\; 1 \;\; 2)^\mathsf{T} \, [t_2\rangle (0 \;\; 0 \;\; 4)^\mathsf{T} \quad \text{and} \quad (0 \;\; 1 \;\; 2)^\mathsf{T} \, [t_3\rangle (1 \;\; 1 \;\; 1)^\mathsf{T}.$$

On the other hand, firing $t_1$ in state $(0 \;\; 1 \;\; 2)^\mathsf{T}$ is not possible, since there is no token on $s_1$. We may also fire from other states. For instance, we have $(7 \;\; 3 \;\; 5)^\mathsf{T} \, [t_1\rangle (7 \;\; 3 \;\; 6)^\mathsf{T}$ and $(2 \;\; 1 \;\; 0)^\mathsf{T} \, [t_2\rangle (2 \;\; 0 \;\; 2)^\mathsf{T}$.

Of course, it is usually possible to execute a sequence of transitions, one after another, instead of just one of them. This naturally leads to two interesting questions:

– How can the set of states that are reachable from the initial state through such sequences be characterised?
– How can the set of firable sequences be characterised?

As we will see, these two questions are of a rather different nature. Moreover, the answers to both of them are non-trivial.

**Definition 7.** *Firing sequence*
Let $N = (S, T, F, M_0)$ be a Petri net. We define inductively: $\forall t \in T, \;\; \sigma \in T^*$:

$$\begin{aligned} M \, [\varepsilon\rangle M' \;\; &\text{if} \;\; M = M' \\ M \, [\sigma t\rangle M' \;\; &\text{if} \;\; \exists M'' \in \mathbb{N}^S : \;\; M \, [\sigma\rangle M'' \, [t\rangle M', \end{aligned}$$

where $M \, [\sigma\rangle M'' \, [t\rangle M'$ is a shorthand notation for $M \, [\sigma\rangle M'' \wedge M'' \, [t\rangle M'$ and $\varepsilon$ is the empty sequence.
We read $M \, [\sigma\rangle M'$ as '$\sigma$ fires from $M$ to $M'$', or '$\sigma$ is executed from $M$ and leads to $M'$', or, more simply, '$\sigma$ leads from $M$ to $M'$'. Here also, $M \xrightarrow{\sigma} M'$ is synonymous to $M \, [\sigma\rangle M'$.

$$M \, [\sigma\rangle \quad :\Longleftrightarrow \quad \exists M' \in \mathbb{N}^S : \;\; M \, [\sigma\rangle M'.$$

A sequence $\sigma \in T^*$ is called a *firing sequence* or an *execution (sequence)* from $M$ (or *executable/firable at* $M$), denoted by $M \, [\sigma\rangle$, if there is some marking $M'$ with $M \, [\sigma\rangle M'$.
Further, we call $\mathcal{E}(M) = \{M' \mid \exists \sigma \in T^*\colon M \, [\sigma\rangle M'\}$ the *reachability set* (or the *state space*) of $M$, and $\mathcal{E}(N) := \mathcal{E}(M_0)$ is the reachability set of $N$. Alternatively, we also write $[M\rangle$ instead of $\mathcal{E}(M)$. ■

In Figure 1, we have the following firing sequences: $\sigma_1 = t_3 t_3 t_1 t_3 t_2 t_3 t_3$, or $\sigma_2 = t_3 t_1 t_1 t_1 t_1 t_3$, amongst others. More precisely, we have $(0 \;\; 1 \;\; 2)^\mathsf{T} \xrightarrow{\sigma_1} (5 \;\; 0 \;\; 0)^\mathsf{T}$ and $(0 \;\; 1 \;\; 2)^\mathsf{T} \xrightarrow{\sigma_2} (2 \;\; 1 \;\; 4)^\mathsf{T}$. By contrast, $\sigma_3 = t_3 t_3 t_1 t_3 t_3 t_2 t_3$ is not firable from $M_0 = (0 \;\; 1 \;\; 2)^\mathsf{T}$, since the fourth instance of $t_3$ cannot be executed.

The reachability set, $\mathcal{E}(M_0)$, of this example is as follows:

$$\{(0\ \ 1\ \ 2)^{\mathsf{T}}, (0\ \ 0\ \ 4)^{\mathsf{T}}\} \cup \{(i\ \ j\ \ k)^{\mathsf{T}} \mid i \geq 1 \wedge j \leq 1 \wedge i + 2j + k \geq 4\}. \qquad (1)$$

The reachability set does not have to be representable so smoothly, even if the given Petri net is small.

Firing enjoys several basic properties that one should know about. The state reached after firing a transition depends only on the previous state, rather than on the (possibly large) history by which this state has been reached; this property is called *memorylessness* or *local determinacy* of firing. If a transition can be fired in some state, then it can also be fired in every 'larger' state (where larger means that no place contains less tokens and at least one place more tokens); this property is called the *monotonicity* of firing. If two transitions can be fired in arbitrary order, then the resulting marking after firing both does not depend on their ordering; this is called the *commutativity* of firing. More formally, we have:

**Lemma 1.** *Properties of firing*
Transition firing is

- *locally determined*, i.e. $\forall t \in T\ \forall M, M', M'' \in \mathbb{N}^S$: $(M[t\rangle M' \wedge M[t\rangle M'' \Rightarrow M' = M'')$,
- *monotonic*, i.e. $\forall M, M', M'' \in \mathbb{N}^S\ \forall t \in T$: $(M\,[t\rangle M' \Longrightarrow (M + M'')\,[t\rangle\,(M' + M''))$,
- and *commutative*, i.e. if $M, M_1, M_2, M_3, M_4$ are states and $t, t'$ are transitions with $M\,[t\rangle M_1\,[t'\rangle M_2$ and $M\,[t'\rangle M_3\,[t\rangle M_4$, then $M_2 = M_4$.

*Proof:* Local determinacy: We have $M' = M - \mathbb{B}(t) + \mathbb{F}(t) = M''$.
Monotonicity: $M \geq \mathbb{B}(t)$ entails $M + M'' \geq \mathbb{B}(t) + M'' \geq \mathbb{B}(t)$ and $M' + M'' = M - \mathbb{B}(t) + \mathbb{F}(t) + M'' = (M + M'') - \mathbb{B}(t) + \mathbb{F}(t)$.
Commutativity: $M_2 = M_1 - \mathbb{B}(t') + \mathbb{F}(t') = M - \mathbb{B}(t) + \mathbb{F}(t) - \mathbb{B}(t') + \mathbb{F}(t') = M - \mathbb{B}(t') + \mathbb{F}(t') - \mathbb{B}(t) + \mathbb{F}(t) = M_3 - \mathbb{B}(t) + \mathbb{F}(t) = M_4$. ■

It is important to note that commutativity does not mean that $t't$ is firable whenever $tt'$ is. Thus, the above concept of commutativity is not exactly the same as that used frequently in mathematics. Mathematical commutativity is not normally satisfied in Petri net firings. More precisely, the following properties are *not* normally satisfied:

- *persistency*, i.e. $\forall t, t' \in T\ \forall M \in \mathbb{N}^S$: $(t \neq t' \wedge M\,[t\rangle \wedge M\,[t'\rangle \Longrightarrow M\,[tt'\rangle)$.
- *confluence*, i.e. $\forall M, M', M'' \in \mathbb{N}^S\ \forall \sigma, \sigma' \in T^*$: $(M\,[\sigma\rangle M' \wedge M\,[\sigma'\rangle M'' \Longrightarrow \exists \widehat{M} \in \mathbb{N}^S\ \exists \sigma'', \sigma''' \in T^*$: $(M'\,[\sigma''\rangle \widehat{M} \wedge M''\,[\sigma'''\rangle \widehat{M}))$.

Both properties can be disproved by the following simple Petri net:

$$S = \{s_1, s_2, s_3\}, T = \{t_1, t_2\}, F = \{(s_1, t_1), (s_1, t_2), (t_1, s_2), (t_2, s_3)\} \text{ and } M(s_1) = 1, M(s_2) = M(s_3) = 0. \qquad (2)$$

Persistency means that an activated transition cannot be de-activated by *other* transitions. In general, however, it is possible for some transition to de-activate another one; such a situation is called a *conflict*. In (2), for example, $t_2$ is activated in marking $M$. However, firing $t_1$ de-activates $t_2$ (as well as $t_1$ itself). In section 4, persistency and the absence of conflicts will be investigated more closely.

Confluence means that executions drifting apart in a net can be brought together again. In general, however, this may not be the case. In (2), the markings $M_1$ and $M_2$ reached after firing $t_1$ and $t_2$ from $M$, respectively, have no common successor marking. For example, if $t_2$ denoted an erroneous exection, then there would be no way of 'correcting' the error by reaching a state that could also have been reached after $t_1$.

### 1.3 Graphs and Multigraphs

The graphical representation of a Petri net indicates that it can be understood as a *graph* in the mathematical sense. In this section, we recall some notions pertaining to (multi-)graphs in general.

A *graph* is a structure $(X, E)$ where $E \subseteq X \times X$. An element of $X$ is called a *vertex*, or a *node*. An element $e = (x, y) \in E$ is called an *arc*, or an *edge*, or sometimes also an *arrow*, leading from $x$ to $y$. An *arc-labelled* graph is a structure $(X, L, E)$ where $E \subseteq X \times L \times X$. For $e = (x, \ell, y) \in E$, $\ell$ is also called the *label* (or the *inscription*) of the arc $e$ from $x$ to $y$. A *multigraph* is a structure $(X, \mathbb{E})$ where $\mathbb{E}$ is a multiset of pairs from $X \times X$. Thus, we may have several arrows in parallel from one node to another one. Likewise, a *labelled multigraph* is a structure $(X, L, \mathbb{E})$ where $\mathbb{E}$ is a multiset of triples from $X \times L \times X$.

A *subgraph* of a (multi-)graph with vertex set $X$ is a graph with vertex set $X' \subseteq X$ and (labelled) arcs restricted to those between nodes in $X'$. By a (directed) *path* we mean a directed sequence of edges, such that the endpoint of one edge is the beginning of the next one. A path is a *cycle* if its starting vertex equals its end vertex. The *length* of a path is the number of edges in it. A special case is just a single node without any edge (called a path of length 0). A path is called *simple* or *elementary* if no vertex appears twice in it, except possibly the very beginning and the very end, in which case it is called a *simple cycle* or an *elementary cycle*.

A (multi-)graph $G$ is called *strongly connected* if for any two nodes $x$ and $y$, there is a directed path from $x$ to $y$. $G$ is called *weakly connected* if for any two nodes $x, y$, there is some sequence of arrows (not necessarily pointing in the same direction) from $x$ to $y$. $G$ is called *covered by (directed) cycles* if for any arrow from $x$ to $y$, there is a directed path from $y$ to $x$. It is clear that if a graph is strongly connected, then it is also weakly connected and covered by cycles. Conversely, if a graph is covered by cycles and weakly connected, then it is also strongly connected. A *strongly connected component* (*weakly connected component*) of a graph $G$ is a maximal subset $X$ of vertices such that the subgraph $G'$ with vertex set $X$ is strongly (weakly) connected.

### 1.4 The Reachability Graph

In this section, we investigate the set of states of a Petri net $(S, T, F, M_0)$ that can be reached during its execution(s). The concept of *reachability set* has already been introduced. It comprises all states that can be reached after the execution of arbitrary firing sequences, including the initial state which is reached after 'firing' the empty sequence. This set can be provided with some structure. Instead of just recording the reachable states, we may also record a relation between them, namely the information which transition has to be fired in order to get from one state to another one. In this way, we obtain the *reachability graph.*

**Definition 8.** *The reachability graph*
Let $N = (S, T, F)$ be a Petri net, let $M \in \mathbb{N}^S$ and let $\mathcal{E}(M)$ be the reachability set of $M$ in $N$. The *reachability graph* $RG(N, M)$ is defined to be an arc-labelled graph $(\mathcal{E}(M), E)$ with the following set of arcs:

$$E \;=\; \{(M_1, t, M_2) \mid M_1 \in \mathcal{E}(M) \;\wedge\; M_1\,[t\rangle M_2\}.$$

If $N = (S, T, F, M_0)$ is a Petri net with an initial marking, then the reachability graph of $N$, $RG(N)$, is defined to be $RG(N) = RG(N, M_0)$. ■

The reachability graph is always weakly, but not necessarily strongly, connected. Ignoring the arc labelling, it is also a multigraph, because $M_1\,[t\rangle M_2$ and $M_1\,[t'\rangle M_2$ does not necessarily imply $t = t'$. Thus there may be two or more arrows leading from $M_1$ to $M_2$.

Let us consider two examples. The Petri net shown in Figure 2 is a modification of the net shown previously in Figure 1. The loop between $s_1$ and $t_1$ was replaced by a single arc. The modified net has the reachability graph shown in Figure 3. It has two strongly connected components and three edges that are not covered by cycles.

If we designate $(0\;\;1\;\;2)^{\mathsf{T}}$ as a start state and define some additional "accepting" states, the Petri net can be viewed as an automaton that accepts the

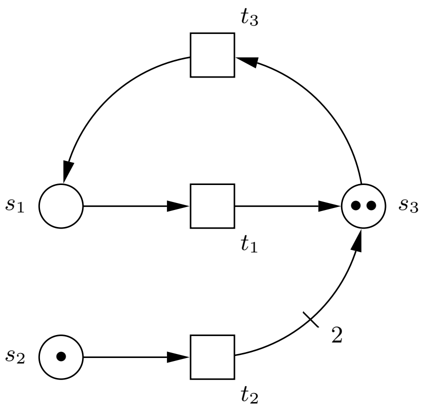



**Fig. 2.** A modification of Figure 1

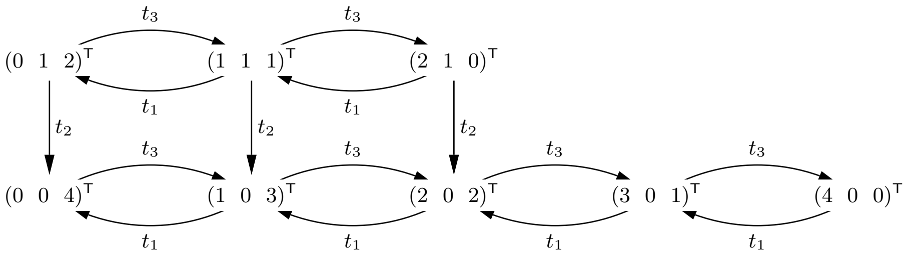



**Fig. 3.** The reachability graph of the Petri net shown in Figure 2; the initial marking is $(0\ \ 1\ \ 2)^{\mathsf{T}}$

language of all firing sequences ending in such an accepting state. With these additions we can study *Petri net languages*; such a study is, however, beyond the scope of the present notes. Suffice it here to say that Petri nets can accept non-regular languages, while, e.g., all languages of finite automata are regular. Consequently, while the reachability graph in our example looks quite similar to such finite automata, there must also be some Petri nets whose reachability graphs cannot be viewed that way. This is indeed the case.

Consider the net which was originally shown in Figure 1. It has a loop (instead of just a single arc) between $s_1$ and $t_1$, and its reachability graph $RG(N, M_0)$ with $M_0 = (0\ \ 1\ \ 2)^{\mathsf{T}}$ is shown in Figure 4. This reachability graph is infinite and can thus not be included fully in the Figure. Its representation in Figure 4 ends at some arbitrarily chosen (sufficiently early) point. The targets of the arcs at the bottom of the figure are not shown explicitly. This graph has infinitely many strongly connected components (each node being one).

An infinitely large reachabilibity graph is, of course, not a finite automaton. Also, in an infinite graph, the task of searching whether a given state is reachable is burdensome. Nevertheless, we may discover quickly that in our example,

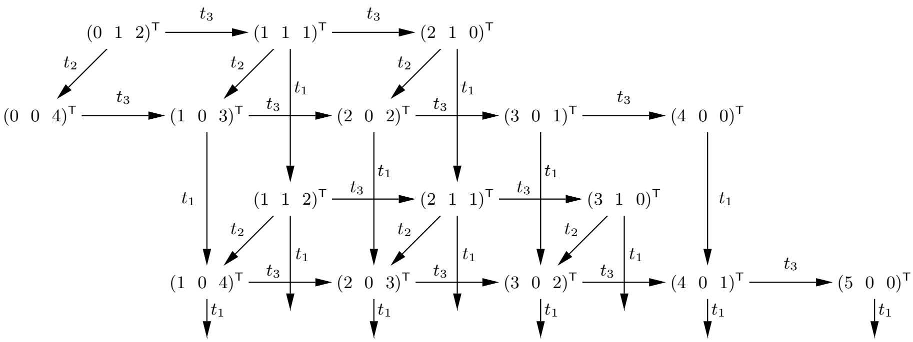



**Fig. 4.** Part of the reachability graph of the net shown in Figure 1; the initial state is $(0\ \ 1\ \ 2)^{\mathsf{T}}$

for instance, state $(1\ \ 0\ \ 9)^{\mathsf{T}}$ is reachable while state $(0\ \ 0\ \ 9)^{\mathsf{T}}$ is not. It suffices to examine the systematic structure of the graph in order to answer this or similar questions. In general, however, reachability graphs are much more complex, and the question whether a given state is reachable is very hard to answer. In fact, it has been an open question for several years whether or not this question is decidable. We cite the known result next.

**Theorem 1.** *Reachability is decidable*
The reachability problem, defined by

$$\mathsf{RP} \ = \ \{ \ (N, M, M') \mid N = (S, T, F) \text{ is a Petri net, } M, M' \in \mathbb{N}^S, \text{ and } M' \in \mathcal{E}(M) \ \},$$

is decidable. ■

## 1.5 Boundedness, Safeness, Liveness, Deadlock-Freeness, and Reversibility

The reachability graph reveals more information than just the set of reachable states. We discuss some interesting and relevant properties that can be inferred from the reachability graph.

**Definition 9.** *Boundedness and safeness*
Let $N = (S, T, F, M_0)$ be a Petri net. A place $s \in S$ is called *safe* if $M(s) \leq 1$ whenever $\sigma$ is a firing sequence and $M$ is a state with $M_0\,[\sigma\rangle M$; $s$ is called *$m$-bounded* (for $m \in \mathbb{N}$), if $M_0\,[\sigma\rangle M$ always entails $M(s) \leq m$. Place $s$ is *bounded* if it is $m$-bounded, for some $m \in \mathbb{N}$, otherwise it is *unbounded*.

A Petri net $N = (S, T, F, M_0)$ is called safe (bounded) if all places $s \in S$ are safe (bounded, respectively). $N$ is called $m$-bounded, for some $m \in \mathbb{N}$, if every place $s \in S$ is $m$-bounded. $N$ is called unbounded if it contains an unbounded place. ■

The safeness ($m$-boundedness) of $N$ can be deduced by inspecting the reachability graph $RG(N)$. Similarly, the safeness ($m$-boundedness) of any place can be deduced by inspection. We simply need to check all states which occur in $RG(N)$. This is practical only if $RG(N)$ is finite (and is, even then, likely to be extremely time-consuming).

**Definition 10.** *Liveness, deadlock-freeness, and reversibility*
Let $N = (S, T, F, M_0)$ be a Petri net.

- A transition $t \in T$ is called *singly live* or *not dead*, if there is a firing sequence $\sigma$ with $M_0\,[\sigma t\rangle$.
- A transition $t$ is called *weakly live*, if there is an infinite word $w \in T^\infty$ such that $t$ occurs infinitely often in $w$ and $M_0\,[w\rangle$ (meaning that $M_0\,[\sigma\rangle$ holds for every finite prefix $\sigma$ of $w$).

- A transition $t$ is called *live* or *strongly live*, if for every reachable marking $M \in \mathcal{E}(M_0)$, there is some firing sequence $\sigma$ with $M\,[\sigma t\rangle$.
- A transition $t$ is called *reversible*, if for every reachable marking $M \in \mathcal{E}(M_0)$, if $M\,[t\rangle M'$, then $M \in \mathcal{E}(M')$.

The Petri net $N$ is called (singly / weakly / strongly) live or reversible, if every transition in the net is (singly / weakly / strongly) live or reversible, respectively. $N$ is called *dead* if $N$ contains no singly live transition. A reachable marking $M \in \mathcal{E}(M_0)$ is called a *deadlock* if no transition is activated at $M$. $N$ is called *deadlock-free* if there is no marking $M \in \mathcal{E}(M_0)$ which is a deadlock. ■

It is easy to check on the reachability graph whether or not a transition is singly live. If it contains some arc $(M, t, M')$, then every path from $M_0$ to $M$ determines a firing sequence $\sigma$ with $M_0\,[\sigma\rangle M$, and in state $M$ we have $M\,[t\rangle$. That is, $t$ is singly live if and only if such an arc occurs in the reachability graph.

If the reachability graph is finite, then it is not hard to check whether a given transition $t$ is weakly live. If it is, then it can be fired arbitrarily often, and since the reachability graph is finite, it must contain a cycle with an arc of the form $(M, t, M')$. Conversely, if the reachability graph contains such a cycle, then $t$ is weakly live, since we can fire into the cycle, and then along the cycle arbitrarily often.

Checking strong liveness of a transition $t$ is not so easy, but it can also, in principle, be done on the reachability graph. For every marking $M$ contained in it, it must be checked whether a path leads from $M$ to an arc inscribed by $t$. For a finite reachability graph this means every *terminal* strongly connected component (i.e., with no arcs going out to other such components) must contain an edge labelled $t$.

Checking deadlock-freeness can be done by examining the reachability graph for vertices which have no output arc. The net is deadlock-free if and only if such vertices are absent.

Reversibility can be checked on the reachability graph as well. Recall that the reachability graph is always weakly connected. For a transition $t$ to be reversible, each edge labelled with $t$ must lie on some cycle (or equivalently, within some strongly connected component) of the reachability graph, and for the whole net to be reversible, the entire reachability graph must be strongly connected. Conversely, if the reachability graph is strongly connected, then the net is reversible. Note that this holds even if the initial marking is a deadlock.

The considerations of this section are subsumed as follows:

**Corollary 1.** *Properties that can be checked on the reachability graph*

If the reachability graph of a Petri net $N = (S, T, F, M_0)$ is finite, then there exist terminating algorithms which decide the following properties:

- whether a place is safe;
- whether a place is $m$-bounded;
- whether a place is bounded;
- whether a transition is dead (i.e., not singly live);
- whether a transition is weakly live;

– whether a transition is strongly live;
– whether a transition is reversible;
– and whether the net $N$ is safe ($m$-bounded, bounded, dead, singly live, weakly live, live, deadlock-free, or reversible). ■

## 1.6 Coverability Graphs

What happens if the reachability graph is not finite? Then none of the above questions can be solved efficiently using the reachability graph. Given that the reachability graph may be infinite and therefore unmanageable, does there perhaps exist a similar construction which always leads to a finite structure from which at least *some* interesting information about safeness, boundedness and liveness can be inferred? Such a structure has indeed been invented. It is called the *coverability graph*. Such a graph (indeed, a number of such graphs with similar properties, all of them finite) can be associated with a Petri net. If the latter is bounded, the coverability graphs defined in the literature would normally coincide with the unique reachability graph.

Some information is lost if the Petri net is unbounded, since coverability graphs are finite. However, the coverability graph(s) can still be used for some, but not all, analysis of the net. For instance, the question whether a net is bounded, can be decided on the coverability graph, as can the question which, if any, places are unbounded. As another example, however, the question whether a transition (or the entire net) is live, or weakly live, cannot be decided on the coverability graph only. For weak liveness it is sufficient to additionally know the full Petri net structure which might be partially hidden in the coverability graph. For liveness we know that it is at least as hard as the reachability problem. The latter is EXPSPACE-hard, and no upper complexity bounds are known at the present time. The reachability problem can also not be decided on the coverability graph, but there exist constructions using sequences of generalised coverability graphs to solve this problem. Further details about coverability graphs are beyond the scope of this tutorial.

## 1.7 Structural Boundedness, Structural Liveness, and Well-Formedness

In this section, we define two notions that are related to boundedness and liveness, but pertain to nets without markings, rather than to net systems as before. Hence there is in general no single reachability graph on which they could be checked. We introduce these properties by means of Figures 5 to 7.

Consider the system shown on the left-hand side of Figure 5. It is 2-bounded but not safe. It is also not live. Its reachability graph is shown on the right-hand side of the figure. The system shown in Figure 6 is unbounded. However, it is live. The system shown in Figure 7 is safe (i.e., 1-bounded) as well as live, but it is not reversible. The reachability graph of this net is shown on the right-hand side of Figure 7. (In order to avoid writing long vectors, the reachable markings were written in a short-hand notation. They are represented as strings of place names, where a place appears as often as there are tokens on it.) The graph is composed of two separate strongly connected components, one containing just $M_0 = s_2 s_5$ and the other containing all other markings.

**Definition 11.** *Structural boundedness, structural liveness, and well-formedness*
A net $N = (S, T, F)$ is called

- *structurally bounded*, if for all markings $M$, the system $(S, T, F, M)$ is bounded;
- *structurally live*, if there is some marking $M$ such that the system $(S, T, F, M)$ is live;
- *well-formed*, if there is some marking $M$ such that the system $(S, T, F, M)$ is bounded and live.

A marking $M$ of a net $(S, T, F)$ is called *live (bounded)* if the system $(S, T, F, M)$ is live (bounded). ■

Note that there is a significant difference between the notions of structural boundedness and structural liveness: while in the first case, boundedness must hold for *all* possible initial markings, in the second case it is sufficient that there exists *some* initial marking for which liveness holds.

If we omit the three tokens from the net shown on the left-hand side of Figure 5, then we get a net which is not structurally live. To see this, we actually have to investigate not just the initial marking which is shown in the figure, but any other initial marking as well, that is, we need to consider an infinite number of reachability graphs and check that none of them belongs to a live net. Indeed, let an arbitrary initial marking be given and consider a maximal sequence that arises by choosing $b$ instead of $a$ whenever both are enabled. It is easy to see that such a sequence always leads to a deadlock, and thus, the net is not live. On the other hand, the net is structurally bounded. To see this, we again need an argument showing that the net is not only bounded for the marking shown in the figure, but

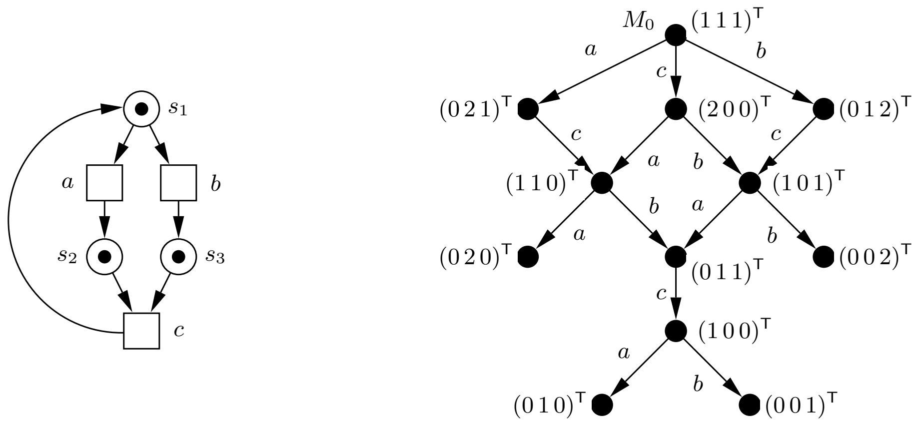



**Fig. 5.** Neglecting the tokens: structurally bounded, but not structurally live

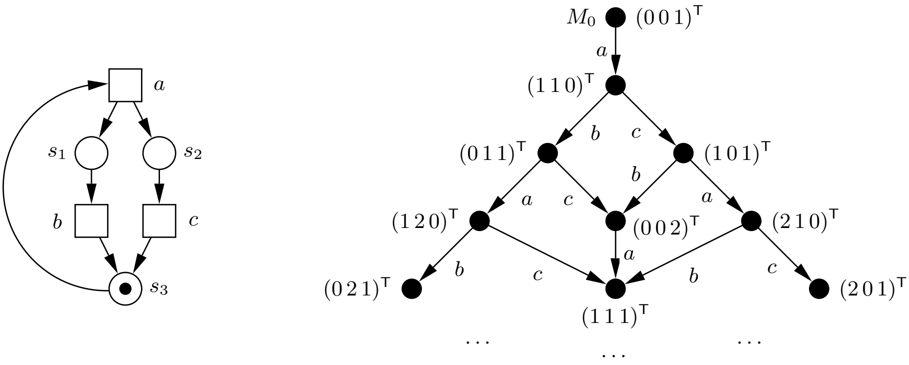



**Fig. 6.** Neglecting the token: structurally live, but not structurally bounded

for all other possible initial markings as well. In fact, this follows easily from the fact that for any arbitrary initial marking, the overall number of tokens of this net can never increase. Finally, this net is not well-formed, because it has no live marking and, *a fortiori*, no bounded and live marking.

The net shown on the left-hand side of Figure 6 (neglecting the token) is structurally live. To see this, it suffices to exhibit a live marking; for instance, the marking shown in the figure is indeed live. The net is, however, not structurally bounded. To see this, it suffices to exhibit a non-bounded marking; for instance, the marking shown in the figure is not bounded. The net is not well-formed, either, because its only bounded marking (which is the empty – i.e., token-free – marking) fails to be live.

The net shown on the left-hand side of Figure 7 (neglecting the two tokens) is well-formed. To see this, it suffices to exhibit a live and bounded marking; for instance, the marking shown in the figure is both live and bounded. It follows that the net is structurally live. It is not clear, at this point, whether the net is also structurally bounded. Later, we will prove that this is indeed the case.

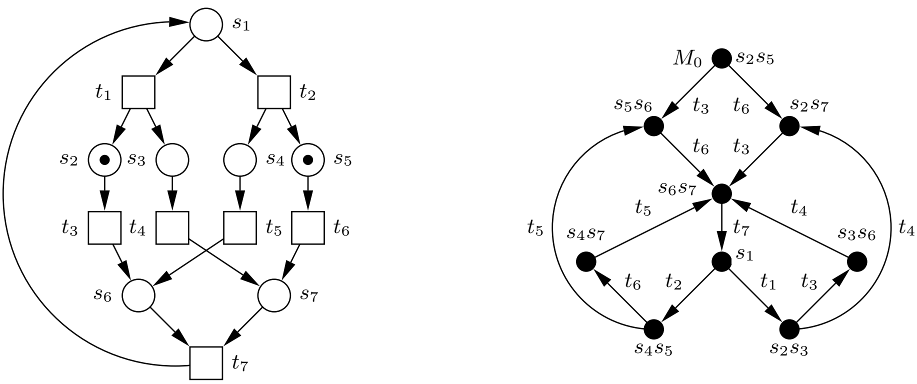



**Fig. 7.** Neglecting the tokens: structurally bounded and structurally live

### 1.8 Bibliographical Remarks and Further Reading

Petri nets were conceived in [Pet62] and brought into the form presented in these notes, called the place/transition nets, by several research teams, prominent amongst whom were Hartmann Genrich, Amir Pnueli, et al. [CHEP71, GL73]. Around the same time, a closely related model, the vector addition systems, was put forward and investigated by Richard Karp and others [KM69]. The notion of a coverability tree was defined in [KM69] to serve as the basis for an algorithm to decide boundedness. Both models have sparked several interesting developments and led to famous results, such as the decidability of reachability which was proved independently by Ernst Mayr in his dissertation [May80, May84] and by Rao Kosaraju [Kos82]. This particular result was made more widely accessible through a new proof by Jean-Luc Lambert [Lam92], by further work by Jérôme Leroux [Ler09], and to a German-speaking audience, by a textbook by Lutz Priese and Harro Wimmel [PW03]. The proof ideas which were contained in these works have proved useful in obtaining further results, such as described in [Wim04] and in [HMW10]. Even before reachability was known to be decidable, a number of properties had been shown to be reducible to reachability [Hac74].

The notions of liveness and boundedness originated, to our knowledge, from the early work cited above [KM69, CHEP71, GL73]. Soon after these publications, Leslie Lamport coined the terminology of liveness versus safety properties in connection with program verification [Lam77]. He later said that he used these terms based on a slight misunderstanding of the similar terms coming from Petri net theory [Lam]. As readers proficient in verification will surely be aware of, they have since led a very meaningful and independent scientific life as well.

There are several textbooks and overview articles on place/transition nets, e.g. by Wolfgang Reisig [Rei85], James Lyle Peterson [Pet81] and Tadao Murata [Mur89]. The reader is also referred to the bibliography entries mentioned in `http://www.informatik.uni-hamburg.de/TGI/PetriNets/bibliographies/`

## 2 Linear-Algebraic Structure of Petri Nets

Both a Petri net $N$ and its reachability graph $RG(N)$ are mathematical structures known as *(multi-)graphs*. However, the graph-theoretical structure of $N$ bears no apparent relationship to the graph-theoretical structure of $RG(N)$. It may be the case that $N$ is a very complicated graph and that $RG(N)$ is extremely simple, but it may also be the case that $RG(N)$ is very complex even though $N$ looks rather innocent.

Normally, there is a large discrepancy between the size of $N$ and the size of $RG(N)$. While $N$ is of the order of a computer program's size (which may vary between a few and several hundreds of millions of lines), $RG(N)$ is often exponentially larger than $N$. Therefore, it has long been one of the objectives of net theory to be able to deduce some properties of $RG(N)$ from properties of $N$ itself. It is particularly interesting to find results that would allow one to check properties such as those of Corollary 1 (e.g., boundedness, or liveness) by checking the structure of the net only, without constructing the reachability graph (or the coverability graph). This idea has been known by the name of *structure theory*, or the *structural analysis*, of Petri nets.

Structure theory will be investigated from different points of view.

In the present section, we will exploit the *linear-algebraic* structure of $N$ in order to deduce properties of its behaviour under a given initial marking. In section 3, by contrast, we will exploit the *graph-theoretical* structure of $N$ for the same purpose. The final section 4 reveals some connections between net properties that are defined partly structurally and partly behaviourally. In most cases, the objective is to use *static properties* of the net $N$, i.e., properties that can be 'read off' its structure, in order to deduce *dynamic properties*, i.e. properties such as boundedness or liveness or the absence of conflicts.

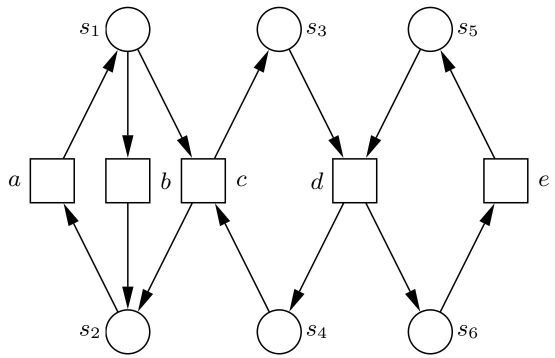
## 2.1 Incidence Matrix, Marking Equation, Marking Inequality, and Realisability

The incidence matrix of a Petri net is derived from the two matrices $\mathbb{F}$ and $\mathbb{B}$, and it forms the basis for linear-algebraic manipulations of Petri nets. An example is shown in Figure 8.

**Definition 12.** *Incidence matrix*
Let $N = (S, T, F)$ be a net. The *incidence matrix*, or *connectivity matrix*, of $N$ is defined as the function $C\colon S \times T \to \mathbb{Z}$ with $C = \mathbb{F} - \mathbb{B}$. ■

If $C$ has no rows, or no columns, or both, linear algebra cannot reasonably be expected to work. Therefore, we will assume that there is at least one transition and at least one place in the nets we consider.

Incidence matrices do not capture loops. In general, $N$ cannot be reconstructed uniquely from $C$, because some information about loops is lost. For example, consider the net which consists only of a place $s$ and a transition $t$, without any arrows.




| $C$ | $a$ | $b$ | $c$ | $d$ | $e$ |
|---|---|---|---|---|---|
| $s_1$ | 1 | −1 | −1 | 0 | 0 |
| $s_2$ | −1 | 1 | 1 | 0 | 0 |
| $s_3$ | 0 | 0 | 1 | −1 | 0 |
| $s_4$ | 0 | 0 | −1 | 1 | 0 |
| $s_5$ | 0 | 0 | 0 | −1 | 1 |
| $s_6$ | 0 | 0 | 0 | 1 | −1 |

**Fig. 8.** An unmarked net (l.h.s.) and its incidence matrix (r.h.s.)

This net has exactly the same incidence matrix as the net that consists of $s$ and $t$ and has two additional arrows, one from $s$ to $t$ and another one from $t$ to $s$.

Using the incidence matrix, the marking reached after a firable sequence can be computed linear-algebraically. To see how this can be done, we need to define Parikh vectors.

**Definition 13.** *Parikh vector*
Let $\tau = t_1 \dots t_k \in T^*$ be a sequence of transitions from $T$. Let $\#(t,\tau)$ denote the number of times transition $t$ occurs in $\tau$. The *Parikh vector* or *occurrence count vector* of $\tau$ is defined as a (column) T-*vector* (that is, a $T$-based column vector) $\mathcal{P}(\tau) \in \mathbb{N}^T$ which contains, at entry $t$, the occurrence count $\#(t,\tau)$. ∎

For example, for four transitions $\{t_1, t_2, t_3, t_4\}$,

$$\mathcal{P}(\varepsilon) = (0\,0\,0\,0)^\mathsf{T}, \quad \mathcal{P}(t_2) = (0\,1\,0\,0)^\mathsf{T}, \quad \text{and } \mathcal{P}(t_1t_2t_4t_2t_3t_4) = (1\,2\,1\,2)^\mathsf{T}.$$

Comparing Definition 12 with the firing rule (section 1), it can be seen that an entry $C(s,t)$ indicates how the token count on $s$ changes through the firing of $t$. That is, if $M_1\,[t\rangle M_2$, then $M_2 = M_1 + C \cdot \mathcal{P}(t) \quad (= M_1 + (\mathbb{F} - \mathbb{B})(\mathcal{P}(t)))$. For example, consider the net shown in Figure 9. The initial marking can be represented as a column vector $M_0 = (1\,0\,0\,2\,1\,0)^\mathsf{T}$. Transition $c$ can fire, and we easily check that indeed,

$$\underbrace{\begin{pmatrix}0\\1\\1\\1\\1\\0\end{pmatrix}}_{M} = \underbrace{\begin{pmatrix}1\\0\\0\\2\\1\\0\end{pmatrix}}_{M_0} + \underbrace{\begin{pmatrix}1 & -1 & -1 & 0 & 0\\ -1 & 1 & 1 & 0 & 0\\ 0 & 0 & 1 & -1 & 0\\ 0 & 0 & -1 & 1 & 0\\ 0 & 0 & 0 & -1 & 1\\ 0 & 0 & 0 & 1 & -1\end{pmatrix}}_{C} \cdot \underbrace{\begin{pmatrix}0\\0\\1\\0\\0\end{pmatrix}}_{\mathcal{P}(c)}$$

This idea can be generalised as follows.

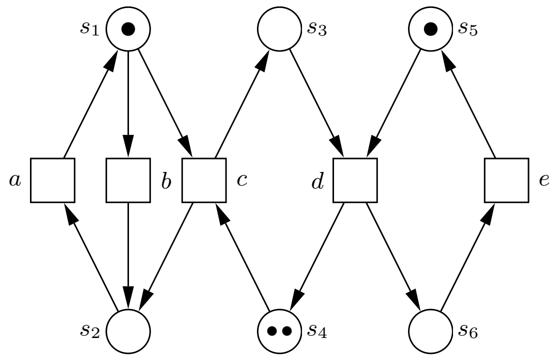



**Fig. 9.** The net of Figure 8 with an initial marking $M_0$

**Lemma 2.** *Firing lemma and marking equation*
If $M_1[\tau\rangle M_2$, then $M_2 = M_1 + C \cdot \mathcal{P}(\tau)$.

*Proof:* The claim follows easily by induction on the length of $\tau$. ∎

The term *marking equation* refers to the conclusion

$$\boxed{M_2 = M_1 + C \cdot \mathcal{P}(\tau)}$$

of the lemma.

For example, $\tau = cabda$ is a firing sequence in the net shown in Figure 9. If we want to calculate the marking $M$ reached after this sequence, we may use Lemma 2 as follows:

$$\underbrace{\begin{pmatrix}1\\0\\0\\2\\1\\0\end{pmatrix}}_{M_0} + \underbrace{\begin{pmatrix}1 & -1 & -1 & 0 & 0\\ -1 & 1 & 1 & 0 & 0\\ 0 & 0 & 1 & -1 & 0\\ 0 & 0 & -1 & 1 & 0\\ 0 & 0 & 0 & -1 & 1\\ 0 & 0 & 0 & 1 & -1\end{pmatrix}}_{C} \cdot \underbrace{\begin{pmatrix}2\\1\\1\\1\\0\end{pmatrix}}_{\mathcal{P}(cabda)} = \underbrace{\begin{pmatrix}1\\0\\0\\2\\0\\1\end{pmatrix}}_{M}$$

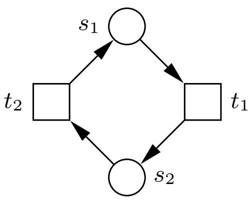
The lemma states that the validity of the marking equation is a *necessary* condition for a given sequence of transitions leading to a certain marking. Stated differently, if the marking equation is not satisfied for a given vector of transition counts, then there is no firing sequence having this vector as a Parikh vector reaching the goal marking.

The marking equation is *not*, in general, a sufficient condition for firability. That is, $M_2 = M_1 + C \cdot y$ and $y \geq 0$ do not necessarily entail $M_1 [\tau\rangle M_2$ for some firing sequence $\tau$ with $\mathcal{P}(\tau) = y$. As a counterexample, we may consider Figure 10 where the initial marking is empty, i.e. equal to $(0\,0)^\mathsf{T}$.

The above may be expressed in a slightly different way. If $M\,[\tau\rangle$ and $y = \mathcal{P}(\tau)$, then the two inequalities

$$\boxed{0 \leq y \quad \text{and} \quad 0 \leq M + C \cdot y}$$

are satisfied. This is called the *marking inequality.* A T-vector $y \in \mathbb{N}^T$ is called *realisable* from some marking $M$ if there is a firing sequence $M\,[\tau\rangle$ with $\mathcal{P}(\tau) = y$.




| $C$ | $t_1$ | $t_2$ |
|---|---|---|
| $s_1$ | $-1$ | $1$ |
| $s_2$ | $1$ | $-1$ |
| $\mathcal{P}(t_1t_2)^\mathsf{T}$ | $1$ | $1$ |

**Fig. 10.** A system in which $(0\;0)^\mathsf{T} = (0\;0)^\mathsf{T} + C{\cdot}\mathcal{P}(t_1t_2)$ but $\tau = t_1t_2$ is not firable

If some arbitrary integer-valued vector $y$ satisfies the marking inequality, it is still not guaranteed to be realisable.

### 2.2 Transposition Lemmata

Apart from the componentwise comparisons of vectors,

$$\begin{array}{rl} & M \geq M' \iff \forall k: \;\; M(k) \geq M'(k) \\ \text{and} & M > M' \iff M \geq M' \wedge M \neq M', \end{array}$$

we will also need the following concept of strict comparison.

**Definition 14.** *Strictly greater*
Let $k \in \mathbb{N}$ and let $M, M' \in \mathbb{Z}^k$ be vectors over $\mathbb{Z}$. Then $M$ is called *strictly greater* than $M'$ (in symbols: $M \gg M'$) if $M(i) > M'(i)$ for all $i \in \{1, \ldots, k\}$. The notion of *strictly less* is defined analogously.

A vector $M$ is called *nonnegative* if $M \geq 0$, *semipositive* if $M > 0$, and *positive* if $M \gg 0$, where 0 denotes the null vector. ■

For example, a Parikh vector is always nonnegative. Some entries may be 0 if the corresponding transition does not occur in the sequence on which the vector is based. However, Parikh vector entries are never negative.

We exploit a useful principle which is sometimes known in Linear Algebra by the name of *transposition principle* or *alternation principle* and which goes back to Gordan [Gor1873], Farkas [Far1902] and others. The following lemma explicates this principle. There are several versions of this lemma, but in the present notes we use only this one.

**Lemma 3.** *A transposition lemma*
Let $A$ be a matrix with rational entries (that is, entries from the set of rational numbers). Then *exactly one* of the following statements is valid:
(i) There exists a (rational) vector $x$ with $x \gg 0$ and $A^\mathsf{T} \cdot x \leq 0$. Here, $A^\mathsf{T}$ denotes the transposed matrix of $A$.
(ii) There exists a (rational) vector $y \geq 0$ with $A \cdot y > 0$.

*Proof:* It is easy to see that (i) and (ii) cannot be true at the same time. This is because (i) $\wedge$ (ii) entails

$$0 \;\geq\; y^\mathsf{T} \cdot A^\mathsf{T} \cdot x \;>\; 0,$$

which is a contradiction. The first inequality comes from $y \geq 0$ (ii) and from $A^\mathsf{T} \cdot x \leq 0$ (i), if the middle product is associated as $y^\mathsf{T} \cdot (A^\mathsf{T} \cdot x)$. The second inequality comes from $x \gg 0$ (i) and $A \cdot y > 0$ (ii), if the product is associated as $(y^\mathsf{T} \cdot A^\mathsf{T}) \cdot x$.

The proof that (i) $\vee$ (ii) holds true is non-trivial; the interested reader is referred to [Schr86]. ■

This lemma can easily be lifted to integers for the vectors $x$ and $y$ (whichever exists).

### 2.3 Structural Boundedness, Infinite Executions, and Dickson's Lemma

This section contains some small examples involving linear-algebraic arguments, including a typical application of the transposition principle. First, we give an elementary linear-algebraic characterisation of structural boundedness. Then, we characterise the existence of infinite firing sequences linear-algebraically.

**Proposition 1.** *Characterisation of structural boundedness*
Let $N$ be a net and let $C$ be the incidence matrix of $N$. The following statements are equivalent:
**(A)** $N = (S, T, F)$ is structurally bounded.
**(B)** There exists a vector $x \in \mathbb{N}^{|S|}$ with $x \gg 0$ and $C^{\mathsf{T}} \cdot x \leq 0$.

*Proof:* **(A)⇒(B)** can be shown by contraposition. Assume $\neg$**(B)**, i.e., there is no vector $x$ as in **(B)**. By Lemma 3, there is some vector $y \in \mathbb{N}^{|T|}$ with $C \cdot y > 0$. We choose some marking $M$ which guarantees that a firing sequence $\tau$ with $\mathcal{P}(\tau) = y$ is firable from it, for instance the following:

$$M(s) = \sum_{t \in s^\bullet} (F(s,t) \cdot y(t)), \text{ for } s \in S.$$

Let $M'$ be defined by $M\,[\tau\rangle M'$. The firing lemma yields

$$M' = (\text{ by Lemma 2 }) \ M + C \cdot \mathcal{P}(\tau) = (\text{ by } \mathcal{P}(\tau) = y \ ) \ M + C \cdot y > (\text{ by } C \cdot y > 0 \ ) \ M.$$

Furthermore, from $M' > M$ we deduce the existence of a place $r$ with $M'(r) > M(r)$. Since $\tau$ can be fired arbitrarily often from $M'$ because of $M' > M$, at least the place $r$ is unbounded. Hence $\neg$**(A)** holds.

To show **(B)⇒(A)**, we choose $x$ such that property **(B)** is satisfied. Let $M_1$ be an arbitrary marking of $N$ and let $M_1\,[\tau\rangle M_2$ with an arbitrary firing sequence $\tau$. Using **(B)** we get:

$$x^{\mathsf{T}} \cdot M_2 \ = \ x^{\mathsf{T}} \cdot (M_1 + C \cdot p(\tau)) \ = \ x^{\mathsf{T}} \cdot M_1 + x^{\mathsf{T}} \cdot (C \cdot p(\tau)) \ \leq \ x^{\mathsf{T}} \cdot M_1$$

where the first equality follows from the firing lemma and the last inequality from **(B)**. For $s \in S$,

$$x(s) M_2(s) \ \leq \ (\text{ by } x \geq 0 \ ) \sum_{r \in S} x(r) M_2(r) \ = \ x^{\mathsf{T}} \cdot M_2 \ \leq \ (\text{ by the above }) \ x^{\mathsf{T}} \cdot M_1.$$

Therefore, $(x^{\mathsf{T}} \cdot M_1)/x(s)$ is an upper bound for the number of tokens on an arbitrary place $s$ in $M_2$, depending neither on $M_2$ nor on $\tau$. Therefore, place $s$ is bounded. Since the above is true for arbitrary $M_1$ and for arbitrary $s$, Property **(A)** is satisfied. ■

In order to characterise the existence of a marking from which an infinite sequence of transitions can be fired, we exploit a lemma which is useful in several other circumstances as well.

**Lemma 4.** *Dickson's lemma*

Let $n \in \mathbb{N}$ and let $x_1, x_2, x_3, \ldots$ be an infinite sequence of vectors in $\mathbb{N}^n$. Then there are indices $i_1 < i_2 < i_3 < \ldots$ with

$$x_{i_1} \leq x_{i_2} \leq x_{i_3} \leq \ldots$$

*Proof:* Consider the case that $n = 1$. Then the sequence of vectors is simply a sequence of natural numbers. If one of them occurs infinitely often in the sequence, we are done, since the corresponding subsequence is (weakly) monotonically increasing. Otherwise we choose $i_1$ as the last occurrence of the minimum, $i_2$ as the *subsequently* last occurrence of the (new) minimum, and so on.

The case that $n > 1$ can be dealt with by induction and componentwise consideration. From an infinite sequence of vectors with $n$ components, we first choose an infinite subsequence that is (weakly) monotonically increasing with respect to components 1 to $n-1$. This can be done by induction hypothesis. From this subsequence, we then choose another subsequence which (weakly) increases with respect to the last ($n$th) component. This can be done as in the case that $n = 1$. The resulting sub-subsequence is (weakly) monotonically increasing with respect to all $n$ components. ■

The lemma can be applied to the sequence of markings occurring in an infinite firing sequence, as follows.

**Proposition 2.** *Existence of an infinite firing sequence*

For an unmarked net $N$, there is some marking $M_0$ such that an infinite firing sequence from $M_0$ exists, if and only if the system of inequalities $C \cdot y \geq 0,\ y > 0$ has a solution.

*Proof:* ($\Rightarrow$): Let $M_0$ be a marking of $N$ with $M_0\,[t_1\rangle M_1\,[t_2\rangle M_2\,[t_3\rangle\ \ldots$ By Lemma 4, there exist indices $i < j$ with $M_i \leq M_j$ and $M_i\,[t_{i+1} \ldots t_j\rangle M_j$. The vector $y$ defined by $y = \mathcal{P}(t_{i+1} \ldots t_j)$ solves the system of inequalities given in the proposition, since: $y \geq 0$, because $y$ is a Parikh vector; $y \neq 0$, because $i < j$; and $C \cdot y \geq 0$ by $M_j = M_i + C \cdot y$ (firing lemma) and by $M_j \geq M_i$.

($\Leftarrow$): Let $y$ be a solution of the system of inequalities given in the lemma. As in the proof of Proposition 1, we can find a 'sufficiently large' marking $M$ that activates a firing sequence $\sigma$ with $\mathcal{P}(\sigma) = y$. Let $M'$ be the marking defined by $M\,[\sigma\rangle M'$. By the firing lemma, $M' = M + C \cdot y$, whence, by $C \cdot y \geq 0$, we have $M \leq M'$. Hence $\sigma$ can be iterated indefinitely, and $\sigma\sigma\sigma\ldots$ is firable from $M$. Moreover, $\sigma\sigma\sigma\ldots$ is an infinite sequence since $\sigma \neq \varepsilon$ because of $y \neq 0$. ■

## 2.4 S-Invariants and T-Invariants

In the previous section, vectors $x$ (Proposition 1) and $y$ (Proposition 2) satisfied some inequalities, $C^{\mathsf{T}} \cdot x \leq 0$ and $C \cdot y \geq 0$, respectively. The special case that these inequalities become actual equalities, viz. $C^{\mathsf{T}} \cdot x = 0$ and $C \cdot y = 0$, is of particular importance:

**Definition 15.** *Place invariants and transition invariants*
A vector $x \in \mathbb{Z}^{|S|}$ is called *S-invariant* or *place invariant*, if $C^{\mathsf{T}} \cdot x = 0$.
A semipositive S-invariant $x$ is *minimal* if there is no S-invariant $x'$ with $0 < x' < x$.

A vector $y \in \mathbb{Z}^{|T|}$ is called *T-invariant* or *transition invariant*, if $C \cdot y = 0$. Minimality is defined similarly as for S-invariants. ■

The semipositive, the positive, and the minimal invariants will turn out to be of primary interest.

As an example, consider Figure 11. $x_1$ is a minimal semipositive S-invariant. $x_2$ is a non-semipositive S-invariant. $x_3$ is a semipositive S-invariant which arises from another semipositive S-invariant, namely $(0, 0, 1, 1, 0, 0)^{\mathsf{T}}$, by multiplication with 2; thus, it is not minimal. $x_3$ is also the sum of $x_1$ and $x_2$. $y_1$ and $y_2$ are minimal semipositive T-invariants.

The following lemma follows directly from the firing lemma.

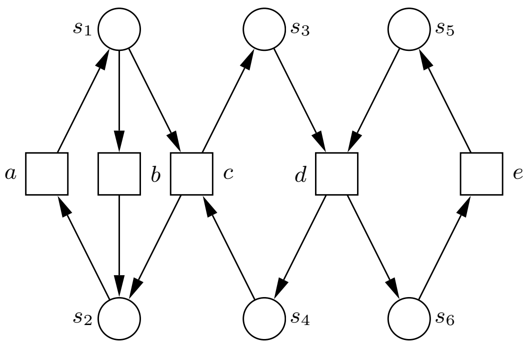
**Lemma 5.** *Basic properties of S- and T-invariants*

Let $x$ be an S-invariant of $N$ and let $M_1, M_2$ be markings of $N$ with $M_1\,[\tau\rangle M_2$, for some sequence $\tau$. Then $x^{\mathsf{T}} \cdot M_1 \;=\; x^{\mathsf{T}} \cdot M_2$.

Let $M$ be a marking of $N$ with $M\,[\tau\rangle M$ for some sequence $\tau$. Then $\mathcal{P}(\tau)$ is a T-invariant of $N$.

Conversely, if $M\,[\tau\rangle$ and $\mathcal{P}(\tau)$ is a T-invariant of $N$, then $M\,[\tau\rangle M$. ■

Informally, the first part of this lemma states that the $x$-weighted marking on any S-invariant $x$ is constant. In particular, if a net has a positive S-invariant, then it is necessarily structurally bounded. The second part of the lemma states that any reproduction sequence generates a T-invariant. In particular, any reproduction sequence containing every transition at least once, generates a positive T-invariant. A weak converse also holds true: if $\mathcal{P}(\tau)$ is a T-invariant, then $M\,[\tau\rangle M$ for any marking $M$ enabling $\tau$.




| $C$ | $a$ | $b$ | $c$ | $d$ | $e$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −1 | −1 | 0 | 0 | 1 | −1 | 0 |
| 2 | −1 | 1 | 1 | 0 | 0 | 1 | −1 | 0 |
| 3 | 0 | 0 | 1 | −1 | 0 | 0 | 2 | 2 |
| 4 | 0 | 0 | −1 | 1 | 0 | 0 | 2 | 2 |
| 5 | 0 | 0 | 0 | −1 | 1 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 |
| $y_1^{\mathsf{T}}$ | 1 | 1 | 0 | 0 | 0 | | | |
| $y_2^{\mathsf{T}}$ | 1 | 0 | 1 | 1 | 1 | | | |

**Fig. 11.** The example of Figure 8, with some S- and T-invariants

### 2.5 Positive S-Invariants and T-Invariants

A positive S-invariant, that is, one which assigns a number $\geq 1$ to every place, is said to *cover* the net. Similarly, a positive T-invariant is said to *cover* a net. As these properties occur frequently, often in connection with well-formedness, we abbreviate them as follows:

(PS): The net under consideration is covered by a positive S-invariant.
(PT): The net under consideration is covered by a positive T-invariant.
(WF): The net under consideration is well-formed, i.e., it has a live and bounded marking.

(WF) is stronger than (PT), because of the following lemma. However, (WF) is not stronger than (PS), since there are Petri nets satisfying (WF) but not (PS). Finding an example is left as an exercise to the reader. (This is not entirely trivial.)

**Proposition 3.** *On the existence of positive T-invariants*
Let $N$ be a well-formed Petri net. Then $N$ has a positive T-invariant.

*Proof:* Let $M$ be a live and bounded marking of $N$. By liveness, there exists an infinite firing sequence $\tau = \tau_1\tau_2\tau_3\ldots$ such that every sequence $\tau_i$ contains *all* transitions of $N$. Define markings $M_i$ by $M[\tau_1 \ldots \tau_i\rangle M_i$. By boundedness, not all markings $M_i$ can be different. Hence there are two indices $k, j$ with $k < j$ and $M_k = M_j$. The subsequence $M_k\,[\tau_{k+1} \ldots \tau_j\rangle M_j$ is repetitive, as it reproduces the marking $M_k{=}M_j$. Thus, $\mathcal{P}(\tau_{k+1} \ldots \tau_j)$ is a T-invariant by the second part of Lemma 5, which is positive since $\tau_{k+1} \ldots \tau_j$ has the nonempty suffix $\tau_j$ and thus contains at all transitions, by definition of $\tau_j$. ■
If (PS) and (PT) are true for some net, then there are repercussions on its graph-theoretical structure.

**Proposition 4.** *Cycle-coveredness of nets covered by positive S- and T-invariants*

Let $N$ be a Petri net satisfying (PS) and (PT). Then $N$ is covered by cycles.

*Proof:* (*Sketch.*) Let $N = (S, T, F)$ and choose $x$ such that $C^{\mathsf{T}}{\cdot}x = 0$ and $x \gg 0$, and $y$ such that $C{\cdot}y = 0$ and $y \gg 0$. Consider an arrow $(u, v)$ in $F$. We want to prove that there is a directed path from $v$ to $u$.

**First Case:** $u \in S$ and $v \in T$.
The basic proof idea is to restrict $y$ to transitions 'after' $v$. Let $y' \colon T \to \mathbb{N}$ be defined as follows:

$y'(t) = y(t)$ if a directed (possibly empty) path leads from $v$ to $t$ in $N$,
$y'(t) = 0$ for all other transitions $t$.

It is then possible to show that (i) $y'$ is a T-invariant, and (ii) some input transition $w \in {}^{\bullet}u$ satisfies $y'(w) > 0$. By the definition of $y'$, a directed path leads in $N$ from $v$ to $w$, and therefore also from $v$ to $u$.

**Second Case:** $u \in T$ and $v \in S$.
Consider the *dual net* $N^d = (T, S, F)$, in which places and transitions are exchanged but arrows are retained. The incidence matrix of $N^d$ is $-C^\mathsf{T}$. Hence $x$ is a positive T-invariant and $y$ is a positive S-invariant of $N^d$, and the second case is reducible to the first case. ■

**Corollary 2.** *Strong connectedness of nets covered by positive S- and T-invariants*

> Let $N$ be a weakly connected Petri net containing a positive S-invariant and a positive T-invariant.
> Then $N$ is strongly connected. ■

### 2.6 Rank, Conflict Clusters, and Sets of Presets

There are some interesting connections between the structural liveness of a Petri net and the rank of its incidence matrix $C$. The *column rank* (*row rank*) of $C$ is defined as the maximal number of linearly independent column (row, respectively) vectors in $C$. Since, as is known from your favourite course on Linear Algebra, the column rank and the row rank of any matrix $C$ are identical, the *rank* of $C$ is simply defined as one of them, say the column rank. The rank of a Petri net $N$ is defined as the rank of its incidence matrix.

> In the remaining part of this section, we will assume that $N$ is weakly connected and *plain*, that is, the function $F$ does not yield values greater than 1 (or, equivalently, the matrices $\mathbb{B}$ and $\mathbb{F}$ have values in the set $\{0, 1\}$).

A first observation is that the rank of $C$ is less than $|T|$, the number of transitions, provided that $N$ is covered by a positive T-invariant. This is simply because $C \cdot y = 0$ and $y \gg 0$ just means that some positive linear combination of the columns of $C$ equals 0, which means that its columns are linearly dependent and the rank of $C$ cannot exceed $|T| - 1$. We may combine this observation with Proposition 3, to see that any well-formed net has column rank $\leq |T| - 1$.

As a special case, consider a simple directed cycle and a function which assigns the number 1 to every transition of the cycle, as shown on the left-hand side of Figure 12. This is already a positive T-invariant, and the column rank of the


**Fig. 12.** A simple cycle (l.h.s.) and a modification (r.h.s.)

net actually equals $|T|-1$. Suppose now that we change such a cycle slightly by putting two successive transitions into conflict rather than sequence, as depicted on the right-hand side of Figure 12. Then, in some reproducing firing sequence, either one or the other can be chosen. In this way, we get two semi-positive, non-positive T-invariants. Their sum is a positive T-invariant covering the net with column rank $|T|-2$. There is one place less in the net (compared to the left hand side of Fig. 12), leading to less potential conflicts between transitions. Thus, one might be led to suspect a connection between the rank of a net and its 'degree of conflict'. The notions of a *conflict cluster* and of a *preset*, as follows, are designed to make this suspicion more precise. Informally, both play a role in capturing the 'degree of conflict' of a net.

**Definition 16.** *Conflict clusters, and the set of presets*
Let $N = (S, T, F)$ be a plain Petri net.

Fot $t, t' \in T$, let $t \sim_0 t'$ if ${}^\bullet t \cap {}^\bullet t' \neq \emptyset$ (i.e., if there is a potential conflict between $t$ and $t'$). Let $\sim \subseteq T \times T$ be the reflexive and transitive closure of $\sim_0$. A *conflict cluster* of $N$ is defined as an equivalence class of the equivalence relation generated by $\sim$. The set of all conflict clusters of $N$ is denoted by $CC_N$.

The set of all non-empty *presets* of $N$ is defined as $PRESETS_N = \{{}^\bullet t \mid t \in T \wedge {}^\bullet t \neq \emptyset\}$. ■

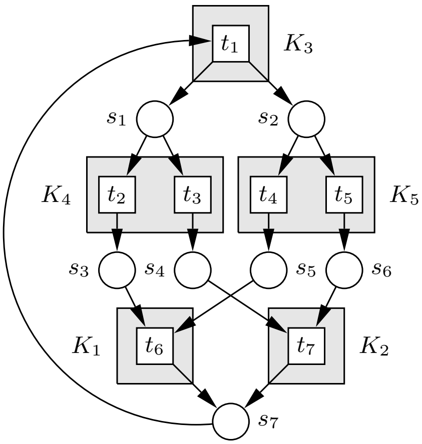
Notice that in the simple cycle on the left-hand side of Figure 12, the relation $\sim_0$ is the identity relation, and we have three conflict clusters. Also, there are three presets. The net has rank $2 = |T|-1$ overall. On the right-hand side of Figure 12, however, $\sim_0$ is not the identity since we have $t_2 \sim_0 t_3$. In all, there are two conflict clusters, as well as two presets. The rank of the net's matrix is $1 = |T|-2$.

### 2.7 Sufficient and Necessary Conditions for Structural Liveness

**Theorem 2.** *Sufficient condition for the existence of a live marking*
Assume that $N$ is a weakly connected, plain net covered by a positive S-invariant and a positive T-invariant.
If the rank of $C$ is strictly less than $|CC_N|$, then there exists a live marking of $N$.

*Proof:* (*Sketch.*)
The proof may be done by contraposition. Supposing that no live marking of $N$ exists, the column rank of $C$ is shown to be $\geq |CC_N|$. The proof proceeds in several steps, starting with a suitably chosen non-live marking $M_1$ and constructing exactly $|CC_N|$ linearly independent column vectors contained in $C$.

First, we consider an initial marking $M_1$ such that all places of all conflict clusters are marked with some token. By assumption, $M_1$ is not live. Using this, and also the strong connectedness obtained by Corollary 2, it may be shown that a firing sequence $M_1[\tau\rangle M_2$ exists, such that in $M_2$, every conflict cluster contains a transition with an unmarked input place. (This argument is invalid if there are arc weights greater than 1.) Linearly independent entries in $C$ can be then constructed as follows. The sequence $\tau$ is scanned backwards, such that for




| $C$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ |
|---|---|---|---|---|---|---|---|
| $s_1$ | 1 | −1 | −1 | 0 | 0 | 0 | 0 |
| $s_2$ | 1 | 0 | 0 | −1 | −1 | 0 | 0 |
| $s_3$ | 0 | 1 | 0 | 0 | 0 | −1 | 0 |
| $s_4$ | 0 | 0 | 1 | 0 | 0 | 0 | −1 |
| $s_5$ | 0 | 0 | 0 | 1 | 0 | −1 | 0 |
| $s_6$ | 0 | 0 | 0 | 0 | 1 | 0 | −1 |
| $s_7$ | −1 | 0 | 0 | 0 | 0 | 1 | 1 |

**Fig. 13.** A non-structurally-live net, its conflict clusters (l.h.s.), and its incidence matrix (r.h.s.)

every conflict cluster, the *last* transition in $\tau$ is recorded. It can be shown that the corresponding entries in $C$ are linearly independent, and since $\tau$ contains at least one transition from every conflict cluster, the number of transitions so obtained equals $|CC_N|$. ■

Consider, for example, the Petri net shown on the left-hand side of Figure 13 and its incidence matrix, shown on the right-hand side. The net is plain, and it satisfies both (PS) and (PT), the verification of which is left to the reader. It has 5 conflict clusters, also shown on the left-hand side of the figure. There exists no live marking for this net. The theorem claims that the rank of $C$ should be $\geq 5$, and indeed, it actually equals 5. For instance, the first five columns are linearly independent, while the remaining two columns can be linearly combined from them.

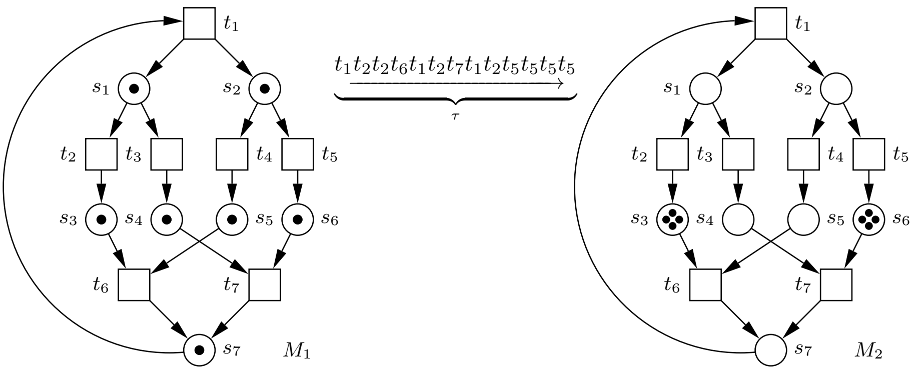
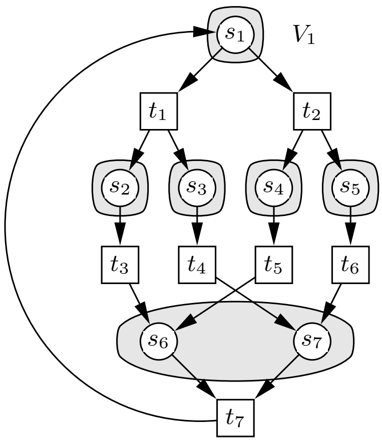


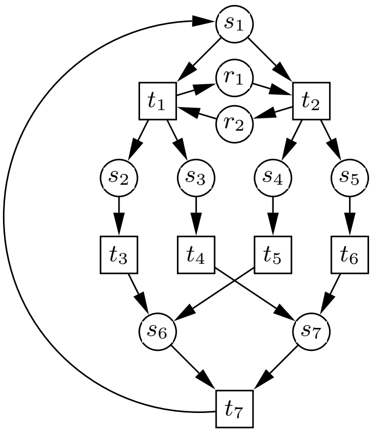

**Fig. 14.** $M_1\,[\tau\rangle M_2$, and in $M_2$, every conflict cluster contains a transition with an unmarked preplace




| $C$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ |
|---|---|---|---|---|---|---|---|
| $s_1$ | −1 | −1 | 0 | 0 | 0 | 0 | 1 |
| $s_2$ | 1 | 0 | −1 | 0 | 0 | 0 | 0 |
| $s_3$ | 1 | 0 | 0 | −1 | 0 | 0 | 0 |
| $s_4$ | 0 | 1 | 0 | 0 | −1 | 0 | 0 |
| $s_5$ | 0 | 1 | 0 | 0 | 0 | −1 | 0 |
| $s_6$ | 0 | 0 | 1 | 0 | 1 | 0 | −1 |
| $s_7$ | 0 | 0 | 0 | 1 | 0 | 1 | −1 |

**Fig. 15.** A structurally live net, its set of presets (l.h.s.), and its incidence matrix (r.h.s.)

To trace the constructions in the proof, we may consider a marking $M_1$ putting exactly one token on each place, such as shown on the left-hand side of Figure 14. This guarantees that every place in every conflict cluster has a token. A marking $M_2$, reachable from $M_1$, such that every transition in every conflict cluster has at least one unmarked input place is shown on the right-hand side of Figure 14, and a firing sequence $\tau$ with $M_1\,[\tau\rangle M_2$ is also shown. In the final step of the proof, the following transitions are recorded, in this order: $t_5$ (for $K_5$), then $t_2$ (for $K_4$), and then, similarly, $t_1, t_7, t_6$. The corresponding entries in $C$ are linearly independent.




| $C$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ |
|---|---|---|---|---|---|---|---|
| $s_1$ | −1 | −1 | 0 | 0 | 0 | 0 | 1 |
| $s_2$ | 1 | 0 | −1 | 0 | 0 | 0 | 0 |
| $s_3$ | 1 | 0 | 0 | −1 | 0 | 0 | 0 |
| $s_4$ | 0 | 1 | 0 | 0 | −1 | 0 | 0 |
| $s_5$ | 0 | 1 | 0 | 0 | 0 | −1 | 0 |
| $s_6$ | 0 | 0 | 1 | 0 | 1 | 0 | −1 |
| $s_7$ | 0 | 0 | 0 | 1 | 0 | 1 | −1 |
| $r_1$ | 1 | −1 | 0 | 0 | 0 | 0 | 0 |
| $r_2$ | −1 | 1 | 0 | 0 | 0 | 0 | 0 |

**Fig. 16.** The net of Figure 15 with a regulation circuit $\{t_1, r_1, t_2, r_2\}$ (l.h.s.), and its incidence matrix (r.h.s.)

**Theorem 3.** *Necessary condition for the existence of a live marking*
Assume that $N$ is a weakly connected, plain net covered by a positive S-invariant. If there exists a live marking of $N$, then the rank of $C$ is strictly less than $|PRESETS_N|$.

*Proof:* (*Sketch.*)
Let $N$ be a net which has a live marking and is covered by a positive S-invariant. By Proposition 3, there exists a positive T-invariant. Thus, $N$ satisfies (PT).

If the number $m = |PRESETS_N|$ is 0, then no transition has any input place. Since there are at least one transition and one place and the net is weakly connected, there is some transition without any input place but with some output place. Such a net cannot satisfy (PS). Hence $1 \leq m \leq |T|$, where $m = |T|$ in case no two transitions have a common preset. The theorem can be proved by induction on $|T| - m \geq 0$.

*Base:* Suppose $m = |T|$. Because $N$ satisfies (PT), the rank of $C$ is less than $|T| = |PRESETS_N|$.

*Step:* Suppose $m < |T|$. Then there exist at least two transitions with the same preset. Let $U$ with $|U| \geq 2$ be some set of transitions all of which have the same preset. (For instance, consider $U = \{t_1, t_2\}$ on the left-hand side of Figure 15.) Define $N[U]$ as $N$, augmented with a *regulation circuit* through the transitions of $U$ (such as shown on the left-hand side of Figure 16, cf. $r_1$ and $r_2$). The following can be observed:

1) $N[U]$ satisfies (PS) and is structurally live. To show structural liveness of $N[U]$, a live marking of $N$ can be augmented by sufficiently many tokens on the places of the regulation circuit; an upper limit for the number of such tokens can be derived from (PS).
2) $|PRESETS_{N[U]}| = |PRESETS_N| + |U| - 1 > |PRESETS_N|$ by properties of $N[U]$ and by $|U| \geq 2$.

Because of the inequality in 2), the induction hypothesis can be applied to $N[U]$, entailing

$$(\text{rank of } N[U]) \quad \leq \quad |PRESETS_N| + |U| - 2. \tag{3}$$

It can moreover be shown that (rank of $N$) $+ |U| - 1 \;\leq\;$ (rank of $N[U]$), which can be combined with (3), yielding (rank of $N$) $\leq |PRESETS_N| - 1$ and ending the inductive proof. ■

As an example, see Figure 15 which shows on its left-hand side the dual of the previous example. This net is also plain and satisfies (PS). It is structurally live as well; a live and bounded (even safe) marking of it has already been shown in Figure 7. The theorem claims that the number of presets should be larger than the rank of the incidence matrix. Indeed, the number of presets is 6, while the rank of the incidence matrix is 5, just as before.

To trace the inductive proof, we may consider the set $U = \{t_1, t_2\}$ in Figure 15. Then $N[U]$ is shown on the left-hand side of Figure 16, and its incidence matrix on the right-hand side of Figure 16 has rank 6. The inequalities claimed in the proof can thus be verified.

To see that weak connectedness is required as a precondition of this theorem, consider the net consisting of an isolated place, an isolated transition, and any marking. Such a net has a positive S-invariant and is live, but its rank is 0 and the number of presets is also 0, so the inequality claimed in the theorem fails to hold.

### 2.8 Bibliographical Remarks and Further Reading

The use of the incidence matrix, of S- and T-invariants, and of Dickson's and Farkas' lemma date back to early work in [KM69, CHEP71, GL73] and also to work by Kurt Lautenbach [Lau73]. Often in the literature, 'Dickson's lemma' denotes a statement which may be more general or slightly different from the one we used. Innocent as it might seem, Dickson's lemma can also be viewed as a (very restricted) special case of one of the most famous new results in graph theory, the *graph minor theorem*, cf. [Die10].

The connections between the rank of the incidence matrix and structural liveness were discovered for free-choice nets – to be defined in the next section – by a group around Manuel Silva in Zaragoza [CCS91]. In the context of free-choice Petri nets, these results are also contained – with improved proofs – in the textbook by Jörg Desel and Javier Esparza [DE95]. In this section, we have presented them independently of the free-choice property, and this presentation is due to Jörg Desel [Des92, Des98].

## 3 Graph-Theoretical Structure of Petri Nets

Graph-theoretically speaking, a Petri net is a *bipartite directed multigraph.* The term 'bipartite' refers to the fact that the set of nodes is divided into two disjoint sets, places and transitions, such that arcs connect nodes from one set with nodes of the other set, but never two nodes of the same set. The term 'multigraph' refers to the possible non-plainness of a Petri net, in the sense that there may be several arcs in the same direction between two nodes. In the remainder of this tutorial, we will neglect such multiplicities, however:

| |
|---|
| Henceforth, all Petri nets considered in definitions or in results, will be assumed to be *plain.* |

In the language of graph theory, this means that we consider *bipartite digraphs.* This section is devoted to exhibiting a few results by which the graph-theoretical structure of such a bipartite digraph (called a Petri net $N$) can be related to properties of the reachability graph of $N$.

In Petri net literature, one finds various constructions for turning non-plain Petri nets into 'behaviourally equivalent' plain ones. These constructions usually involve new places, new transitions and new tokens. In each individual case, however, it has to be checked whether the properties one is interested in are stable with respect to such transformations, or whether definitions and results one wishes to apply can be transferred easily from the plain to the non-plain case. For the definitions and results described in this and in the next section, such considerations can be non-trivial.

### 3.1 Some Simple Observations

Let us start our considerations with isolated places (i.e., places $s$ such that ${}^\bullet s = \emptyset = s^\bullet$) and isolated transitions (i.e., transitions $t$ such that ${}^\bullet t = \emptyset = t^\bullet$). An isolated place neither loses any of the tokens it has initially, nor does it gain any new tokens. Therefore, it impacts neither on liveness nor on boundedness properties, and we might as well (and will) exclude such places. An isolated transition can occur indefinitely often, neither needing any input tokens nor producing any output tokens. As such, it also has no effect on (strong or weak) liveness or on boundedness and we will exclude isolated transitions from consideration as well:

| Henceforth, all Petri nets considered will have no isolated places and no isolated transitions. |
|---|

Next, we consider places and transitions with mixed empty and non-empty pre- or postsets. A place $s$ with ${}^\bullet s \neq \emptyset = s^\bullet$ destroys either liveness or boundedness, because if the transitions in ${}^\bullet s$ are live, then $s$ is most certainly not bounded. A place $s$ with ${}^\bullet s = \emptyset \neq s^\bullet$ destroys liveness, because the transitions in $s^\bullet$ can fire at most as many times as there are tokens on $s$ initially. A transition $t$ with ${}^\bullet t \neq \emptyset = t^\bullet$ destroys either liveness or boundedness, because it is not live, unless unboundedly many tokens can be assembled on the places in ${}^\bullet t$. A transition $t$ with ${}^\bullet t = \emptyset \neq t^\bullet$ destroys boundedness, because it can fire indefinitely often in isolation, putting unboundedly many tokens on every place in $t^\bullet$.

These observations can, in fact, be extended to the following, which is a counterpart of Proposition 4:

**Proposition 5.** *Cycle-coveredness of well-formed nets*
Let $N$ be a well-formed plain Petri net. Then $N$ is covered by cycles.

*Proof:* (*Sketch.*)
Let $(u, v)$ be some arc in $N$.
If $u \in T$, then $v \in S$. If there was no path from $v$ to $u$, then liveness would allow the part of the net which does not depend on $v$ to fire sufficiently many times in order to put arbitrarily many tokens on $v$, contradicting boundedness.
If $u \in S$, then $v \in T$. If there was no path from $v$ to $u$, then the liveness of $v$ could only be guaranteed if arbitrarily many tokens could be assembled on $u$, again prompting a contradiction. ■

**Corollary 3.** *Strong connectedness of well-formed nets*
Let $N$ be a weakly connected, well-formed plain Petri net. Then $N$ is strongly connected. ■

Weak connectedness is not, usually, a strong requirement. For example, if a net is not weakly connected, one may analyse its weakly connected components separately. Sometimes, however, it is useful to consider non-weakly-connected substructures of an otherwise, weakly or strongly, connected net.

### 3.2 S-Nets, T-Nets, and Free-Choice Nets

Using the graph-theoretical structure of a Petri net, it is possible to define restrictive, but meaningful *Petri net classes*. Such classes can be useful in practice, but also in theory, since problems which are hard in general can sometimes be made more tractable by studying them in one of the restricted net classes first. We shall analyse problems such as liveness and boundedness for a range of structurally restricted classes of Petri nets.

S-nets forbid synchronisation and 'splitting' as shown on the left-hand side of Figure 17. T-nets prevent 'merging' and conflicts as shown on the right-hand side of Figure 17.

**Definition 17.** *S-nets and T-nets*
A plain net $N = (S, T, F)$ is called an *S-net* if $\forall t \in T\colon |{}^\bullet t| \leq 1 \geq |t^\bullet|$.
A plain marked net $N = (S, T, F, M_0)$ is an *S-system* if $(S, T, F)$ is an S-net.
A plain net $N = (S, T, F)$ is called a *T-net* if $\forall s \in S\colon |{}^\bullet s| \leq 1 \geq |s^\bullet|$.
A plain marked net $N = (S, T, F, M_0)$ is a *T-system* if $(S, T, F)$ is a T-net. ■

T-systems satisfy a basic token conservation property. For a marking $M$ and a place set $S' \subseteq S$, let

$$M(S') \;=\; \sum_{s \in S'} M(s),$$

and for a cycle (that is, a simple, closed path) $\gamma$, let

$$M(\gamma) = M(S'), \;\; \text{where } S' \text{ is the set of places on } \gamma.$$

We say that $S'$ is *token-empty* (*token-free*) or *marked*, depending on whether $M(S') = 0$ or $M(S') > 0$.

**Lemma 6.** *Elementary property of T-systems*
Let $N = (S, T, F, M_0)$ be a T-system and let $M \in [M_0\rangle$. For every cycle $\gamma$ of $(S, T, F)$, $M(\gamma) = M_0(\gamma)$.

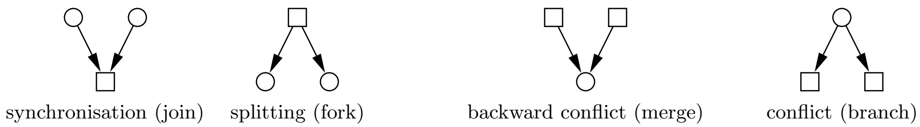



**Fig. 17.** Forbidden structures: S-nets (l.h.s.) and T-nets (r.h.s.)

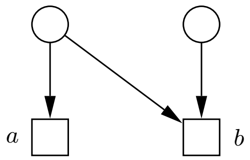



**Fig. 18.** Forbidden structure in FC-nets

*Proof:* Consider a cycle $\gamma$ and the effect of firing a transition $t$. If $t$ lies on $\gamma$, firing $t$ moves exactly one token on $\gamma$. If $t$ does not lie on $\gamma$, firing $t$ does not affect the tokens on $\gamma$. ∎

Note that whilst Definition 17 could be applied *verbatim* to non-plain nets, this is not necessarily meaningful. In particular, Lemma 6 would no longer be true.

Next, we will define a class of nets that encompasses both S-nets and T-nets called *free-choice nets* (or *FC-nets*, for short). FC-nets can be viewed as a 'smallest common generalisation' of S-nets and T-nets. In FC-nets, all structures shown in Figure 17 are allowed, but a combination of two of them, such as shown in Figure 18, is disallowed.

**Definition 18.** *Free-choice nets (FC-nets)*
A plain net $N = (S, T, F)$ is called an *FC-net* if

$$\forall t_1, t_2 \in T\colon {}^\bullet t_1 \cap {}^\bullet t_2 \neq \emptyset \Rightarrow {}^\bullet t_1 = {}^\bullet t_2. \tag{4}$$

A plain marked net $N = (S, T, F, M_0)$ is an *FC-system* if $(S, T, F)$ is an FC-net. ∎

The free-choice property is not satisfied in Figure 18, since ${}^\bullet a \cap {}^\bullet b \neq \emptyset$ and ${}^\bullet a \neq {}^\bullet b$. Originally, the class of free-choice nets was defined more restrictively. A (plain) net $(S, T, F)$ was called free-choice if

$$\forall t_1, t_2 \in T\colon {}^\bullet t_1 \cap {}^\bullet t_2 \neq \emptyset \;\Rightarrow\; |{}^\bullet t_1| = |{}^\bullet t_2| = 1. \tag{5}$$

The class of nets defined in 18 was originally called *extended free-choice*. Since most important properties and results either hold for both classes or are easily transferred from one to the other, we feel justified in ignoring this distinction for the time being. When we explicitly refer to the class defined by (5) (and this will occur only once, in section 3.7), then we speak of fc-nets rather than FC-nets.

Every free-choice net satisfies the following property which is symmetric to its defining property (4):

$$s_1{}^\bullet \cap s_2{}^\bullet \neq \emptyset \;\Rightarrow\; s_1{}^\bullet = s_2{}^\bullet \tag{6}$$

In fact, (6) is equivalent to (4) and could have been used as an alternative definition of the FC-property.

The nomenclature 'free choice' can be explained in the following way. Suppose that in a free-choice net, some marking $M$ activates a transition $t$. By (4), all transitions in the conflict cluster $({}^{\bullet}t)^{\bullet}$ of $t$ are activated, and one may freely choose between firing any of them.

Another interesting property of a free-choice net $N$ is that, unless there are transitions $t$ with ${}^{\bullet}t = \emptyset$, its conflict clusters $CC_N$ and its presets $PRESETS_N$ are in 1-1 correspondence with each other. For a conflict cluster $U \in CC_N$, the set ${}^{\bullet}U$ is well-defined because any two transitions in $U$ have the same presets by (4), and if it is nonempty, then it is a preset in $PRESETS_N$. Conversely, for any preset $R \in PRESETS_N$, the set $R^{\bullet}$ is well-defined by (6), and it is a conflict cluster. Moreover, $U = ({}^{\bullet}U)^{\bullet}$ and $R = {}^{\bullet}(R^{\bullet})$ for clusters $U$ with ${}^{\bullet}U \neq \emptyset$ and for presets $R$, which means that the correspondence is indeed one-to-one, unless there are transitions with empty presets.

If (PS) holds for an FC-net, or if it is well-formed, transitions $t$ with ${}^{\bullet}t = \emptyset$ are absent, and then the pleasant property $|CC_N| = |PRESETS_N|$ is valid. Theorems 2 and 3 can therefore be combined for FC-nets as follows:

**Corollary 4.** *Characterisation of the existence of a live marking in FC-nets*
Let $N$ be a weakly connected plain FC-net satisfying (PS) and (PT).
$N$ has a live marking if and only if its rank is strictly less than $|CC_N|$. ■

All S-nets and all T-nets are free-choice nets. Therefore, the last corollary also applies to such nets. The reader is encouraged to verify and simplify it separately for S-nets and for T-nets. However, the class of FC-nets is larger than the union of the classes of S-nets and T-nets. In the previous sections, several FC-nets which are neither S-nets nor T-nets were exhibited, such as, for example, in Figures 13 and 15.

### 3.3 A Liveness Criterion for FC-Systems

Corollary 4 gives a structural criterion for the *existence* of a live marking in an unmarked FC-net. Next, we characterise the circumstances under which a *given* marking is live in an FC-system. The characterisation uses two graph-theoretical structures that can meaningfully be defined for any Petri net, siphons and traps.

**Definition 19.** *Siphons and traps*
Let $N = (S, T, F)$ be a plain Petri net. A set $D \subseteq S$ is called *siphon* or *d-set* if ${}^{\bullet}D \subseteq D^{\bullet}$.
A set $Q \subseteq S$ is called *trap* or *t-set* if $Q^{\bullet} \subseteq {}^{\bullet}Q$. ■

According to this definition, the empty set $\emptyset \subseteq S$ is both a siphon and a trap. Moreover, it is not difficult to see that the union of two siphons (traps) is also a siphon (a trap, respectively). This property is, however, not valid for the intersection.

**Lemma 7.** *Elementary properties of siphons and traps*
Let $D$ be a siphon, let $M(D) = 0$ and let $M' \in [M\rangle$. Then $M'(D) = 0$.
Let $Q$ be a trap, let $M(Q) > 0$ and let $M' \in [M\rangle$. Then $M'(Q) > 0$.

*Proof:* Assume $M[t\rangle M'$ with $M(D)$=0 and $M'(D)$>0. Then necessarily $t \in {}^\bullet D$. By ${}^\bullet D \subseteq D^\bullet$, also $t \in D^\bullet$, contradicting $M[t\rangle \wedge M(D)$=0. Thus if $D$ is token-empty, $D$ remains token-empty.

Assume $M[t\rangle M'$ with $M(Q)$>0 and $M'(Q)$=0. Then necessarily $t \in Q^\bullet$. By $Q^\bullet \subseteq {}^\bullet Q$, also $t \in {}^\bullet Q$, contradicting $M[t\rangle \wedge M'(Q)$=0. Thus once $Q$ is marked, $Q$ remains marked. ■

This lemma can be applied in a special circumstance. Consider a net with initial marking $M_0$ which has some siphon $D$, and inside $D$, some trap $Q$ with $M_0(Q)$>0. By Lemma 7, such a siphon $D$ can never be completely emptied of tokens. It turns out that for FC-nets, liveness is already guaranteed if this condition holds for *every* siphon $D \neq \emptyset$:

**Theorem 4.** *The Commoner/Hack Criterion* CHC

Let $N = (S, T, F, M_0)$ be a free-choice system. The following two properties are equivalent:

| | | |
|---|---|---|
| **(i)** | For all siphons $D \subseteq S$ with $D \neq \emptyset$ there is a trap $Q \subseteq D$ such that $M_0(Q) > 0$ | } CHC |
| **(ii)** | $N$ is live | |

*Proof:* (*Sketch.*)
(i)⇒(ii) can be proved by contraposition. Suppose that $M_0$ is not live. Then there are $t \in T$ and $M \in [M_0\rangle$ such that $t$ is dead at $M$. By the FC property, it can easily be shown that there is a place $s \in {}^\bullet t$ which is token-empty at all markings reachable from $M$. Then every transition in ${}^\bullet s$ is also dead at $M$. By a backtracking (i.e., repeating this argument), a siphon which is token-empty at $M$ can be constructed. The siphon constructed by this algorithm cannot contain a trap which is marked at $M_0$, because such a trap could not have been emptied of tokens completely. That is, CHC fails to hold.

(ii)⇒(i) can be proved by contradiction. Suppose that $N = (S, T, F, M_0)$ is an FC-system which does *not* satisfy the Commoner/Hack Criterion CHC and, at the same time, is live. We deduce a contradiction.

Because of $\neg$ CHC, there exists a siphon $D \neq \emptyset$ which does not contain a trap marked at $M_0$. In particular, the (set-theoretically) largest trap $Q$ in $D$ is unmarked at $M_0$; note that $Q$ always exists because $\emptyset$ is a trap, and that it is unique because the union of two traps is again a trap. It may be the case that $q = \emptyset$.

For the contradiction, we wish to show that this particular siphon $D$ can be made completely free of tokens, because it is then that all of its output transitions (and there is at least one, due to the absence of isolated places) are dead, contradicting liveness. In this respect, the set $D \backslash Q$ is of critical importance, because it could contain tokens and it might be possible to move some of them onto $Q$. Once $Q$ has a token, the chance of obtaining token-emptiness of $D$ is obliterated. We need to show that starting from $M_0$, we can find some firing sequence which removes all tokens from $D \backslash Q$ without, at the same time, putting any tokens on $Q$. This is done by means of *allocations*.

An allocation is essentially a conflict resolution rule, picking exactly one transition out of a conflict cluster. If the transitions of the cluster are all enabled simultaneously, firing *according to an allocation* means that the allocated transition will be chosen, rather than any other. Now for the proof, it can be shown that there exists an allocation $\alpha$ which keeps removing tokens from $D \setminus Q$ without continually putting tokens back there, and also not onto $Q$. Firing according to $\alpha$ will eventually make $D \setminus Q$ token-free while keeping $Q$ token-free. Eventually, both $Q$ and $D \setminus Q$ are empty, and a token-free nonempty siphon with at least one output transition is obtained, contradicting the liveness of $N$. ∎

Note that Condition CHC mentions only the initial marking and the two graph-theoretical structures of trap and siphon. In particular, it does not refer to the reachability set $[M_0\rangle$ or to the reachability graph of $N$. When property CHC is tested algorithmically, it suffices to consider only the minimal siphons and in each of them, the maximal trap. Still, in the worst case there may be exponentially many minimal siphons.

The next examples demonstrate that the premise of free-choiceness cannot be omitted in any of the two directions of the liveness theorem. The left-hand side of Figure 19 presents a non-FC-system satisfying condition CHC but failing to be live. Note, that this system is deadlock-free, though, a property that holds for all systems satisfying CHC in general, even if they are non-FC-systems. The right-hand side of Figure 19 shows a non-FC-system which is live but fails to satisfy condition CHC. To see this, note that $\{s_1, s_2, s_3, s_4\}$ is a siphon which does not contain any marked trap. The free-choice property is violated at place $s_4$ and its output transitions.

The reader is invited to check what becomes of CHC in the special cases of S-systems. For T-systems, one has the following result:

**Theorem 5.** *Liveness and realisability of Parikh vectors in T-systems*
Let $N = (S, T, F, M_0)$ be a plain T-system. The following are equivalent:

a) $N$ is live;
b) all places $s \in S$ satisfy ${}^\bullet s \neq \emptyset$ and all (elementary) cycles carry at least one token under $M_0$;

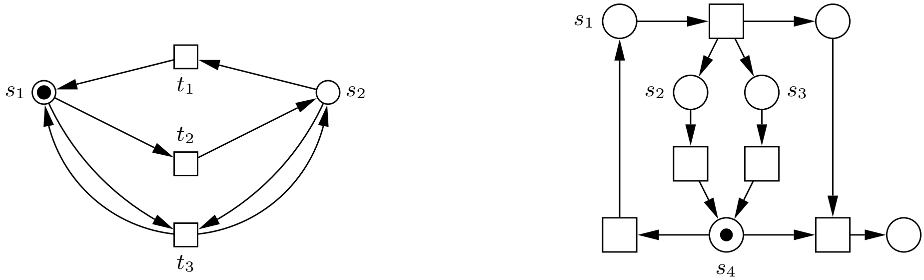



**Fig. 19.** Non-free-choice counterexamples to Theorem 4(⇒) and 4(⇐), respectively

c) all places $s \in S$ satisfy ${}^\bullet s \neq \emptyset$ and the Parikh vector 1 is *realisable*, that is, there is some firing sequence $\tau$ such that $M_0 [\tau\rangle M$ and every transition occurs exactly once in $\tau$.

*Proof:* (*Sketch.*)
In a T-system, minimal siphons are either singletons $\{s\}$ for a place $s \in S$ with ${}^\bullet s = \emptyset$, or elementary cycles. The former cannot contain any trap because of the absence of isolated places, and the maximal trap in an elementary cycle is the cycle itself. Thus condition b) is exactly what CHC reduces to for T-systems, and the equivalence between a) and b) turns out to be the counterpart of Theorem 4 for T-systems.

c)⇒b): If 1 is realisable, there can be no token-free cycles.

a)⇒c): If a place $s \in S$ satisfies ${}^\bullet s = \emptyset$, consider $t \in s^\bullet$. This transition exists due to the absence of isolated places. Then $t$ can fire at most $M_0(s)$ times, i.e., there is some reachable marking at which $t$ is dead.

The firability of 1 can be shown by induction on the number of transitions. If $T = \{t\}$, liveness implies that $t$ can be fired once from $M_0$. Suppose $|T| > 1$ and $t \in T$ such that $M_0[t\rangle$. Then $N$ can be transformed into another live T-system $N'$ by erasing $t$ and 'merging' input places and output places of $t$ in an appropriate way. By induction hypothesis, a suitable firing sequence $\tau'$ exists in $N'$. Then $\tau = t\tau'$ is a suitable firing sequence in $N$. ■

### 3.4 A Boundedness Criterion, and Some Coverability Results, for Live FC-Systems

In the previous section, an exact liveness criterion for FC-systems was described. In this section, this discussion is extended by presenting an exact characterisation of the boundedness of a live FC-net. We still assume all nets to be plain, and more graph-theoretical concepts are needed. In particular, we define two particular kinds of subnets.

**Definition 20.** *S-components and T-components*
Let $N = (S, T, F)$ be a plain net and let $N_1$ be the subnet $N(S_1, T_1)$ for some $S_1 \subseteq S$ and $T_1 \subseteq T$.

$N_1$ is called an *S-component* of $N$ if $T_1 = {}^\bullet S_1 \cup S_1^\bullet$ (taking the preset and the postset in $N$) and $\forall t \in T_1 \colon |{}^\bullet t \cap S_1| \leq 1 \geq |t^\bullet \cap S_1|$.

$N_1$ is called a *T-component* of $N$ if $S_1 = {}^\bullet T_1 \cup T_1^\bullet$ and $\forall s \in S_1 \colon |{}^\bullet s \cap T_1| \leq 1 \geq |s^\bullet \cap T_1|$.

$N_1$ is called *strongly connected* (inside $N$) if $N_1$ is strongly connected (as a separate net). ■

On the left-hand side of Figure 20, a net $N$ is shown. The right-hand side of the figure shows three of its subnets, $N_1$, $N_1'$ and $N_1''$. $N_1$ is a (non-strongly-connected) T-component but not an S-component. $N_1'$ is a strongly connected S-component, but not a T-component, since the property $S_1 = {}^\bullet T_1 \cup T_1^\bullet$, is

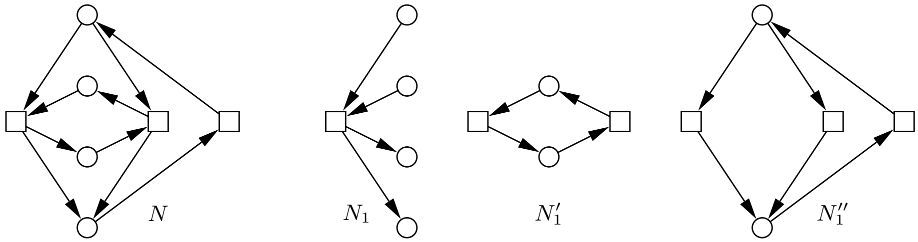



**Fig. 20.** Sample S- and T-components

violated in $N$. $N_1''$ is another strongly connected S-component. In $N$ there are no strongly connected T-components.

Let $N_1$, with place set $S_1$, be a strongly connected S-component of $(S, T, F)$. It is easy to verify that the S-vector having a 1 at places in $S_1$ and a 0 at places in $S \backslash S_1$ is an S-invariant. Similarly, every strongly connected T-component defines a binary T-invariant with entries in $\{0, 1\}$. In the following, unless specified otherwise, we consider only strongly connected S- and T-components.

**Theorem 6.** *Covering by S-components, and a boundedness criterion for live FC-systems*

Let $N = (S, T, F, M_0)$ be a (plain) live FC-system and let $s \in S$.

(1) A place $s$ is $m$-bounded ($m \in \mathbb{N}$, $m \geq 1$) if and only if there exists a strongly connected S-component $(S_1, T_1, F_1)$ with $s \in S_1$ and $M_0(S_1) \leq m$.

(2) There exists a marking $M \in [M_0\rangle$ satisfying $M(s) = m$ ($m \geq 1$) if and only if $M_0(S_1) \geq m$ is true for all strongly connected S-components $(S_1, T_1, F_1)$ with $s \in S_1$. ■

In both (1) and (2), one of the two directions is easy to prove, using the properties of S-components and their derived S-invariants. The nontrivial part of (1) states that the boundedness of $s$ entails the existence of an S-component covering $s$. The nontrivial part of (2) states that the least bound for the number of tokens on $s$ that can be derived from the S-components can actually be realised by some firing, that is, that there exists a reachable marking placing as many tokens on $s$ as are allowed by the S-components covering $s$.

As a corollary, it follows that a live FC-system is $m$-bounded if and only if it is covered by a set of strongly connected S-components with $m$ or less tokens. Consider, as an example, Figure 21. The initial marking shown on the left-hand side is live, and the system is also safe. According to the proposition, it should therefore be covered by strongly connected S-components carrying one token each. There exist two such S-components. One of them is shown in the middle of the figure, the other one is symmetrical.

The strongly connected S-components of a T-system are precisely its simple cycles. Hence a live T-system is $m$-bounded if and only if there exists a covering by simple cycles which carry $m$ or less tokens. For example, in Figure 22, places $s_1$ and $s_6$ are on an S-component with 2 tokens, and they are, moreover, not on

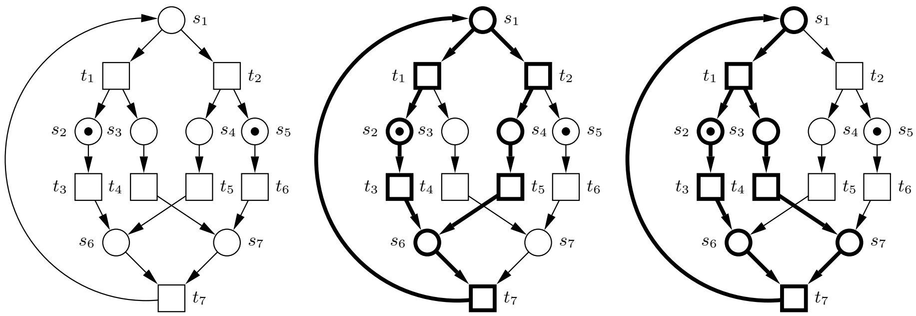



**Fig. 21.** An initially marked FC-net (l.h.s.); an S-component (middle); a T-component (r.h.s.)

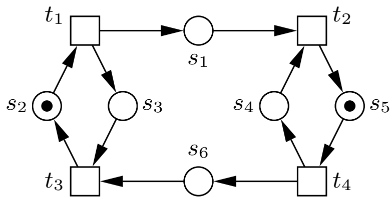



**Fig. 22.** A 2-bounded T-System

any other (strongly connected) S-component. Hence by Theorem 6 (part (2)$\Leftarrow$), there must be some firing sequence putting two tokens on $s_1$, and another firing sequence putting two tokens on $s_6$. Indeed, $t_1t_4t_3t_1$ results in two tokens on $s_1$, while $t_1t_4t_2t_4$ results in two tokens on $s_6$.

Well-formed FC-systems also satisfy a T-component covering property, as follows.

**Theorem 7.** *Covering by T-components*
A live and bounded FC-system $N$ is covered by strongly connected T-components. Moreover, for every strongly connected T-component $N_1$ in the cover, there exists a reachable marking $M$ such that $M$, restricted to $N_1$, is a live and bounded marking of $N_1$ (as a separate net). ■

As an example, consider Figure 21. The net is covered by two strongly connected T-components, and one of them is shown in bold on the right-hand side of the figure.

There are various ways in which Theorems 6 and 7 can be proved. One can show that in a well-formed FC-net, minimal semipositive S-invariants, minimal nonempty siphons and strongly connected S-components essentially agree with each other and that every place is contained in one of them. One can also make a connection between minimal semipositive T-invariants, minimal reproduction sequences and strongly connected T-components. Alternatively, one can prove one from the other theorem using the duality principle explained in the next section (which would then, in turn, need an independent proof).

## 3.5 Well-Formedness Criteria, the Duality Theorem, and Net Reductions

The coverability theorems of the previous section and Corollary 4 yield an exact condition for well-formedness, as follows.

**Corollary 5.** *Well-formedness criterion for FC-nets*
For a plain, weakly connected FC-net $N$, the following are equivalent:
(i) $N$ is well-formed
(ii) $N$ satisfies (PS) and (PT), and the rank of its incidence matrix is $\leq |CC_N| - 1$. ∎

This corollary directly leads to the following duality theorem. Let the *reverse* of a net $N$ be obtained by changing the directions of all arcs, the *dual* by exchanging places and transitions, and the *reverse-dual* by changing directions of all arcs as well as exchanging places and transitions. For example, consider the two nets shown in Figures 13 and 15 which are reproduced in Figure 23. These nets are duals and reverses of each other, and both are self-reverse-dual.

**Corollary 6.** *Duality theorem for FC-nets*
A plain, weakly connected net is a well-formed FC-net if and only if its reverse-dual is a well-formed FC-net.

*Proof:* The FC property and conditions (PS),(PT) are invariant with respect to reverse-duality. Moreover, the rank of $C$ (i.e. of $N$) equals the rank of $C^\mathsf{T}$ (i.e. of the reverse-dual of $N$), and the number of clusters is the same in $N$ and in its reverse-dual. The claim then follows with Corollary 5. ∎

An almost fully graph-oriented way of characterising well-formed FC-nets can be achieved by *net reductions*. Only three rules are needed. Suppose in the following that $(S, T, F)$ is a plain and weakly connected net.

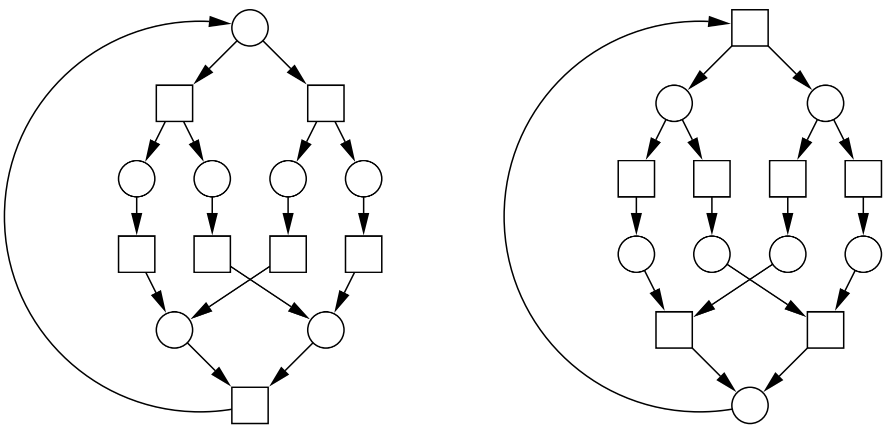


**Fig. 23.** The nets shown in Figure 15 (l.h.s.) and Figure 13 (r.h.s.)

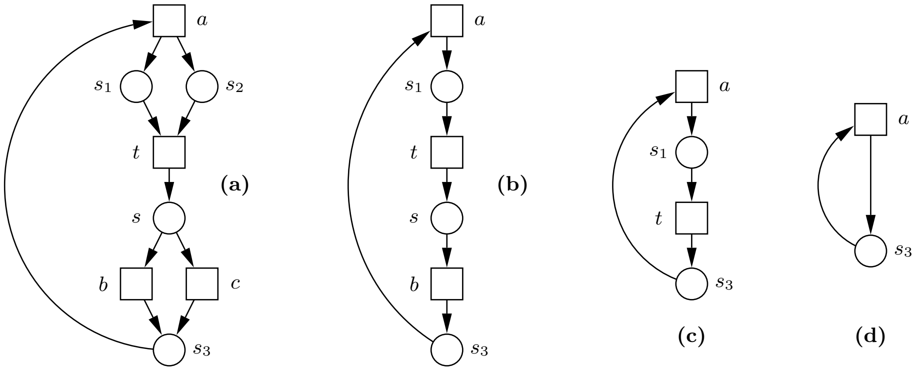



**Fig. 24.** A sample Petri net reduction

**ST-reduction:** Suppose $s \in S$ and $t \in T$ such that ${}^\bullet s \neq \emptyset$, $t^\bullet \neq \emptyset$, $s^\bullet = \{t\}$, ${}^\bullet t = \{s\}$, and $({}^\bullet s \times t^\bullet) \cap F = \emptyset$. Then omit $s$ and $t$ and all arrows around $s$ and $t$, while introducing a new arrow from every $u \in {}^\bullet s$ to every $r \in t^\bullet$.

**S-reduction:** Suppose a place $s$ is nonnegatively linearly dependent on a set of other places. Then omit $s$, along with all arrows around it.

**T-reduction:** Suppose a transition $t$ is nonnegatively linearly dependent on a set of other transitions. Then omit $t$, along with all arrows around it.

A simple example is shown in Figure 24. Rules are applied in this example as follows:

**(a) to (b):** S-reduction and T-reduction. The vector for place $s_2$ is $1\times$ the vector for place $s_1$, whence $s_2$ depends linearly and nonnegatively on $s_1$. Similarly, transition $c$ is $1\times$ transition $b$. (In this case, the two places and the two transitions actually duplicate each other, which amounts to a special case of the S- (and T-, respectively) reduction rule.)

**(b) to (c):** ST-reduction with $s$ and $b$

**(c) to (d):** ST-reduction with $s_1$ and $t$.

Note that at the end (in Figure 24(d)), a loop consisting of a single place and a single transition is obtained. Call this net the *loop net*.

**Theorem 8.** *Reduction theorem for FC-nets*
A plain, weakly connected FC-net is well-formed if and only if it can be reduced to the loop net by the three reduction rules defined above. ■

### 3.6 Home States in Free-Choice Nets

The initial marking in Figure 21 is live and safe, but cannot be reproduced by any nonempty firing sequence. As soon as one of the initially activated transitions $t_3$ or $t_6$ occur, the initial marking is no longer reachable. The property of being reachable from arbitrary reachable markings is called the *home state property*. In Figure 21, the initial marking is not a home state.

**Definition 21.** *Home state*
Let $N = (S, T, F, M_0)$ be a marked net. A marking $M \in [M_0\rangle$ is called *home state* or *home marking* if for all $M' \in [M_0\rangle$, $M \in [M'\rangle$ holds true. ■

There are various ways of convincing oneself that $M_0$, in Figure 21, is not a home state. One possibility is to construct the reachability graph (which is actually depicted on the right-hand side of Figure 7). This graph has a unique 'last' strongly connected component, but $M_0$ is not contained in it. In fact, *only* $M_0$ is not contained in it, so that all reachable markings except $M_0$ are home states.

Another possibility is to use a trap of the net, as follows. Consider in particular the trap $Q = \{s_1, s_3, s_4, s_6, s_7\}$ (cf. Figure 21). $Q$ is token-free initially. However, both $t_3$ and $t_6$ put a token on $Q$, and by the trap property, Lemma 7, $Q$ can never again become empty of tokens. Hence $M_0$ cannot possibly be a home state. In general:

> If there is a nonempty trap which is token-empty in some marking $M$ and the net is live, then $M$ cannot be a home state.

For live and bounded FC-nets, the converse is also true.

**Theorem 9.** *Trap theorem for FC-nets*
Let $N = (S, T, F, M_0)$ be a live and bounded FC-system.
$M_0$ is a home state if and only if all traps $Q \neq \emptyset$ satisfy $M_0(Q) > 0$.

*Proof:* (*Sketch.*)
Proving ($\Rightarrow$) is easy; the proof was already sketched.
($\Leftarrow$):

Note first that a marking may be live and safe even if every strongly connected T-component contains a token-free cycle, that is, even if *no* T-component is live when seen as a separate T-system. This is indeed the case in Figure 21. The T-component shown there has the token-free cycle $\{s_1, t_1, s_3, t_4, s_7, t_7\}$. The other strongly connected T-component also has a token-free cycle.

If a strongly connected T-component has no token-free cycle, we call it *activated*. Taken in isolation, an activated T-component is a live T-system, to which Theorem 5 applies. Inside an FC-system, transitions of an activated T-component can always be chosen by the free choice property in favour of others that would effect token loss on it. It is known that every marking of a strongly connected, live T-system is a home state. Therefore, if $t$ lies inside an activated T-component in a live FC-net, and if $M\,[t\rangle M'$, then $M \in [M'\rangle$; that is, the firing of $t$ can be reversed. This argument extends inductively to firing sequences. If, in a firing sequence $M_0\,[t_1\rangle M_1\,[t_2\rangle M_2 \ldots M_{n-1}\,[t_n\rangle M_n$, every transition $t_i$ is inside some activated T-component, then $M_0 \in [M_n\rangle$.

Now suppose that $M_0\,[t\rangle M$. We want to prove that $M_0 \in [M\rangle$.

If $t$ is inside some strongly connected T-component which is activated at $M_0$, then by the argument just given, $M_0 \in [M\rangle$, and we are done.

However, there might not be *any* activated T-components containing $t$. We show that nevertheless, $M_0$ can be reached from $M$ as follows:

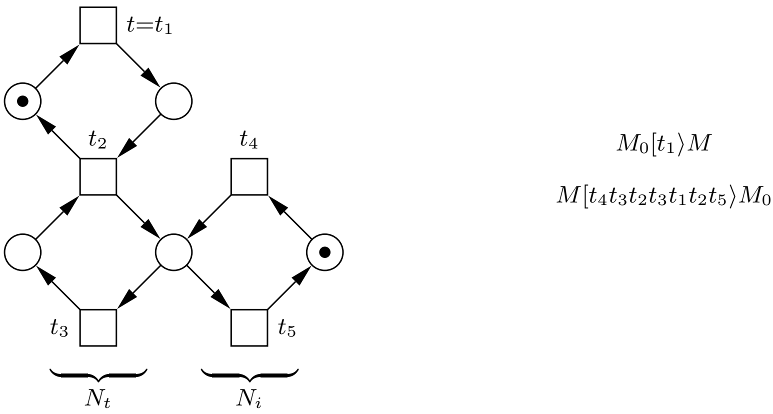



**Fig. 25.** An FC-system whose initial state is a home state; transition $t$ is not inside an activated T-component

- Start with *some* initially activated strongly connected T-component $N_i$. From the fact that all nonempty traps are marked initially, it can be proved that an initially activated, strongly connected T-component exists.
- Using the covering theorem, pick any T-component $N_t$ containing $t$.
- Now execute activated T-component(s) as much as possible, but without firing either $t$ or any other transition in its conflict cluster. Do this in such a way that tokens are 'moved towards' $N_t$. This can be achieved by a suitable allocation, as in the liveness theorem. Say, $M_0\,[\tau\rangle\widetilde{M_0}$ with a maximal sequence $\tau$ satisfying this property. It can be shown that this can be done in such a way that at $\widetilde{M_0}$, $N_t$ is activated and $t$ is still enabled. Thus $M_0\,[\tau t\rangle$ and also $M_0\,[t\tau\rangle$.
- Let $M_0\,[\tau\rangle\widetilde{M_0}\,[t\rangle\widetilde{M}$ and also $M_0\,[t\rangle M\,[\tau\rangle\widetilde{M}$. Then $\widetilde{M}\,[\tau'\rangle M_0$ with some sequence $\tau'$, because both $\tau$ and the subsequent firing of $t$ can be reversed (for they all take place within activated T-components). But then also $M_0\,[t\rangle M\,[\tau\rangle\widetilde{M}\,[\tau'\rangle M_0$, showing that $M_0 \in [M\rangle$. ■

An example explaining this proof is shown in Figure 25. Suppose $M_0[t_1\rangle M$. We want to show that $M_0$ can be reached from $M$. Transition $t = t_1$ is inside the T-component $N_t$ shown in the figure, but $N_t$ is not activated. However, there is another, initially activated, T-component $N_i$ whose transitions can be executed in a reversible way. A maximal sequence activating $N_t$ and not containing $t$ itself is $\tau = t_4t_3$ which can be fired from $M_0$ and also from $M$ (note that $t_3$ was chosen rather than $t_5$). The sequence $\tau'$ constructed in the last step of the proof is $\tau' = t_2t_3t_1t_2t_5$. Hence $M_0$ can be reached from $M$ by $t_4t_3t_2t_3t_1t_2t_5$. Note how $\tau'$ 'undoes' first $t$, by $t_2t_3$; then $t_3$, by $t_1t_2$; and then $t_4$, by $t_5$.

**Corollary 7.** *Confluence*

Let $N = (S, T, F, M_0)$ be a live and bounded FC-system and let $M_1, M_2 \in [M_0\rangle$.

Then $[M_1\rangle \cap [M_2\rangle \neq \emptyset$. ■

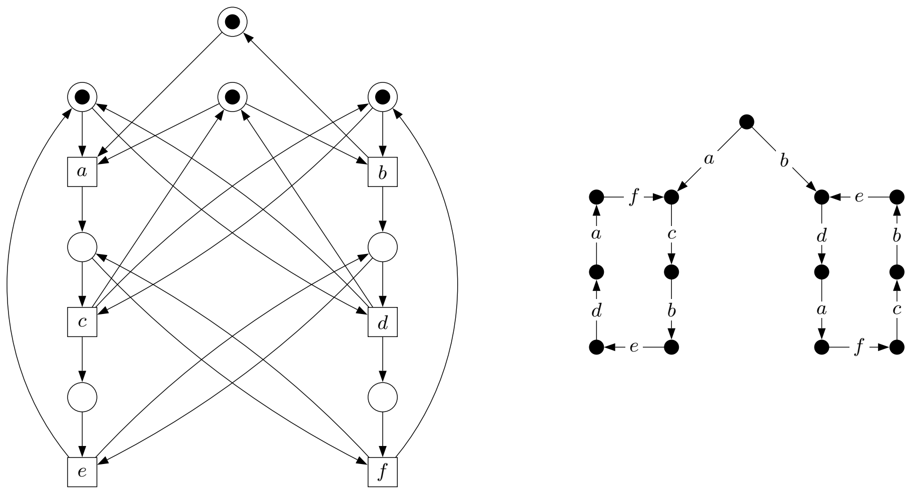



**Fig. 26.** A live and 2-bounded net without home states (l.h.s.); its reachability graph (r.h.s.)

**Corollary 8.** *Existence of home states*
Let $N = (S, T, F, M_0)$ be a live and bounded FC-system. There exists a home state $M \in [M_0\rangle$. ∎

A firing sequence containing every transition at least once necessarily generates a home state, since every trap $Q \neq \emptyset$ has at least one incoming transition. Such a firing sequence exists by liveness. The free-choice property is important. If it is omitted, we can find counterexamples such as the one shown in Figure 26.

The next result about blocking markings shows that it is in general possible to find home states, even without necessarily firing *all* transitions.

**Definition 22.** *Blocking marking*
Let $N$ be a plain net with initial marking $M_0$ and let $K \in CC_N$ be a conflict cluster. A *blocking marking for* $K$ is a marking $M_K \in [M_0\rangle$ such that every transition in $K$ is enabled by $M_K$ and no other transitions are enabled by $M_K$. ∎

**Theorem 10.** *Existence and uniqueness of blocking markings*
Let $N = (S, T, F, M_0)$ be a plain, live and bounded FC-net and let $K \in CC_N$. There exists a unique blocking marking $M_K \in \mathcal{E}(M_0)$ associated with $K$.

*Proof:* (*Sketch.*)
Consider the net which remains if the T-component shown in bold on the right-hand side of Figure 21 is erased. Note that it is an acyclic T-system with a unique *starting transition*, $t_2$ ('starting' is meant in the sense of the flow relation). It so happens that one can always find a minimal cover of $N$ with some strongly

connected T-components in such a way that taking away a carefully chosen one of them makes this is true in general.

Let $t_{in}$ be the starting transition of such a subnet. It can be shown that in any blocking marking for the cluster of $t_{in}$, there exists a token-free path from $t_{in}$ to any other transition in the subnet. If such a situation is given in some T-system, then it can be shown that the marking is unique (on the subnet). This yields a lever in order to prove the theorem (in particular, the uniqueness of the blocking marking) by induction on the number of T-components in a minimal strongly connected T-component covering of $N$. ■

Existence and uniqueness of a blocking marking implies that any such marking is a home state. Thus, in a weakly connected FC-system, a home state can be reached by the following procedure: (a) fix some enabled transition $t$ (then by the FC property, all transitions of the cluster $({}^\bullet t)^\bullet$ of $t$ are enabled); (b) fire transitions in $T \setminus ({}^\bullet t)^\bullet$ until it is no longer possible.

### 3.7 Realisability and Reachability Analysis

In section 2.1, it was emphasised that the marking equation or the marking inequality are necessary, but not sufficient for realisability or reachability. In the present section, this predicament will be discussed in more detail. It will be seen that nevertheless, under certain conditions, one may get some sort of converse of the firing lemma (Lemma 2).

Before starting the discussion, let us reconsider the non-structurally-live and non-structurally-bounded nets from section 1.7, reproduced here in Figure 27(a) and (b). Consider some elementary directed cycle in a Petri net. Any elementary directed path which starts at some place of the cycle and ends at some transition of the cycle but does not touch the cycle at any point in between is called a PT-*handle* (for this cycle). The notion of TP-*handle* is defined symmetrically. Note that there exist PT-handles in Figure 27(a) and TP-handles in 27(b).

Intuitively, a PT-handle is detrimental for liveness, because some tokens needed for firing the cycle's transitions could 'get lost' on it, such as in Figure 27(a). TP-handles, on the other hand, seem to be detrimental for boundedness, because

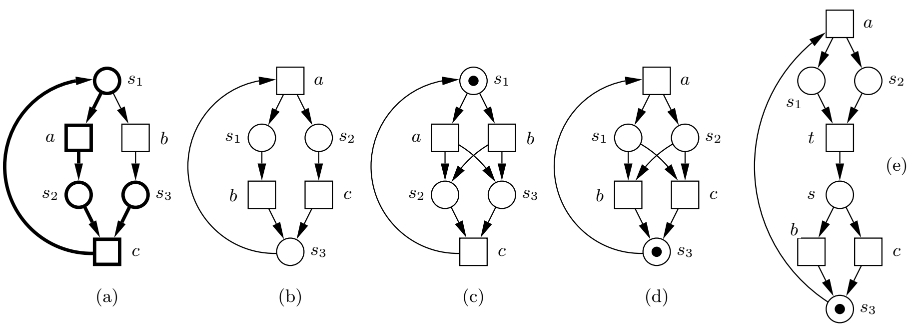



**Fig. 27.** Handle examples

tokens not needed for a cycle's liveness may be produced indefinitely, such as in Figure 27(b). However, one should be careful because not every PT-handle or TP-handle leads to non-liveness or to non-boundedness. The systems shown in Figure 27(c) and (d) contain handles, but they are perfectly live and bounded (even safe) FC-systems. Note that (c) is also an fc-system while (d) is not.

In order to state a realisability theorem, we need to define subnets generated by T-vectors. Let $y \in \mathbb{N}^T$ be any semipositive T-vector and consider the *support* of $y$, which is defined as $supp(y) = \{t \in T \mid y(t) > 0\}$. The *subnet generated by* $y$, $N_y$, is defined as the subnet $N(T_y, S_y)$ where $T_y = supp(y)$ and $S_y$ is the set of all places which are either input or output places of $supp(y)$. For example, the subnet generated by the T-vector $(2, 0, 1)$ is shown in bold in Figure 27(a).

**Theorem 11.** *First realisability criterion*

Let $N = (S, T, F)$ be a plain, pure net without PT-handles and let $y \in \mathbb{N}^T$ be a semipositive T-vector.

Then $y$ is realisable from a marking $M$ if and only if $M + C{\cdot}y \geq 0$ and $N_y$ has no token-free (under $M$, restricted to $N_y$) nonempty siphons.

*Proof:* (*Sketch.*)
The problematic direction is ($\Leftarrow$). This can be proved (a) for FC-nets where every place has at most two output transitions, then (b) for arbitrary FC-nets by reducing them to (a), and finally (c) for arbitrary nets by reducing them to FC-nets. ■

Part (c) of the proof relies on a construction replacing every arc from a place to a transition by a sequence arc-transition-arc-place-arc. Such a construction transforms every net into an FC-net; it works for the present purpose, but not for all purposes, as it may, e.g., introduce new deadlocks.

In Figure 27(a) with an initial token on $s_1$, the vector $(1\,1\,1)$ is not realisable even though it satisfies the marking inequality $M_0 + C{\cdot}y \geq 0$ and there are no (nonempty) token-free siphons. This shows that the premise of there not being any PT-handles is necessary for the theorem to hold. By duality (more precisely: considering the reverse net), the following theorem is a corollary:

**Theorem 12.** *Second realisability criterion*

Let $N = (S, T, F)$ be a plain, pure net without TP-handles and let $y \in \mathbb{N}^T$ be a semipositive T-vector.

Then $y$ is realisable from a marking $M$ if and only if $M + C{\cdot}y \geq 0$ and $N_y$ has no token-free (under $M + C{\cdot}y$, restricted to $N_y$) nonempty traps. ■

Both theorems can be turned into exact reachability criteria for nets without PT-handles or nets without TP-handles. Such a class is given by live and bounded fc-systems, since it is known that such nets do not have TP-handles. Consider Figure 27(d). It shows a live and bounded FC-system with TP-handles. A systematic transformation of it into an fc-system would just insert between a conflict cluster $U$ and its preset ${}^\bullet U$ a single transition followed by a single place, provided that $|{}^\bullet U| \geq 2$. The result for Figure 27(d) is shown in part (e) of the figure. The TP-handles have disappeared.

The reachability criterion for live and bounded fc-systems can be formulated as follows:

– Given: a live and bounded fc-system $(S, T, F, M_0)$ and a marking $M \in \mathbb{N}^S$.
– Solve, if possible, the following system of linear (in)equations for the unknown vector $y$:

$$\begin{array}{l} y \in \mathbb{N}^T \\ M_0 + C \cdot y \;=\; M \end{array}$$

under the further constraint that $N_y$ has no token-free trap under $M$.
– If this is possible, $M \in [M_0\rangle$, otherwise $M \notin [M_0\rangle$.

The correctness of this procedure follows from Theorem 12. The theorems are applicable to other classes of systems as well.

### 3.8 Bibliographical Remarks and Further Reading

T-systems have traditionally been called *marked graphs* [CHEP71] or *synchronisation graphs* [GL73]. The definitive book on the structure theory of free-choice systems is [DE95], by Jörg Desel and Javier Esparza, which contains several of the results and arguments described in this section, such as the Commoner/Hack liveness theorem [Hac72], the home state theorems [BV84, BDE92], the coverability and duality theorems, and the reduction theorem. Since the publication of this meritorious piece of work, further structural results in a similar vein have been discovered, e.g.: [Esp98] (NP-completeness of reachability in live and safe FC-systems); the blocking theorem described in section 3.6 [GHM03, Weh10]; general reachability criteria as described in section 3.7, which are due to Hideki Yamasaki, Jeng S. Huang and Tadao Murata [YHM01], with related results in [LR94, MM98, YY03]; and the proof, by Joachim Wehler [Weh09], of an old conjecture on a subclass of free-choice systems by Hartmann Genrich and P.S. Thiagarajan [GT84], making a connection to the almost equally old notion of frozen tokens [BM85]; not to mention many generalisations and extensions of these results, e.g. by the active research group around Manuel Silva [RTS98]. The literature also offers generalisations of definitions and results for arc-weighted T-systems [TCCS92] and for arc-weighted FC-systems [TS96].

## 4 Conflict Structure of Petri Nets

In a T-system, tokens cannot be removed from a place but by the – unique, if any – output transition of such a place. This implies what at the end of section 1.2 has somewhat loosely been called the absence of conflicts, or persistency. Symmetrically, in S-systems, transitions cannot be hindered from firing except by the – unique, if existing – input place of such a transition. This can loosely be called the absence of synchronisation, or communication-freeness.

Both notions, the absence of conflicts and the absence of synchronisation, give rise to a variety of structural constraints that partially overlap with those considered in the previous sections. For example, we might relax the notion of a T-system by requiring only $|s^\bullet| \leq 1$ or $|s^\bullet| = 1$, but not necessarily also $|^\bullet s| \leq 1$, for all $s \in S$. Such a class of nets might be called *place-output-nonbranching.* Symmetrically, we might relax the notion of an S-system by requiring only $|^\bullet t| \leq 1$ or $|^\bullet t| = 1$, not necessarily also $|t^\bullet| \leq 1$, for all $t \in T$. Such a class of nets might be called *transition-input-nonbranching.*

In the present section, we shall concentrate on restrictions related to persistency and to the absence of conflicts. Amongst others, we examine the class of place-output-nonbranching Petri nets, just called *output-nonbranching nets*, for short. Symmetrical restrictions related to the absence of synchronisation will not be examined in detail.

In the last part of the paper, a property known as *separability* is studied. This property indicates that a system can be viewed as a superimposition of independent subsystems, and it is a desirable feature in some applications. It turns out that in persistent systems, some (partly structural) conditions guaranteeing separability can be given.

| We continue to assume that every net is plain. |
|---|

### 4.1 A Hierarchy of Petri Nets without Conflicts

There is a surprising variety of classes of nets, all of which could (more or less) be called 'conflict-free'. By historical developments, the privilege of bearing the actual name, '*conflict-free nets*', has been bestowed onto one of these classes. The next definition introduces this class, as well as several related ones.

**Definition 23.** *Output-nonbranching, conflict-free, and persistent nets*
Let $N = (S, T, F, M_0)$ be a net with an initial marking.

- $N$ is called *output-nonbranching* (ON) if all places $s$ satisfy $|s^\bullet| \leq 1$.
- $N$ is called *conflict-free* (CF) if all places $s$ satisfy $|s^\bullet| > 1 \Rightarrow s^\bullet \subseteq {}^\bullet s$.
- $N$ is called *behaviourally conflict-free* (BCF) if for any two transitions $t, t' \in T$ with $t \neq t'$ and for every $M \in \mathcal{E}(M_0)$, if $M\,[t\rangle$ and $M\,[t'\rangle$ then ${}^\bullet t \cap {}^\bullet t' = \emptyset$.
- $N$ is called *binary-conflict-free* (BiCF) if for any two transitions $t, t' \in T$ with $t \neq t'$ and for every $M \in \mathcal{E}(M_0)$, if $M\,[t\rangle$ and $M\,[t'\rangle$ then $\forall s \in S\colon M(s) \geq F(s,t){+}F(s,t')$.
- A transition $t \in T$ is called *persistent*, if for every reachable marking $M \in \mathcal{E}(M_0)$, and for every transition $t' \in T$ with $t \neq t'$, if $M\,[t\rangle$ and $M\,[t'\rangle$ then $M\,[tt'\rangle$ and $M\,[t't\rangle$. $N$ is called *persistent* if every transition is persistent.
- A transition $t \in T$ is called *weakly persistent*, if for every reachable marking $M \in \mathcal{E}(M_0)$ and for every sequence $\sigma \in T^*$, if $M\,[t\rangle$ and $M\,[\sigma t\rangle$ then $M\,[t\sigma'\rangle$ for some permutation $\sigma'$ of $\sigma$. $N$ is called *weakly persistent* if every transition is weakly persistent. ■

Whether a net is output-nonbranching or conflict-free depends only on its structure. These properties can be checked without necessarily constructing the reachability graph. The other properties can be detected on the reachability graph.

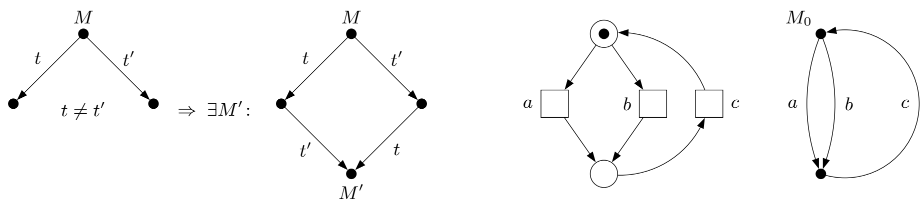



**Fig. 28.** Illustration of persistency (l.h.s.); a non-persistent net and its reachability graph (r.h.s.)

For instance, behavioural conflict-freeness can be checked as follows. Whenever a vertex is encountered from which two or more arcs labelled $t$ and $t'$, respectively, emanate, we check any pair of such arcs for the property ${}^\bullet t \cap {}^\bullet t' = \emptyset$, which can be read off the net. In order to check the persistency of a transition $t$, it is sufficient to check the property indicated in Figure 28, for every transition $t' \neq t$ and for every vertex $M$ in the reachability graph.

The BiCF condition, $\forall s \in S\colon M(s) \geq F(s,t)+F(s,t')$, indicates that $t$ and $t'$ are *concurrently enabled*. We shall therefore use the shorthand $M[\{t,t'\}\rangle$ in order to denote $\forall s \in S\colon M(s) \geq F(s,t)+F(s,t')$. The difference between BCF and BiCF (and persistency) can be seen on Figure 29.

From Definition 23, one gets the hierarchy shown in Figure 30. Thus, the class of T-systems is the smallest class under consideration while the class of weakly persistent systems is the largest class (i.e., all others lie inside). Actually, it is hard to call the class of weakly persistent systems ‘conflict-free’ since it contains systems that clearly exhibit conflicts in the intuitive sense, such as the one on the right-hand side of Figure 28. Nevertheless, as we will show in the next section, weakly persistent systems do enjoy some properties normally associated with persistent and conflict-free systems.

*Proof sketch* of the implications shown in Figure 30:
Every T-system is also ON: this follows directly from the definitions.

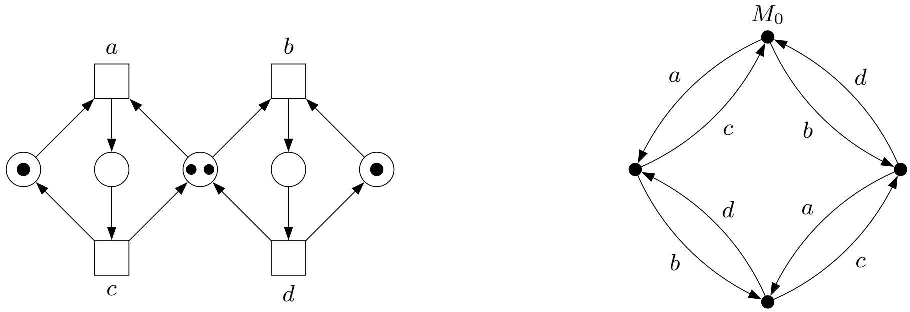



**Fig. 29.** A net which is persistent and BiCF but not BCF (l.h.s.), and its reachability graph (r.h.s.)

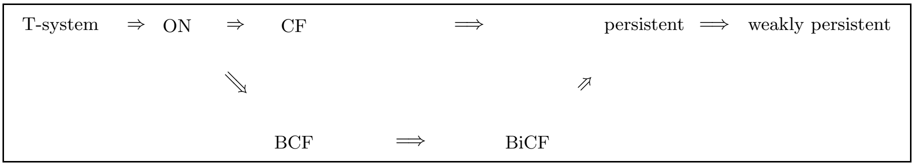



**Fig. 30.** A hierarchy of conflict-free and persistent Petri nets

Every ON system is CF: the CF condition is trivially true for ON systems.
Every CF system is persistent: if ${}^\bullet t \cap {}^\bullet t' \neq \emptyset$ and $t \neq t'$, then for any $s \in {}^\bullet t \cap {}^\bullet t'$, $|s^\bullet| > 1$. Then by the CF condition, both $t \in {}^\bullet s$ and $t' \in {}^\bullet s$, so that the occurrence of either of them cannot disable the other one. (Note that this argument is invalid if arc weights can be greater than 1.)
Every ON system is BCF: for ON systems, $t \neq t'$ already suffices to imply ${}^\bullet t \cap {}^\bullet t' = \emptyset$.
Every BCF system is BiCF: suppose that $M$ enables both $t$ and $t'$ with $t \neq t'$; by the BCF property, ${}^\bullet t \cap {}^\bullet t' = \emptyset$, and then, $M[\{t, t'\}\rangle$ is necessarily true because of $M[t\rangle$ and $M[t'\rangle$.
Every BiCF system is persistent: $M[\{t, t'\}\rangle$ implies that both $M[tt'\rangle$ and $M[t't\rangle$.
Every persistent system is weakly persistent: this follows from Keller's theorem, to be stated below. ■

Some of the implications of Figure 30 can be reversed under (relatively) weak conditions. They were described in the figure by short arrows. Implications indicated by long arrows cannot be reversed so nicely. To see this, we examine more closely the case that $M[tt'\rangle$ and $M[t't\rangle$ for some marking $M$ and two transitions $t \neq t'$. Because of its shape in the reachability graph, such a situation is called a *diamond*. A diamond comes in two varieties. If $t \neq t'$ and $M[\{t, t'\}\rangle$, then it is called a *concurrent diamond*. (Note that both $M[tt'\rangle$ and $M[t't\rangle$ are implied by $M[\{t, t'\}\rangle$). If $t \neq t'$ and $M[tt'\rangle$ and $M[t't\rangle$ but $\neg M[\{t, t'\}\rangle$, then the diamond is called a *conflicting diamond*. Figure 31 shows the difference. Self-loops are necessary for the existence of conflicting diamonds. If a Petri net is free of self-loops, all diamonds are concurrent.

The next proposition shows that properties ON and CF are essentially equivalent to each other, and both properties are also close to T-systems. Moreover, property BiCF is almost equivalent to persistency, except for the difference


**Fig. 31.** A concurrent diamond (l.h.s.) and a conflicting diamond (r.h.s.)

between the two types of diamonds just explained. In particular, in self-loop-free Petri nets, BiCF is the same as persistency.

**Proposition 6.** *Some relationships between classes of systems without conflicts*

1. For every CF net $N$ with initial marking $M_0$, an ON net $N'$ with initial marking $M_0'$ can be constructed such that the two reachability graphs are isomorphic.
2. Every live and bounded ON system is a T-System.
3. A net is BiCF if and only if it is persistent and there is no conflicting diamond.

*Proof:* To prove 1., consider an arbitrary place $s$ with $|s^\bullet| > 1$. By the CF property, $s^\bullet \subseteq {}^\bullet s$, that is, $s$ is a side condition for every output transition in $s^\bullet$. Replace $s$ by $|s^\bullet|$ new places which are marked and connected as $s$, except that a side condition connects it to only one (not all) of the $|s^\bullet|$ output transitions. The reachability graph of this new net is isomorphic to the old one. The construction can be repeated until there are no more places $s$ with $|s^\bullet| > 1$.

To prove 2., let $N = (S, T, F, M_0)$ be a live and bounded ON system. By Proposition 5, every weakly connected component of $N$ is strongly connected. Since $N$ is an ON net, it is also an FC-net. Thus it is covered by S-components by Theorem 6, and hence, also structurally bounded because it is covered by a positive S-invariant. By Proposition 1 and Farkas' lemma (Lemma 3), there is, therefore, no vector $y \in \mathbb{N}^{|T|}$ with $C \cdot y > 0$. Because $N$ is an ON net and is covered by cycles, $C \cdot y \geq 0$ where $y$ is the all-ones T-vector **1**. Suppose, for a contradiction, that $N$ is not a T-net. Then $C \cdot y \neq 0$, because there is at least one place with more than one input transition, contradicting the fact, just proved, that no such $y$ exists.

To prove 3.($\Rightarrow$), we have already seen that BiCF implies persistency. BiCF also implies the absence of conflicting diamonds: if there is such a diamond with $M$, $t$ and $t'$, then the BiCF property is violated with the same transitions at the same marking. To prove 3.($\Leftarrow$), assume $M[t\rangle$ and $M[t'\rangle$ with $t \neq t'$. By persistency, we get the diamond $M[tt'\rangle$ and $M[t't\rangle$. By the absence of conflicting diamonds, this is a concurrent diamond; thus $M[\{t, t'\}\rangle$. ■

It is perhaps illuminating to compare this Petri net hierarchy with the one defined in the previous section. T-systems (marked graphs) and ON-nets are both free-choice and persistent. But Proposition 6 notwithstanding, there exist CF nets in the sense of Definition 23 which are not FC in the sense of Definition 18, and conversely. The same is true for BCF nets. In fact, S-systems are free-choice but, in general, not even persistent. An example can be found on the right-hand side of Figure 28. Persistent nets are considerably less well-behaved than T-systems. For instance, even if they are strongly connected, there may be reproducing T-vectors which are different from multiples of **1** (as an example, see Figure 33 below). While in a T-system, every live marking is a home state, there exist live and bounded persistent systems whose initial marking is not a home state. The net shown in Figure 16 can be turned into an example by putting tokens on $s_2$, $s_5$ and $r_1$.

### 4.2 Weak Persistency and Semilinearity

Some times, the reachability set of a Petri net has a *semilinear* representation.

**Definition 24.** *Semilinear sets*
A set $W \subseteq \mathbb{N}^n$ is called *linear* if there are vectors $b, p_1, \ldots, p_\ell \in \mathbb{N}^n$ (with $\ell \in \mathbb{N}$) such that

$$W = \{b + \sum_{i=1}^{\ell} n_i \cdot p_i \mid n_i \in \mathbb{N}\}.$$

A set $W \subseteq \mathbb{N}^n$ is called *semilinear* if it is the finite union of linear sets. ■

The vectors named $b$ are the *bases*, and the vectors named $p$ are the *periods*.

For instance, the reachability set in Equation (1) of section 1.2 has a semilinear representation as follows:

$$\begin{array}{l} \{\,(0\ \ 1\ \ 2)^\mathsf{T}\,\} \cup \{\,(0\ \ 0\ \ 4)^\mathsf{T}\,\} \\ \cup\,\{\,(1\ \ 0\ \ 3)^\mathsf{T} + n_1\cdot(0\ \ 0\ \ 1)^\mathsf{T} \mid n_1 \in \mathbb{N}\,\} \cup \{\,(2\ \ 0\ \ 2)^\mathsf{T} + n_1\cdot(0\ \ 0\ \ 1)^\mathsf{T} \mid n_1 \in \mathbb{N}\,\} \\ \cup\,\{\,(3\ \ 0\ \ 1)^\mathsf{T} + n_1\cdot(0\ \ 0\ \ 1)^\mathsf{T} \mid n_1 \in \mathbb{N}\,\} \cup \{\,(4\ \ 0\ \ 0)^\mathsf{T} + n_1\cdot(1\ \ 0\ \ 0)^\mathsf{T} + n_2\cdot(0\ \ 0\ \ 1)^\mathsf{T} \mid n_1, n_2 \in \mathbb{N}\,\} \\ \cup\,\{\,(1\ \ 1\ \ 1)^\mathsf{T} + n_1\cdot(0\ \ 0\ \ 1)^\mathsf{T} \mid n_1 \in \mathbb{N}\,\} \cup \{\,(2\ \ 1\ \ 0)^\mathsf{T} + n_1\cdot(1\ \ 0\ \ 0)^\mathsf{T} + n_2\cdot(0\ \ 0\ \ 1)^\mathsf{T} \mid n_1, n_2 \in \mathbb{N}\,\} \end{array}$$

with eight bases and two periods combined in appropriate ways. In general, however, there is no such easy representation of the reachability set, since it is known that there exist nets whose reachability sets are not semilinear.

One of the seminal results about persistent nets (and later also weakly persistent nets) is that they always have semilinear reachability sets. Even more, weak persistency is decidable, and if the decision is positive, the semilinear reachability set can be constructed.

For the proof a simple conclusion needs to be drawn from the definition of weak persistency. If for some weakly persistent net $N = (S, T, F, M_0)$ there are $\sigma, \sigma' \in T^*$ with $M\,[\sigma\rangle$, $M\,[\sigma'\rangle$ and $\mathcal{P}(\sigma) \geq \mathcal{P}(\sigma')$ there is also some $\sigma''$ with $M\,[\sigma'\sigma''\rangle$ and $\mathcal{P}(\sigma) = \mathcal{P}(\sigma'\sigma'')$. This follows directly by induction over the length of $\sigma'$. For persistent nets this conclusion can be strengthened; this will be discussed in the next section.

**Theorem 13.** *Weak persistency is decidable*
Let $N = (S, T, F, M_0)$ be a Petri net. It is decidable if $N$ is weakly persistent. Furthermore, weakly persistent nets have semilinear reachability sets.

*Proof:* (*Sketch.*) Construct a set $EM$ of extended markings $(x, M) \in \mathbb{N}^T \times \mathbb{N}^S$ where $M_0\,[\sigma\rangle M$ with $\mathcal{P}(\sigma) = x$ holds. The construction of $EM$ starts with $EM = \{(0, M_0)\}$, then consecutively some $(x + 1\cdot t, M')$ is added to $EM$ if $(x, M) \in EM$

and $M\,[t\rangle M'$. Take some stage $k$ where $EM_k = \{(x_1, M_1), \ldots, (x_k, M_k)\} \subseteq EM$ has been computed so far. Define nonnegative difference sets $D_i$ for $1 \le i \le k$ by

$$D_i = \{(x_j - x_i, M_j - M_i) \mid 1 \le j \le k,\ M_j \ge M_i\}.$$

Then,

$$S_k = \bigcup_{1 \le i \le k} \Big\{ (x_i, M_i) + \sum_{d \in D_i} n_d d \;\Big|\; n_d \in \mathbb{N} \Big\}$$

is a semilinear set whose projection on the second component approaches $\mathcal{E}(M_0)$ from below with increasing $k$. Two decidable formulae can be built using $S_k$, formula $A$ checking if for all $(x, M) \in S_k$ with $M\,[t\rangle M'$ also $(x + 1{\cdot}t, M') \in S_k$. If so, $S_k$ is complete ($S_k = EM$) and its projection to the markings is the set $\mathcal{E}(M_0)$ which is also semilinear. Formula $B$ checks if for any pair $(x, M), (x', M') \in S_k$ with $x \le x'$ no transition $t \in T$ with $x(t) < x'(t)$ is enabled. This would contradict the above conclusion from weak persistency. If $N$ is weakly persistent it can be shown that the number of different sets $D_i$ is finite even for $k \to \infty$, i.e. for the complete set $EM = \lim_{k\to\infty} S_k$. The set of all $(x_i, M_i)$ with the same $D_i$ may be infinite, but the minimal elements of this set suffice when building $\lim_{k\to\infty} S_k$ and by Dickson's Lemma there are only finitely many of those. We can conclude $EM$ is semilinear then, so at some finite stage $k$ either formula $A$ or $B$ must hold, deciding if $N$ is weakly persistent or not. If $N$ is weakly persistent, its reachability set is the projection of the final $S_k$ to its second component. ■

Weakly persistent nets share with persistent nets the property of having semilinear reachability sets. However, they do not share the property of, intuitively, being 'nets without conflict'. Consider the net shown on the right-hand side of Figure 28. This net is weakly persistent, but it exhibits a conflict between $a$ and $b$ in its initial state.

### 4.3 Keller's Theorem

A seminal result about persistent Petri nets (which does not hold for weakly persistent nets in general) is based on the notion of the *residue* of a sequence $\tau$ of transitions with respect to another sequence $\sigma$, denoted by $\tau \dot{-} \sigma$. By definition, $\tau \dot{-} \sigma$ is what is left of $\tau$ after cancelling successively all symbols from $\sigma$ (if possible), read from left to right. Formally, $\tau \dot{-} \sigma$ can be defined by induction on the length of $\sigma$:

$$\begin{aligned}
\tau \dot{-} \varepsilon &= \tau \\
\tau \dot{-} t &= \begin{cases} \tau, \text{ if there is no transition } t \text{ in } \tau \\ \text{the sequence obtained by erasing the leftmost } t \text{ in } \tau, \text{ otherwise} \end{cases} \\
\tau \dot{-} (t\sigma) &= (\tau \dot{-} t) \dot{-} \sigma.
\end{aligned}$$

We now formalise that two firing sequences are permutations of each other from a marking. Two sequences $\sigma \in T^*$ and $\sigma' \in T^*$ are said to arise from each other by a *transposition from M* if both are activated at $M$ and if they are the same, except for the order of an adjacent pair of transitions, thus:

$$M[\sigma\rangle \quad \text{and} \quad M[\sigma'\rangle \quad \text{and} \quad \sigma = t_1 \ldots t_k t t' \ldots t_n \quad \text{and} \quad \sigma' = t_1 \ldots t_k t' t \ldots t_n.$$

Essentially, this means that $\sigma$ and $\sigma'$ are the same except for some (not necessarily concurrent) diamond reached after $t_1 \ldots t_k$. Two sequences $\sigma$ and $\sigma'$ are said to be *permutations of each other from M* (written $\sigma \equiv_M \sigma'$) if they are both activated at $M$ and if they arise out of each other through a sequence of transpositions from $M$.

**Theorem 14.** *Keller's theorem*

Let $(S, T, F, M_0)$ be a persistent Petri net. Let $\tau$ and $\sigma$ be two firing sequences activated at some reachable state $M \in \mathcal{E}(M_0)$. Then $\tau(\sigma \stackrel{\bullet}{-} \tau)$ and $\sigma(\tau \stackrel{\bullet}{-} \sigma)$ are also activated from $M$, and $\tau(\sigma \stackrel{\bullet}{-} \tau) \equiv_M \sigma(\tau \stackrel{\bullet}{-} \sigma)$. Furthermore, the marking reached after $\tau(\sigma \stackrel{\bullet}{-} \tau)$ equals the marking reached after $\sigma(\tau \stackrel{\bullet}{-} \sigma)$.

*Proof:* (*Sketch.*) By induction on the length of $\tau$. If $\tau = \varepsilon$, both $\tau(\sigma \stackrel{\bullet}{-} \tau)$ and $\sigma(\tau \stackrel{\bullet}{-} \sigma)$ are equal to $\sigma$, and the result follows directly from the premise that $\sigma$ is activated at $M$, the definition of $\equiv_M$, and persistency. If $\tau = t\tau'$, two cases can be distinguished: $\sigma$ does not contain $t$ or $\sigma$ contains $t$, i.e., $\mathcal{P}(\sigma)(t) = 0$ or $\mathcal{P}(\sigma)(t) > 0$, respectively. In either case, after some manipulations involving permutations of sequences in persistent nets, the induction hypothesis yields the desired result. ■

Note that part of this theorem is a confluence statement. That is, if $M_0\,[\sigma\rangle M$ and $M_0\,[\tau\rangle M'$, then $\mathcal{E}(M) \cap \mathcal{E}(M') \neq \emptyset$. The other part of the theorem asserts that Parikh vectors of sequences leading to a common successor marking of two reachable markings can actually be computed explicitly, using residues.

Using Keller's theorem, we can easily prove that persistency implies weak persistency. Suppose that $N$ is persistent and that some reachable marking $M$ enables both $t$ and $\sigma t$. By Keller's theorem, $M$ also enables $t((\sigma t) \stackrel{\bullet}{-} t)$. But $\sigma' = (\sigma t) \stackrel{\bullet}{-} t$ has the same Parikh vector as $\sigma$ by the definition of $\stackrel{\bullet}{-}$. Hence both $M[\sigma t\rangle$ and $M[t\sigma'\rangle$, with $\mathcal{P}(\sigma t) = \mathcal{P}(t\sigma')$. Again by Keller's theorem, $\sigma t \equiv_M t\sigma'$. Thus $N$ is also weakly persistent.

## 4.4 Cycle Decompositions, *k*-Nets, and Separability

In this section, several recent results about persistent Petri nets will be described. Theorem 14, in combination with other structural Petri net techniques, is used all over their proofs. We will not sketch these proofs, but illustrate the properties of the definitions and the statements of the results by means of examples.

For the next result, we introduce the concept of a *smallest cycle*. Consider Figure 29. The reachability graph shown on the right-hand side has a cycle $M_0[abcd\rangle M_0$ which is elementary in the sense of section 1.3. However, there is also a non-empty cycle $M_0[ac\rangle M_0$ which has a smaller Parikh vector. Note that both $\mathcal{P}(abcd)$ and $\mathcal{P}(ac)$ are T-invariants, by Lemma 5, showing that the former is not a minimal one. Inspired by this example, call a cycle in the reachability graph, starting at some marking $M$, *smallest* (around $M$) if there is no non-empty cycle around $M$ which has a smaller Parikh vector. Thus, every smallest cycle is elementary, but the converse need not be true. Smallest cycles correspond to minimal T-invariants.

**Theorem 15.** *Decomposing cycles of bounded, reversible, and persistent nets*
Let $N = (S, T, F, M_0)$ be a bounded, reversible, and persistent Petri net. There exists a finite set $\{X_1, \ldots, X_n\}$ of semipositive T-invariants such that they are transition-disjoint and every cycle $M[\rho\rangle M$ in the reachability graph of $N$ can be decomposed, up to permutations, to some sequence

$$M[\rho_1\rangle M[\rho_2\rangle M \ldots [\rho_n\rangle M$$

of cycles with all Parikh vectors $\mathcal{P}(\rho_i)$ in $\{X_1, \ldots, X_n\}$. Moreover, $\{X_1, \ldots, X_n\}$ can be chosen as the set of Parikh vectors of smallest cycles through any fixed reachable marking of $N$. ■

To appreciate the relevance of transition-disjointness, reconsider Figure 29. The reachability graph shown there has two transition-disjoint cycles, $(ac)^*$ and $(bd)^*$, which are executable from $M_0$. They correspond to two realisable, minimal, transition-disjoint T-invariants, $(1\,0\,1\,0)^\mathsf{T}$ and $(0\,1\,0\,1)^\mathsf{T}$. From every state and for each one of these T-invariants, a cycle can be executed which has it as a Parikh vector. By contrast, consider the right-hand side of Figure 28. The reachability graph also has two cycles, $(ac)^*$ and $(bc)^*$. However, they are not transition-disjoint, because $c$ belongs to both.

In essence, in a bounded, reversible, persistent Petri net, realisable minimal T-invariants describe 'independent' repetitive behaviours, and this observation can be extended: Suppose that $X_1, \ldots, X_n$ are as in the previous theorem. Then there are $n$ bounded, persistent and reversible nets $N_1, \ldots, N_n$, such that each net $N_i$ has exactly *one* minimal realisable T-invariant $X_i$ and the reachability graph of $N$ is isomorphic to the reachability graph of the place-disjoint union of the nets $N_1, \ldots, N_n$.

As a consequence, the case in which a persistent net has exactly one minimal realisable T-invariant $X$ is of special interest and needs to be scrutinized. It may still be the case that (unlike in a connected T-system, cf. Theorem 5) such a T-invariant is not a multiple of 1. However, there are special conditions under which this is indeed the case, called *k-multiply marked nets*, or $k$-nets for short.

Let $N$ be a net and let $k \geq 1$ be some positive integer number. For a marking $M$, the $k$-multiple marking $k{\cdot}M$ is defined by $(k{\cdot}M)(s) = k{\cdot}(M(s))$ for every place $s$. The net $k{\cdot}N$ is the same as the net $N$ except that the initial marking $k{\cdot}M_0$ replaces the initial marking $M_0$ of $N$. The net $k{\cdot}N$ is called a *k-net.* It turns out that initial $k$-markings $k{\cdot}M_0$ have particularly pleasant properties (partly generalising those of Theorem 5) provided that $k \geq 2$.

**Theorem 16.** *Smallest cycles in $k{\cdot}N$ have Parikh vector* 1 *if* $k \geq 2$
Let $k \geq 2$ and let $(N, k{\cdot}M_0)$ be a plain, bounded, reversible and persistent $k$-net with exactly one minimal realisable T-invariant $X$. Then $X \leq 1$ and for any transition $t$, $X(t) = 0$ if and only if $t$ is dead at $k{\cdot}M_0$. ■

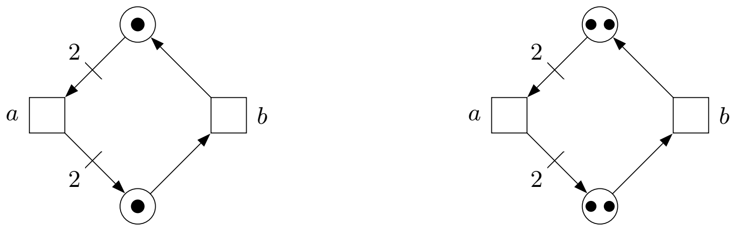



**Fig. 32.** A non-plain Petri net (l.h.s.) and its 2-multiple (r.h.s.) with minimal realisable T-invariant $(1\,2)^{\mathsf{T}}$

Plainness is important for Theorem 16 to hold. In Figure 32, all smallest cycles of the net on the right-hand side have Parikh vector $X = \mathcal{P}(abb)$, but $X > 1$, contrary to the conclusion of Theorem 16.

The statement made in Theorem 16 would not hold under the weaker assumption that $N$ instead of $k{\cdot}N$ is persistent. For instance, let $k = 2$ and consider Figure 33. On the left-hand side, $X = (a \mapsto 1, b \mapsto 1, c \mapsto 2) = (1\,1\,2)^{\mathsf{T}}$ is the unique minimal realisable T-invariant, and it can be realised by the firing sequence $M_0[acbc\rangle M_0$. Note that $X \not\leq 1$. On the right-hand side, $X$ is also the unique minimal realisable T-invariant, so that the conclusion of Theorem 16 is not true for this net. However, also one of the conditions of Theorem 16 is not satisfied, since the net is not persistent: executing $a$ in the initial marking leads to a marking in which both $a$ and $b$ are enabled although their shared input place $s$ carries only one token, hence producing a true conflict and destroying persistency. Thus, both requirements that $k{\cdot}N$ be persistent and that $k \geq 2$ are crucial for Theorem 16 to hold.

Next, we define an operation on transition sequences, called the *shuffle* or *arbitrary interleaving*. Intuitively, one may imagine some pack of cards to be divided into two halves and the second half be merged into the first. Instead of cards we may think of transitions, while the two half-packs correspond to sequences. Shuffling two sequences leaves the order of transitions stemming from

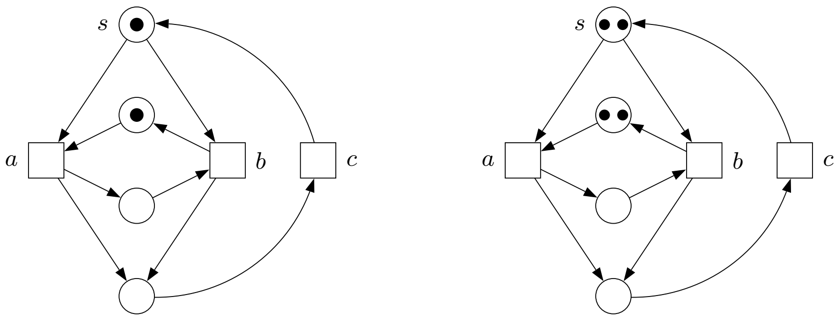



**Fig. 33.** A persistent Petri net (l.h.s.) and its non-persistent 2-multiple (r.h.s.)

one of the sequences unchanged. If two letters are not from the same sequence, however, we cannot predict their order in the resulting sequence.

Formally, the shuffle of two sequences $v$ and $w$ is a *set* of sequences written as $v \sqcup\!\sqcup w$. As an example, let $v = ab$ and $w = cbd$. Then

$$v \sqcup\!\sqcup w \quad = \quad \{abcbd,\ acbbd,\ acbdb,\ cabbd,\ cabdb,\ cbabd,\ cbadb,\ cbdab\}.$$

The shuffle can be extended canonically to sets of sequences. In general, it is associative and commutative.

Using shuffle, we define separability. This notion arises naturally in the context of $k$-nets.

**Definition 25.** *Weak and strong separability*
Let $k \geq 1$ and let $(N, k{\cdot}M)$ be any net with an initial $k$-marking $k{\cdot}M$.

A firing sequence $k{\cdot}M[\sigma\rangle$ is *weakly $k$-separable* from $k{\cdot}M$ (or just weakly separable if $k$ and $M$ are understood from the context) if there exist $k$ sequences $\sigma_1, \ldots, \sigma_k$ such that

$$(\forall j, 1{\leq}j{\leq}k\colon\ M[\sigma_j\rangle \text{ in } (N, M)) \quad \text{and} \quad (\sum_{j=1}^{k} \mathcal{P}(\sigma_j)) = \mathcal{P}(\sigma). \qquad (7)$$

A firing sequence $k{\cdot}M[\sigma\rangle$ is *strongly $k$-separable* from $k{\cdot}M$ if there exist $k$ sequences $\sigma_1, \ldots, \sigma_k$ such that

$$(\forall j, 1{\leq}j{\leq}k\colon\ M[\sigma_j\rangle \text{ in } (N, M)) \quad \text{and} \quad \sigma \ \in\ \sigma_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup \sigma_k. \qquad (8)$$

A $k$-net is weakly (strongly) separable if every sequence firable in its initial marking is weakly (strongly) separable from this $k$-marking. ■

Separability can be useful in verifying Petri nets. If some $k$-net $k{\cdot}N$ is separable, then it is sufficient to verify $N$ rather than $k{\cdot}N$ because, as a rule, properties of the latter can easily be deduced from properties of the former. This can increase efficiency considerably if $k$ is large, because the reachability graph of $k{\cdot}N$ could be much larger than that of $N$.

As an example, consider the two nets in Figure 34. On the left-hand side, a 2-marking is shown, where the set of tokens was split evenly into a hollow part and a solid part. The hollow tokens constitute $M_0$, and the solid tokens also constitute $M_0$. Thus, the whole marking is $2{\cdot}M_0$. Consider the firing sequence

$$2{\cdot}M_0\,[t_1t_2tt_1t_2\rangle.$$

This sequence can actually be fired in $M_0$, using only one of the two sorts of tokens, either just the hollow ones or just the solid ones. Hence $2{\cdot}M_0\,[\underbrace{t_1t_2tt_1t_2}_{\sigma}\rangle$ is strongly 2-separable by $M_0\,[\underbrace{t_1t_2tt_1t_2}_{\sigma_1}\rangle$ and $M_0\,[\underbrace{\ \varepsilon\ }_{\sigma_2}\rangle$. Consider the slightly longer firing sequence

$$2{\cdot}M_0\,[t_1t_2tt_1t_2t\rangle.$$

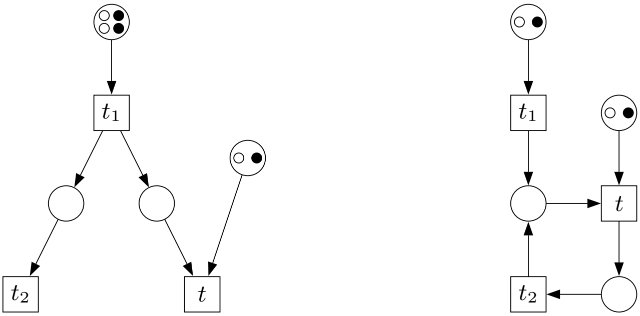



**Fig. 34.** Two 2-nets with hollow and solid tokens; separable (l.h.s.) and not separable (r.h.s.)

This sequence cannot be fired using only hollow tokens or only solid tokens, that is, we have $\neg\, M_0\,[t_1t_2tt_1t_2t\rangle$. If we try to 'prolong' the existing separation, we will fail, because $t$ alone is not firable from $M_0$. Nevertheless, $2{\cdot}M_0\,[t_1t_2tt_1t_2t\rangle$ can be strongly 2-separated by $M_0\,[t_1t_2t\rangle$ and $M_0\,[t_1t_2t\rangle$, since $t_1t_2tt_1t_2t \in (t_1t_2t \sqcup t_1t_2t)$. Intuitively, this corresponds to using hollow tokens for the first half and solid tokens for the second half of $t_1t_2tt_1t_2t$ (or the other way round).

Consider the net on the right-hand side of Figure 34. It again shows a 2-marking, and we have $2{\cdot}M_0\,[t_1tt_2t\rangle$. But no matter what kinds of tokens are used in order to fire $t_1tt_2$, the resulting marking activates $t$ with mixed types of tokens, both a hollow one and a solid one. Formally, there are no two sequences $\sigma_1$ and $\sigma_2$ satisfying $M_0\,[\sigma_1\rangle$ and $M_0\,[\sigma_2\rangle$ and $t_1tt_2t \in (\sigma_1 \sqcup \sigma_2)$. This shows that $t_1tt_2t$ is not strongly 2-separable. There are not even two sequences $\sigma_1$ and $\sigma_2$ satisfying $M_0\,[\sigma_1\rangle$ and $M_0\,[\sigma_2\rangle$ and $\mathcal{P}(\sigma_1) + \mathcal{P}(\sigma_2) = \mathcal{P}(t_1tt_2t)$. Thus, $t_1tt_2t$ is not even weakly separable.

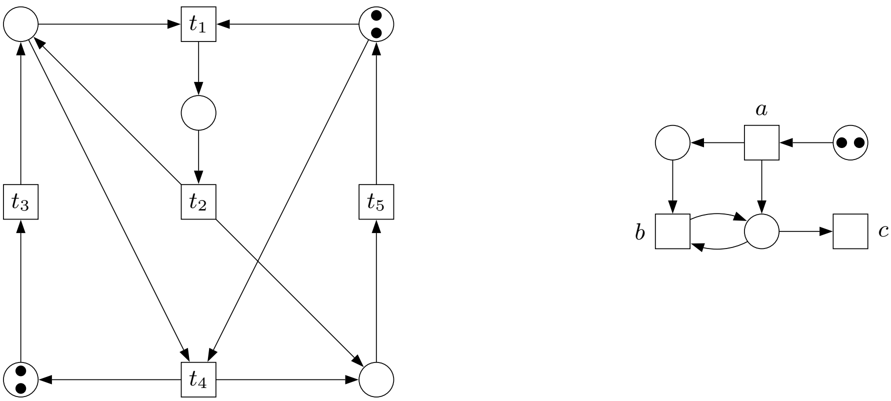



**Fig. 35.** Live, 2-bounded and not 2-separable (l.h.s.); weakly but not strongly separable (r.h.s.)

Figure 35(l.h.s.) depicts a 2-marking $2{\cdot}M$ which is live and bounded but not 2-separable. The reader is encouraged to find a non-separable firing sequence $2{\cdot}M\,[\sigma\rangle$. The 2-net shown on the right-hand side of Figure 35 is not strongly 2-separable from the indicated marking $2{\cdot}M$ since $2{\cdot}M[aacbbc\rangle$ cannot be obtained by shuffling two firing sequences from $M$. However, this 2-net is weakly 2-separable from $2{\cdot}M$. In particular, $\mathcal{P}(aacbbc) = \mathcal{P}(abc) + \mathcal{P}(abc)$, and clearly, $M[abc\rangle$. This 2-net is neither reversible nor persistent; e.g., $2{\cdot}M[acab\rangle$ and $2{\cdot}M[acac\rangle$ but $acacb$ cannot be fired from $2{\cdot}M$.

Separability cannot easily be checked directly on the reachability graph, because it is necessary to find a method for capturing all (possibly infinitely many) paths in the reachability graph of a $k$-marked net $(N, k{\cdot}M)$ and check them against $k$ paths of $k$ copies of the reachability graph of $(N, M)$.

Nevertheless, both weak and strong separability can be deduced for persistent Petri nets under some further premises.

**Theorem 17.** *Weak and strong separability*

Let $N$ be plain. Let $k \geq 1$ and let $k{\cdot}N$, with initial marking $k{\cdot}M_0$, be bounded, reversible, and persistent. If $k{\cdot}N$ has only one minimal realisable T-invariant, then $(N, k{\cdot}M_0)$ is weakly and strongly $k$-separable. ■

The proof of this result works roughly as follows. For $k = 1$, nothing has to be proved because every net is trivially 1-separable. For $k \geq 2$, Theorem 16 is exploited in an essential way. As a next step, weak separability is proved. Finally, in order to prove strong separability, the property of weak separability is used. All parts of this proof (except the case $k = 1$) are non-trivial.

Reversibility, plainness and persistency are important for Theorem 17 to hold. Figure 34 shows on the right-hand side a plain, bounded, non-reversible, persistent Petri net with a 2-marking $2{\cdot}M_0$ such that the firing sequence $2{\cdot}M_0[t_1tt_2t\rangle$ is not weakly 2-separable. The right-hand side of Figure 32 displays a non-plain, bounded, reversible, persistent 2-net with a 2-marking $2{\cdot}M_0$ in which the firing sequence $2{\cdot}M_0[a\rangle$ cannot be separated for obvious reasons. The net shown on the left-hand side of Figure 35 is not persistent but live, bounded, reversible and FC, showing that persistency cannot be omitted and that Theorem 17 does not hold for live and bounded FC-nets.

With the help of Theorem 15, Theorem 17 can be extended to bounded, reversible and persistent nets with several incomparable (mutually transition-disjoint) realisable T-invariants:

**Theorem 18.** *Strong separability for general bounded, reversible and persistent $k$-nets*

Let $N$ be plain. Let $k \geq 1$ and let $k{\cdot}N$, with initial marking $k{\cdot}M_0$, be bounded, reversible, and persistent. Then $(N, k{\cdot}M_0)$ is weakly and strongly separable. ■

## 4.5 Bibliographical Remarks and Further Reading

The class of (place-)output-nonbranching nets is intimately related to system classes also known as *context-free processes* [CHS95] or *basic process algebra*

(BPA) [BW90]. The class of transition-input-nonbranching nets, also known as *communication-free nets* [Esp97], is intimately related to system classes otherwise known as *basic parallel processes* (BPP) [CHS93]. The class of conflict-free Petri nets has been introduced in [LR78] and studied, amongst others, in [HR89]. The class of BiCF nets has been studied in [GGS11].

Sections 4.2 and 4.3 are based on [LR78, Gra80, Yam81, HI92] and on [Kel75]. The class of persistent Petri nets has been studied from different perspectives and extended in various ways; see, e.g., [BO09]. A net with non-semilinear reachability graph can be found in [LR78]. In the context of workflow systems, (weak) separability was introduced in [HSV03] for Petri nets, as follows:

> For business applications, separability is important because it formalises the idea of independent cases... If we associate to each firing the consumption of some resource, like money or energy, then separability implies that the consumption of a batch of cases equals the sum of the individual consumptions.

In the area of security kernels, a related concept has been known for some time, cf. the seminal paper [Rus82]. The results quoted in section 4.4 are based on [BDW07, BD09, BD11]. Figure 35(l.h.s.) is due to Karsten Wolf.

**Acknowledgements.** The authors are grateful to two reviewers for their comments.

## References


[BW90] Baeten, J.C.M., Weijland, W.P.: Process algebra. In: Cambridge Tracts in Theoretical Computer Science, vol. 18. Cambridge University Press (1990)

[BO09] Barylska, K., Ochmański, E.: Levels of Persistency in Place/Transition Systems. Fundamenta Informaticae 93, 33–43 (2009)

[BD09] Best, E., Darondeau, P.: A decomposition theorem for finite persistent transition systems. Acta Informatica 46, 237–254 (2009)

[BD11] Best, E.: P Darondeau: Separability in persistent Petri nets. Fundamenta Informaticae 112, 1–25 (2011)

[BDE92] Best, E., Desel, J., Esparza, J.: Traps characterize home states in free choice systems. Theoretical Computer Science 101, 161–176 (1992)

[BDW07] Best, E., Darondeau, P., Wimmel, H.: Making Petri nets safe and free of internal transitions. Fundamenta Informaticae 80, 75–90 (2007)

[BM85] Best, E., Merceron, A.: Frozen tokens and D-continuity: a study in relating system properties to process properties. In: Rozenberg, G. (ed.) APN 1984. LNCS, vol. 188, pp. 48–61. Springer, Heidelberg (1985)

[BV84] Best, E., Voss, K.: Free choice systems have home states. Acta Informatica 21, 89–100 (1984)

[CCS91] Campos, J., Chiola, G., Silva, M.: Properties and performance bounds for closed free choice synchronized monoclass queueing networks. IEEE Transactions on Automatic Control 36(12), 1368–1381 (1991)

[CHS93] Christensen, S., Hirshfeld, Y., Stirling, C.: Bisimulation equivalence is decidable for basic parallel processes. In: Best, E. (ed.) CONCUR 1993. LNCS, vol. 715, pp. 143–157. Springer, Heidelberg (1993)

[CHS95] Christensen, S., Hüttel, H., Stirling, C.: Bisimulation equivalence is decidable for all context-free processes. Information and Computation 121(2), 143–148 (1995)

[CHEP71] Commoner, F., Holt, A., Even, S., Pnueli, A.: Marked directed graphs. Journal of Computer and System Sciences 5, 511–523 (1971)

[Des92] Desel, J.: Struktur und Analyse von Free-Choice-Petrinetzen. In: Dissertation. Deutscher Universitätsverlag, Vieweg (1992)

[Des98] Desel, J.: Petrinetze, lineare Algebra und lineare Programmierung. Teubner-Texte zur Informatik, vol. 26. Band (1998)

[DE95] Desel, J., Esparza, J.: Free choice Petri nets. Cambridge Tracts in Theoretical Computer Science, vol. 40. Cambridge University Press (1995)

[Die10] Diestel, R.: Graph theory, `http://diestel-graph-theory.com/`

[Esp97] Esparza, J.: Petri nets, commutative context-free grammars, and basic parallel processes. Fundamenta Informaticae 31, 13–26 (1997)

[Esp98] Esparza, J.: Reachability in live and safe free-choice Petri nets is NP-complete. Theoretical Computer Science 198(1-2), 211–224 (1998)

[Far1902] Farkas, J.: Über die Theorie der einfachen Ungleichungen. Journal für die Reine und Angewandte Mathematik 124, 1–24 (1902)

[GHM03] Gaujal, B., Haar, S., Mairesse, J.: Blocking a transition in a free choice net and what it tells about its throughput. Journal of Computer and System Sciences 66, 515–548 (2003)

[GL73] Genrich, H.J., Lautenbach, K.: Synchronisationsgraphen. Acta Informatica 2, 143–161 (1973)

[GT84] Genrich, H.J., Thiagarajan, P.S.: A theory of bipolar synchronization schemes. Theoretical Computer Science 30, 241–318 (1984)

[GGS11] van Glabbeek, R.J., Goltz, U., Schicke, J.-W.: On causal semantics of Petri nets. In: Katoen, J.-P., König, B. (eds.) CONCUR 2011. LNCS, vol. 6901, pp. 43–59. Springer, Heidelberg (2011)

[Gor1873] Gordan, P.: Über die Auflösung linearer Gleichungen mit reellen Coefficienten. Mathematische Annalen 6, 23–28 (1873)

[Gra80] Grabowski, J.: The decidability of persistence for vector addition systems. Information Processing Letters 11, 20–23 (1980)

[HMW10] Habermehl, P., Meyer, R., Wimmel, H.: The downward-closure of Petri net languages. In: Abramsky, S., Gavoille, C., Kirchner, C., Meyer auf der Heide, F., Spirakis, P.G. (eds.) ICALP 2010. LNCS, vol. 6199, pp. 466–477. Springer, Heidelberg (2010)

[Hac72] Hack, M.H.T.: Analysis of production schemata by Petri nets. TR–94, MIT–MAC (1972); Corrections (1974)

[Hac74] Hack, M.H.T.: Decision problems for Petri nets and vector addition systems. Computation Structures Memo 95, Project MAC, MIT (1974)

[HSV03] van Hee, K., Sidorova, N., Voorhoeve, M.: Soundness and separability of workflow nets in the stepwise refinement approach. In: van der Aalst, W.M.P., Best, E. (eds.) ICATPN 2003. LNCS, vol. 2679, pp. 337–356. Springer, Heidelberg (2003)

[HI92] Hiraishi, K., Ichikawa, A.: On structural conditions for weak persistency and semilinearity of Petri nets. Theoretical Computer Science 93(2), 185–199 (1992)

[HR89] Howell, R.R., Rosier, L.E.: Problems concerning fairness and temporal logic for conflict-free Petri nets. Theoretical Computer Science 64(3), 305–329 (1989)

[KM69] Karp, R.M., Miller, R.E.: Parallel program schemata. Journal of Computer and System Sciences 3, 147–195 (1969)

[Kel75] Keller, R.M.: A fundamental theorem of asynchronous parallel computation. In: Tse-Yun, F. (ed.) Parallel Processing. LNCS, vol. 24, pp. 102–112. Springer, Heidelberg (1975)

[Kos82] Kosaraju, S.R.: Decidability of reachability in vector addition systems. In: Proceedings of the 14th Annual ACM STOC, pp. 267–281 (1982)

[Lam92] Lambert, J.L.: A structure to decide reachability in Petri nets. Theoretical Computer Science 99, 79–104 (1992)

[Lam77] Lamport, L.: Proving the correctness of multiprocess programs. IEEE Transactions on Software Engineering 3(2), 125–143 (1977)

[Lam] Lamport, L.: Personal communication (around 1995), `http://research.microsoft.com/en-us/um/people/lamport/pubs/pubs.html` entry 23 (Last accessed April 14, 2012)

[LR78] Landweber, L.H., Robertson, E.L.: Properties of conflict-free and persistent Petri nets. Journal of the ACM 25(3), 352–364 (1978)

[Lau73] Lautenbach, K.: Exakte Bedingungen der Lebendigkeit für eine Klasse von Petri-Netzen. GMD, Bericht Nr. 82 (Dissertation) (1973)

[LR94] Lautenbach, K., Ridder, H.: Liveness in bounded Petri nets which are covered by T-invariants. In: Valette, R. (ed.) ICATPN 1994. LNCS, vol. 815, pp. 358–375. Springer, Heidelberg (1994)

[Ler09] Leroux, J.: The general vector addition system reachability problem by Presburger inductive invariants. In: 24th IEEE Symposium on Logic in Computer Science (LICS 2009), Los Angeles, California, USA, August 11-14, pp. 4–13 (2009)

[MM98] Matsumoto, T., Mayano, Y.: Reachability criterion for Petri nets with known firing vectors. IEICE Trans. Fundamentals E81-A(4), 628–634 (1998)

[May80] Mayr, E.W.: Ein Algorithmus für das allgemeine Erreichbarkeitsproblem bei Petrinetzen und damit zusammenhängende Probleme. Technischer Bericht der Technischen Universität München TUM-I8010 (Dissertation). Institut für Informatik (1980)

[May84] Mayr, E.W.: An algorithm for the general Petri net reachability problem. SIAM Journal of Computing 13(3), 441–460 (1984)

[Mur89] Murata, T.: Petri nets: properties, analysis and applications. Proceedings of the IEEE 77(4), 541–580 (1989)

[Pet81] Peterson, J.L.: Petri net theory and the modeling of systems. Prentice-Hall (1981)

[Pet62] Petri, C.A.: Kommunikation mit Automaten. Dissertation, Institut für Instrumentelle Mathematik, Bonn, Schriften des IMM Nr. 2 (1962)

[PW03] Priese, L., Wimmel, H.: Petri-Netze, 376 pages. Springer (2003) ISBN: 3-540-44289-8

[RTS98] Recalde, L., Teruel, E., Silva, M.: On linear algebraic techniques for liveness analysis of P/T Systems. Journal of Circuits, Systems, and Computers 8(1), 223–265 (1998)

[Rei85] Reisig, W.: Petri nets. EATCS Monographs on Theoretical Computer Science, vol. 4. Springer (1985)

[Rus82] Rushby, J.: Proof of Separability – a Verification Technique for a Class of Security Kernels. In: Dezani-Ciancaglini, M., Montanari, U. (eds.) Programming 1982. LNCS, vol. 137, pp. 352–367. Springer, Heidelberg (1982)

[Schr86] Schrijver, A.: Theory of linear and integer programming. Wiley (1986)

[TCCS92] Teruel, E., Chrząstowski-Wachtel, P., Colom, J.M., Silva, M.: On weighted T-systems. In: Jensen, K. (ed.) ICATPN 1992. LNCS, vol. 616, Springer, Heidelberg (1992)

[TS96] Teruel, E., Silva, M.: Structure theory of equal conflict systems. Theoretical Computer Science 153(1-2), 271–300 (1996); Special volume on Petri nets

[Weh09] Wehler, J.: Free-choice Petri nets without frozen tokens, and bipolar synchronization systems. Fundamenta Informaticae 96, 1–38 (2009)

[Weh10] Wehler, J.: Simplified proof of the blocking theorem for free-choice Petri nets. Journal of Computer and System Sciences 76, 532–537 (2010)

[Wim04] Wimmel, H.: Infinity of intermediate states is decidable for petri nets. In: Cortadella, J., Reisig, W. (eds.) ICATPN 2004. LNCS, vol. 3099, pp. 426–434. Springer, Heidelberg (2004)

[Yam81] Yamasaki, H.: On weak persistency of Petri nets. Information Processing Letters 13(3), 94–97 (1981)

[YHM01] Yamasaki, H., Huang, J.S., Murata, T.: Reachability analysis of Petri nets via structural and behavioral classifications of transitions. Petri Net Newsletter 60, 4–21 (2001); Gesellschaft für Informatik

[YY03] Yen, H., Yu, L.: Petri nets with simple circuits. In: Warnow, T.J., Zhu, B. (eds.) COCOON 2003. LNCS, vol. 2697, pp. 149–158. Springer, Heidelberg (2003)