EXTENSIONS OF TORI IN SL(2)

JEFFREY ADAMS

Let \( \widetilde{SL}(2, \mathbb{F}) \) be the metaplectic two-fold cover of \( SL(2, \mathbb{F}) \), the special linear group in two variables over a local field \( \mathbb{F} \) of characteristic 0. The inverse image \( \widetilde{T} \) of a maximal torus \( T \) in \( SL(2, \mathbb{F}) \) is an abelian extension of \( T \) by \( \pm 1 \). We consider the question of whether this extension is trivial. More generally we find the minimal subgroup \( A \) of the circle for which the extension is split when considered with coefficients in \( A \). We see that \( |A| = 2, 4 \) or 8 in the \( p \)-adic case. We also find an explicit splitting function for the cocycle.

Introduction

Let \( \widetilde{SL}(2, \mathbb{F}) \) be the metaplectic two-fold cover of \( SL(2, \mathbb{F}) \), the special linear group in two variables over a local field \( \mathbb{F} \) of characteristic 0. The inverse image \( \widetilde{T} \) of a maximal torus \( T \) in \( SL(2, \mathbb{F}) \) is an abelian extension of \( T \) by \( \pm 1 \). We consider the question of whether this extension is trivial. We exclude the case \( \mathbb{F} = \mathbb{C} \), which is trivial.

More generally suppose \( A \) is a subgroup of the circle \( \mathbb{T} \) containing \( \pm 1 \). The inclusion of \( \pm 1 \) in \( A \) induces a map on cohomology, and defines an extension

\[ 1 \to A \to T_A \to T \to 1. \]

We say \( A \) is a splitting group for \( \widetilde{T} \) if the extension \( T_A \to T \) splits. It is well-known that \( \mathbb{T} \) is a splitting group. We say a splitting group \( A \) is a minimal splitting group if no proper subgroup of \( A \) is a splitting group. It is easy to see the order of a minimal splitting group is a power of 2, and hence unique, if it is finite.

Let \((, )_\mathbb{F}\) be the Hilbert symbol of \( \mathbb{F} \), and let \( \mu_n \) be the \( n \)th roots of unity in \( \mathbb{C} \).

**Theorem 1.** The minimal splitting group \( A_{\text{min}} \) for \( T \) is given by:

(a) Suppose \( T \simeq \mathbb{F}^* \). Then

\[
A_{\text{min}} = \begin{cases} 
\mu_2 & (\mathbb{F}^-)^* = 1 \\
\mu_4 & (\mathbb{F}^-)^* = -1.
\end{cases}
\]

257
(b) Suppose $T \simeq \mathbb{E}^1$ for $\mathbb{E}$ a quadratic extension of $\mathbb{F}$. Then

$$A_{\text{min}} = \begin{cases} 
\mu_2 & (-1,-1)_{\mathbb{F}} = 1 \\
\mu_4 & (-1,-1)_{\mathbb{F}} = -1, \text{ $\mathbb{F}$ non-archimedean}, \; -1 \notin \mathbb{E}^{\ast 2} \\
\mu_8 & (-1,-1)_{\mathbb{F}} = -1, \text{ $\mathbb{F}$ non-archimedean}, \; -1 \in \mathbb{E}^{\ast 2} \\
T & \mathbb{F} = \mathbb{R}.
\end{cases}$$

**Remark 2.** It is well-known that $(-1,-1)_{\mathbb{F}} = 1$ unless $\mathbb{F} = \mathbb{R}, \mathbb{Q}_2$, or an extension of $\mathbb{Q}_2$ of odd degree.

Theorem 1 is proved in Sections 3, 4 and 5. Here is an alternative realization of $\tilde{T}$. A character of $\tilde{T}$ is said to be genuine if it does not factor to $T$.

**Theorem 3.** Let $\tau(z) = z^2$ ($z \in \mathbb{C}^*$). Let $\tilde{\alpha}$ be a genuine character of $\tilde{T}$. Then $\tilde{\alpha}^2$ factors to a character $\alpha$ of $T$, and $\tilde{T}$ is isomorphic to the pullback of $\tau$ via $\alpha$. In other words $\tilde{T}$ is isomorphic to the $\sqrt{\alpha}$-extension of $G$.

From this we obtain an interpretation of the minimal splitting group of Theorem 1. Let $n(\tilde{T})$ be the minimal order of a genuine character of $\tilde{T}$. Set $\mu_{\infty} = T$.

**Corollary 4.** The minimal splitting group for $\tilde{T}$ is $\mu_{n(\tilde{T})}$.

For the proofs of Theorem 3 and Corollary 4 see Lemma 1.4.

We also give an explicit splitting of this extension, i.e., a function $\zeta : T \to A_{\text{min}}$ whose coboundary is the cocycle defining $\tilde{T}$ (see §3 and Theorem 5.7).

These questions arise from the theory of the oscillator representation and dual pairs. The splitting plays a role in this context, for example see [11]. The case of $\mathbb{F}^*$ is well-known ([4], p. 42, attributed to J. Klose), as is the existence of a $\mathbb{T}$-splitting in general [2]. General results about the splitting of the metaplectic cover over subgroups are due to Kudla [7], and a splitting of the extension of an elliptic torus is found in [7], Proposition 4.8 (in the non-archimedean case it is easy to see this can be taken to be a $\mu_8$-splitting).

This paper grew out of an effort to simplify Kudla’s formula. In the case of a p-adic field of odd residual characteristic a formula for a $\mu_2$-splitting in some cases may be deduced from [6], cf. ([4], p. 43).

Many of the arguments, especially those of Section 1 apply to other abelian extensions of abelian groups, for example a maximal torus in the two-fold cover of $Sp(2n, \mathbb{F})$. If $\tilde{G}$ is a non-linear $n$-fold cover of the $\mathbb{F}$ points of an algebraic group $G$, then the inverse image $\tilde{T}$ of a maximal torus in $G$ is typically not abelian. However similar arguments apply to the center of $\tilde{T}$.

Throughout $\mathbb{F}$ denotes a local field of characteristic zero, and $(x, y)_{\mathbb{F}} \in \mu_2$ is the Hilbert symbol. For $x \in \mathbb{F}^*$ and $\psi_{\mathbb{F}}$ a non-trivial additive character of
\( \mathbb{F} \), \( \gamma_{\mathbb{F}}(x, \psi_{\mathbb{F}}) \in \mu_4 \) is the Weil index. We use basic properties of the Hilbert symbol and the Weil index without further comment, see ([10], Appendix) for details. We make repeated use of the identities

\begin{align*}
(1) \quad & \gamma_{\mathbb{F}}(x, \psi_{\mathbb{F}}) \gamma_{\mathbb{F}}(y, \psi_{\mathbb{F}}) = (x, y)_{\mathbb{F}} \gamma_{\mathbb{F}}(xy, \psi_{\mathbb{F}}) \\
(2) \quad & \gamma_{\mathbb{F}}(x, \psi_{\mathbb{F}})^2 = (-1, x)_{\mathbb{F}}.
\end{align*}

If \( \mathbb{E} \) is a quadratic extension of \( \mathbb{F} \) then

\begin{align*}
(3) \quad & (x, z)_{\mathbb{E}} = (x, Nz)_{\mathbb{F}} \quad (x \in \mathbb{F}^*, z \in \mathbb{E}^*) \\
(4) \quad & (x, y)_{\mathbb{E}} = 1 \quad (x, y \in \mathbb{F}^*) \\
(5) \quad & \gamma_{\mathbb{E}}(x, \psi_{\mathbb{E}}) \gamma_{\mathbb{E}}(y, \psi_{\mathbb{E}}) = \gamma_{\mathbb{F}}(xy, \psi_{\mathbb{F}}) \quad (x, y \in \mathbb{F}^*).
\end{align*}

I would like to thank Steve Kudla and Jonathan Rosenberg for many useful discussions, and Jim Schafer for assistance with the arguments in Section 1.

### 1. Abstract Groups.

In this section we ignore the topology on \( T \) and consider it as an abstract group. We recall some standard facts from group cohomology and establish some notation. For example see [1].

Suppose \( G \) is a group, \( A \) is an abelian group, and \( G \) acts trivially on \( A \). The equivalence classes of central extensions of \( G \) by \( A \) are parametrized by the group cohomology \( H^2(G, A) \). Given an extension \( p : H \to G \) let \( s : G \to H \) be a section, i.e., \( p \circ s = 1 \). The cohomology class of the extension is represented by the 2-cocycle \( c_s(g, h) = s(gh)s(h)^{-1}s(g)^{-1} \). When there is no danger of confusion we do not distinguish between \( c_s \) and its image \( c_s \) in \( H^2(G, A) \). Any other such splitting \( s' \) is given by \( s'(g) = s(g)\zeta(g) \) for some map \( \zeta : G \to A \), and then \( c_{s'}(g, h) = c(g, h)\zeta(gh)\zeta(h)^{-1}\zeta(g)^{-1} \). Thus \( c_{s'} = c_s d\zeta, \) and \( \tau_s = \tau_{s'} \).

Conversely given a cocycle \( c \) we define \( H \) to be equal to \( G \times A \) as a set, with multiplication \((g, a)(g', a') = (gg, aa'c(g, g'))\). The cocycle \( c \) is trivial in cohomology if and only if

\begin{equation}
(6) \quad c(g, h) = \zeta(g)\zeta(h)\zeta(gh)^{-1}
\end{equation}

for some \( \zeta \), i.e., \( d\zeta = c \). We say \( \zeta \) is a splitting of the cocycle. Equivalently the splitting map \( s(g) = (g, \zeta^{-1}(g)) \) is a homomorphism. Any other splitting is then of the form \( \zeta' = \zeta \alpha \) with \( \alpha : G \to A \) a homomorphism.

Suppose \( A = \mu_2 \), with cocycle \( c \), and \( A \subset \mu_{ab} \) with \( b \) odd. If \( \zeta : G \to \mu_{ab} \) is a splitting of \( c \), then \( \zeta^b \) is a \( \mu_a \) splitting. Therefore we will restrict consideration to \( \mu_n \) with \( n \) a power of 2.

Now suppose \( G \) is abelian. The universal coefficient theorem for group cohomology gives an exact sequence:

\[ 1 \to \text{Ext}(G, A) \to H^2(G, A) \xrightarrow{\phi} \text{Hom}(\Lambda^2 G, A) \to 1. \]
Here $G$ and $A$ are considered as $\mathbb{Z}$-modules, $\text{Hom} = \text{Hom}_\mathbb{Z}$, $\text{Ext} = \text{Ext}_\mathbb{Z}$, and $\text{Hom}(\Lambda^2 G, A)$ consists of alternating, bilinear maps $G \times G \to A$.

If $c$ is a 2-cocycle, representing the class $\bar{c} \in H^2(G, A)$, then $\phi(\bar{c})(g, h) = c(g, h) c(h, g)^{-1}$. In terms of the group, suppose $p : H \to G$ is the corresponding extension. For $g, h \in G$ and any section $s$ let $\{g, h\}$ be the commutator $s(g)s(h)s(g)^{-1}s(h)^{-1}$. This is contained in $A$, is independent of the choice of $s$, and $\phi(\bar{c})(g, h) = \{g, h\}$. In particular $\phi(\bar{c}) = 1$ if and only if $H$ is abelian, so $\text{Ext}(G, A) \subset H^2(G, A)$ parametrizes the abelian extensions of $G$ by $A$.

Let $G^n = \{g^n | g \in G\}$ and $nG = \{g \in G | g^n = 1\}$. The next result is presumably well-known to the experts.

**Lemma 1.1.** For any positive integer $n$, inclusion $i : nG \hookrightarrow G$ induces an isomorphism:

$$\text{Ext}(G, \mu_n) \simeq \text{Ext}(nG, \mu_n).$$

**Proof.** Consider the maps

$$G \xrightarrow{\alpha} G^n \xrightarrow{\beta} G$$

where $\alpha(g) = g^n$ and $\beta$ is inclusion. The induced map $\alpha^* \beta^* : \text{Ext}(G, \mu_n) \to \text{Ext}(G, \mu_n)$ is induced by the $n$th power map $g \to g^n$ on $G$. This is the same map as that induced by the $n$th power map on $\mu_n$, and therefore $\alpha^* \beta^* = 0$.

Now the long exact cohomology sequence corresponding to $0 \to G^n \xrightarrow{\beta} G \to G/G^n \to 0$ has final two terms $\text{Ext}(G, A) \xrightarrow{\beta^*} \text{Ext}(G^n, A) \to 0$. Therefore $\beta^*$ is surjective, which implies $\alpha^* = 0$. On the other hand the short exact sequence

$$0 \to nG \xrightarrow{i} G \xrightarrow{\alpha} G^n \to 0$$

gives rise to the long exact sequence

$$0 \to \text{Hom}(G^n, A) \to \text{Hom}(G, A) \to \text{Hom}(nG, A) \to \text{Ext}(G^n, A) \xrightarrow{\alpha^*} \text{Ext}(G, A) \xrightarrow{i^*} \text{Ext}(nG, A) \to 0.$$ 

Since $\alpha^* = 0$, $i^*$ is an isomorphism. $\square$

**Remark 1.2.** In our setting $\mu_2 = \pm 1$. For the $\mu_2$ extension $\bar{T}$ to split it is necessary that it splits over $\pm 1$. Perhaps surprisingly the converse holds as well by the Lemma.

For later use we note an explicit formula for a splitting of $\alpha^* \beta^* c$. We drop the assumption that $H$ is abelian, so let $p : H \to G$ be an extension, with section $s$ and corresponding cocycle $c$.

**Lemma 1.3.** Let

$$\tau(g) = s(g^n)s(g)^{-n} = c(g, g)^{-1}c(g, g^2)^{-1}\ldots c(g, g^{n-1})^{-1} \in A.$$
Then
\[ c(g^n, h^n) = \tau(g)\tau(h)\tau(gh)^{-1}\{g, h\}^{n(n-1)/2}. \]

If $H$ is abelian then $d\tau = \alpha^\ast \beta^\ast c$.

Note that $\{g, h\}^{n(n-1)/2} = \pm 1$, and is identically 1 if $n$ is odd. Compare ([3], p. 130) and ([5], §4).

Proof. This follows from the identity
\[ [s(g)s(h)]^n = s(g)^n s(h)^n \{h, g\}^{n(n-1)/2}. \]
Using $s(g)s(h) = s(gh)c(g, h)$ and $s(g)^n = s(g^n)\tau(g)^{-1}$, the left hand side is equal to
\[ s(gh)^n = s(g^n h^n)\tau(gh)^{-1}. \]
The right hand side is
\[ s(g^n)s(h^n)\tau(g)^{-1}\tau(h)^{-1}\{h, g\}^{n(n-1)/2} = s(g^n h^n)c(g^n, h^n)\tau(g)^{-1}\tau(h)^{-1}\{h, g\}^{n(n-1)/2} \]
and the first assertion follows. Since $\alpha^\ast \beta^\ast c(g, h) = c(g^n, h^n)$, the second assertion is equivalent to
\[ c(g^n, h^n) = \tau(g)\tau(h)\tau(gh)^{-1} \]
which is (8) for $H$ abelian.

For $H$ an abelian extension (9) implies $\tau$ is a character when restricted to $nG$. In terms of the exact sequence (7), $c \in \text{Ext}(G, A)$, $\beta^\ast c \in \text{Ext}(G^n, A)$, $\alpha^\ast \beta^\ast c = 0$, and $\beta^\ast c$ is the image of $\tau \in \text{Hom}(nG, A)$. Thus $\beta^\ast c = 0$ if $\tau$ extends to an element of $\text{Hom}(G, A)$.

More generally, we try to find a splitting subgroup $A$ of $\beta^\ast c$, i.e., $c$ restricted to $G^n$, together with an explicit formula. Note that (9) does not necessarily define such a splitting since the function $g^n \rightarrow \tau(g)$ is not necessarily well-defined. Let $\alpha$ be a character of $G$ whose restriction to $nG$ is equal to $\tau^{-1}$. Then $\zeta_\alpha(g^n) := \tau(g)\alpha(g)$ is well-defined, and $d\zeta_\alpha = c$. The minimal splitting subgroup for $\beta^\ast c$ is thus the minimal subgroup $A$ of $\mathbb{T}$, containing $\mu_n$, such that $\tau$ restricted to $nG$ can be extended to a character of $G$ with values in $A$.

Characters and the $\sqrt{\alpha}$ extension.

For the remainder of this section let $p : \tilde{G} \rightarrow G$ be a $\mu_2$ extension of a group $G$. We do not assume that $G$ or $\tilde{G}$ is abelian. If $\alpha$ is a character of $G$, and $\tau(z) = z^2$ ($z \in \mathbb{C}^\ast$) then the pullback of $\tau$ via $\alpha$ is a $\mu_2$ extension of $G$, and may be realized as the subgroup of $G \times \mathbb{C}^\ast$ given by $\{(g, z) | \alpha(g) = \tau(z)\}$. Projection on the second factor is a genuine character $\tilde{\alpha}$ of $\tilde{G}$ satisfying $\tilde{\alpha}^2 = \alpha \circ p$. This is sometimes denoted the $\sqrt{\alpha}$-extension of $G$. It may or may not be the trivial extension.
We see that $\tilde{G}$ has a $T$-splitting if and only if there is a genuine character of $\tilde{G}$. More precisely:

**Lemma 1.4.** Suppose there is a genuine character $\tilde{\alpha}$ of $\tilde{G}$. Then $\tilde{\alpha}^2$ factors to a character $\alpha$ of $G$, and $\tilde{G}$ is isomorphic to the $\sqrt{\alpha}$ extension of $G$. If $\text{Image}(\tilde{\alpha}) \subset A \subset T$ then $G_A \simeq G \times A$. The minimal splitting group for $G$ is $\mu_n(\tilde{G})$ where $n(\tilde{G})$ is the minimal order of a genuine character of $\tilde{G}$.

Conversely if $G_A \simeq G \times A$ then there is a genuine character of $\tilde{G}$ with values in $A$.

Note that there exists a genuine character $\tilde{\alpha}$ of $\tilde{G}$ if and only if $z \not\in [\tilde{G}, \tilde{G}]$ where $z$ is the non-trivial element in the inverse image of 1. In particular this holds if $\tilde{G}$ is abelian, which proves the existence of a $T$-splitting (cf. [2]).

**Proof.** The map $\phi : g \to (p(g), \tilde{\alpha}(g)) \subset G \times T$ is an isomorphism of $\tilde{G}$ with the pullback of $\tau$ via $\alpha$. This is a subgroup of $G \times A$, and $\phi$ extends to an isomorphism of $G_A$ with $G \times A$. The final two assertions are immediate. □

**Remark 1.5.** In the setting of the Lemma, suppose $\zeta$ is a $T$-splitting of the cocycle defining $\tilde{G}$ (with respect to a section $s$). Then $\alpha := \zeta^2$ is a character of $G$, and $\tilde{G}$ is isomorphic to the $\sqrt{\alpha}$ cover of $G$.

Theorem 3 and Corollary 4 are immediate consequences of the Lemma. Theorem 1 also follows from the Lemma, from a computation of $n = n(T)$: $n$ is the minimal power of 2 such that $z \not\in \tilde{T}^n$. We follow a different approach, by giving explicit formulas for the minimal splitting $\zeta$ in Sections 3-5.

## 2. Moore cohomology and $\tilde{SL}(2, \mathbb{F})$.

Now suppose $G$ and $A$ are locally compact topological groups, $A$ is abelian, and $G$ acts continuously on $A$. In our applications $A$ will either be $\mu_n$ or $T$, with trivial $G$ action. C. Moore has defined cohomology groups $H^2_{\text{top}}(G, A)$ using measurable cochains [8]. In the case of a totally disconnected group it is equivalent to use continuous cochains. Viewing $G$ and $A$ as abstract groups, there is a natural homomorphism $H^2_{\text{top}}(G, A) \to H^2(G, A)$. In general it is neither surjective nor injective.

Now let $G = GL(2, \mathbb{F})$. We recall the definition of the standard cocycle on $G$ [6], cf. ([4], p. 41). For $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ let $x(g) = c$ (resp. $d$) if $c \neq 0$ (resp. $c = 0$). Then

\begin{equation}
(10) \quad c(g, h) = (x(g)x(gh), x(h)x(gh))_{\mathbb{F}}(\det(g), x(g)x(gh))_{\mathbb{F}}.
\end{equation}
This defines a $\mu_2$-extension $\widehat{GL(2,F)}$ of $GL(2,F)$, with distinguished section $s$. We write $\widehat{GL(2,F)}$ in cocycle notation as usual. The restriction to $SL(2,F)$ is isomorphic to $SL(2,F)$.

Up to conjugation $GL(2,F)$ contains one hyperbolic torus isomorphic to $F^* \times F^*$, and for each quadratic extension $E$ of $F$ one elliptic torus isomorphic to $E^*$. The commutator of two elements $z,w$ of $E^*$ is given by ([3], p. 128)

\begin{align}
\{z,w\} &= (z,\bar{w})_E \\
&\quad = (z,w)_E(Nz,Nw)_F.
\end{align}

Here and elsewhere we suppress the map $\iota : E^* \hookrightarrow GL(2,F)$ from the notation, and write $\{z,w\} := \{\iota(z),\iota(w)\}$. The commutator is trivial when restricted to $E_1 = E^* \cap SL(2,F)$.

3. The hyperbolic torus.

Let $\iota : F^* \hookrightarrow SL(2,F)$, so $T = \iota(F^*)$ is a hyperbolic torus. After conjugation we may assume $\iota(x) = \text{diag}(x,x^{-1})$. We drop $\iota$ from the notation and identify $x$ with $\iota(x)$. We prove Theorem 1 in this case, together with a formula for the minimal splitting. While these results are well-known they are not easy to find in the literature, and the calculation illustrates some of the ideas in the next section.

The restriction of $SL(2,F)$ to $T$ defines a $\mu_2$ extension $\tilde{T}$ of $T$ with cocycle

\[ c(x,y) = (x,y)_F. \]

In particular $(-I,e)^2 = (I,(-1,-1)_F)$, so the restriction of the extension to $\pm I$ is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ if $(-1,-1)_F = 1$, or $\mathbb{Z}/4\mathbb{Z}$ if $(-1,-1)_F = -1$. By Lemma 1.1 the minimal splitting group for $\tilde{T}$, considered as an abstract group, is $\mu_2$ (resp. $\mu_4$) if $(-1,-1)_F = 1$ (resp. $-1$). It remains to show this splitting is measurable. We do this by computing it explicitly.

Fix $\psi$ and write $\gamma(x) = \gamma_F(x,\psi)$. The key point is that properties (1) and (2) of the Weil index shows that $\delta\gamma = c$ and $\delta^4 = 1$, so $\gamma$ is a measurable $\mu_4$ splitting of $c$. This completes the proof in case $(-1,-1)_F = -1$, so assume $(-1,-1)_F = 1$. Let $\alpha$ be a character of $F^*$ satisfying $\alpha(x^2) = (-1,x)_F$. To see that such a character exists, define $\alpha$ restricted to $F^2$ by this formula; it is well-defined since $\alpha((-x)^2) = (-1,-x)_F = (-1,-1)_F \alpha(x^2) = \alpha(x^2)$. Extend arbitrarily from $F^2$ to $F^*$. By (2) $\alpha(x)^2 \gamma(x)^2 = (-1,x)^2_F = 1$. Let $\zeta_\alpha = \gamma\alpha$, i.e.,

\[ \zeta_\alpha(x) = \gamma(x,\psi_F)\alpha(x). \]

Then $d\zeta_\alpha = d\gamma = c$ and $\zeta_\alpha$ is a $\mu_2$-splitting of $c$.

Choose representatives $a_1,a_2,\ldots,a_m \in F^*$ of generators of $F^*/F^2 \simeq (\mathbb{Z}/2\mathbb{Z})^m$. (By [4], Lemma 0.3.2, $2^m = 4/|F|$.) Given any choice of signs
Here we may choose \( \alpha \) so that \( \zeta(a_i) = \epsilon_i \), and \( \alpha(x^2) = (-1, x)_F \) for all \( x \in F \). Then \( \zeta \) extends uniquely to a splitting.

For example, if \(-1 \in F^* \) then we may take \( \alpha = 1 \) and \( \zeta(x) = \gamma_F(x, \psi_F) = \pm 1 \). On the other hand, suppose \(-1 \notin F^* \) and the residual characteristic of \( F \) is odd. We may take representatives \( \pm 1, \pm \varpi \) for \( F^*/F^* \) (\( \varpi \) is a uniformizing parameter) and then choose \( \zeta \) satisfying:

\[
\zeta(\pm x^2) = (-1, x)_F \\
\zeta(\pm \varpi x^2) = \pm (-1, x)_F.
\]

4. Elliptic tori.

Let \( T \) be an elliptic torus of \( SL(2, F) \) as in \S 2. Thus \( E \) is a quadratic extension of \( F \), \( \iota : E^1 \hookrightarrow SL(2, F) \) is an embedding, and \( T = \iota(E^1) \). As in \S 3 we fix \( \iota \) and drop it from the notation.

As in the case of the hyperbolic torus, \((-I, \epsilon)\) has order 2 or 4 depending on whether \((-1, -1)_F \) is odd or even. By Lemma 1.1 this proves \( \mu_2 \) is a splitting group if and only if \((-1, -1)_F = 1 \), so assume \((-1, -1)_F = -1 \).

Let \( F = \mathbb{R}, E = \mathbb{C} \). Since \( T = T^2 \), it is enough to find a splitting of \( \beta \cdot c \) as in \S 2. Choose a character \( \alpha \) of \( E^1 \) such that \( \alpha(-1) = -1 \), i.e., \( \alpha(z) = z^n \) for \( n \) odd. It is easy to see \( c(-z, -z) = -c(z, z) \), and the discussion in \S 2 shows that

\[
\zeta_\alpha(z^2) := c(z, z)\alpha(z)
\]

is a well-defined (measurable) \( T \)-splitting. Since \( \zeta_\alpha \) is surjective onto \( T \) for any \( \alpha \), this shows that there is no \( A \)-splitting for any proper subgroup \( A \) of \( T \).

We now assume \( F \) is non-archimedean. Since \((-1, -1)_F = -1 \), \( F \) is an extension of \( Q \) of odd degree, and \(-1 \notin F^* \). If \(-1 \notin F^* \) then \( 4T = \{ \pm 1 \} \). Since \((-I, \zeta)^4 = (I, \pm \zeta^2)^2 = I \) for all \( \zeta \in \mu_4 \), the \( \mu_4 \)-extension splits over \( 4T \).

Suppose \(-1 = \delta^2 \) (\( \delta \in E^* \)). We claim \( 4T = 8T = \{ \pm 1, \pm \delta \} \). It is enough to show \( E^* \) does not contain a primitive eighth root of unity, or equivalently \( \delta \notin E^{*2} \). Since \( F \) is an extension of \( Q \) of odd degree, \( 2 \notin E^{*2} \). But then \((a + b\delta)^2 = \delta \) implies \( a^2 = \pm \frac{1}{2} \), which is a contradiction.

For any \( \zeta \in \mu_8 \) we compute \( (\delta, \zeta)^8 = (\delta, \pm \zeta^2)^4 = (I, -\zeta^4)^2 = I \), which implies the extension splits over \( 8T \).

Therefore \( \mu_4 \) (resp. \( \mu_8 \)) is a minimal splitting group for \( T \) if \( F \) is non-archimedean, \((-1, -1)_F = -1 \), and \(-1 \notin E^{*2} \) (resp. \(-1 \notin E^{*2} \)). It remains to show these splittings can be chosen to be measurable. In the next section we give explicit such splittings.

**Remark 4.1.** We have shown the cohomology class \( \sigma \in H^2_{\text{top}}(G, A) \) has image 0 in \( H^2(G, A) \) in the given cases. An argument due to Jonathan
Rosenberg shows that the map \( H^2_{\text{top}}(G, A) \to H^2(G, A) \) is injective in this situation. Since we are interested in explicit formulas for the splittings in any case, we do not pursue this approach. For \( G \) perfect (not at all the case here!) the injectivity of \( \phi \) is known ([9], Theorem 2.3).

5. Explicit splittings for elliptic tori.

We continue with the notation of the previous section. The embedding \( \iota : \mathbb{E}^1 \to SL(2, \mathbb{F}) \) extends to an embedding \( \iota : \mathbb{E}^* \to GL(2, \mathbb{F}) \). We will make use of the non-abelian \( \mu_2 \) extension of \( \mathbb{E}^* \) obtained by restricting the extension \( GL(2, \mathbb{F}) \) of \( GL(2, \mathbb{F}) \).

We proceed as follows. Since \( \{ z, w \} = (z, w)_{\mathbb{E}} \), the commutator is trivial on \( \mathbb{E}^* \). By the method of §1 we find a splitting of this extension. As in §3 we also find a splitting of the extension of \( \mathbb{F}^* \subset \mathbb{E}^* \). The extension of \( \mathbb{E}^2 \mathbb{F}^* \) is abelian, and by an explicit version of the Mayer-Vietoris sequence we obtain a splitting of this extension. Finally \( \mathbb{E}^1 \) is contained in \( \mathbb{E}^2 \mathbb{F}^* \), and we restrict to obtain a splitting of the extension of \( \mathbb{E}^1 \).

Fix non-trivial additive characters \( \psi_{\mathbb{F}} \) of \( \mathbb{F} \) and \( \psi_{\mathbb{E}} \) of \( \mathbb{E} \). Recall (5) the restriction of \( \gamma_{\mathbb{E}}(\cdot, \psi_{\mathbb{E}}) \) to \( \mathbb{F}^* \) is a quadratic character.

**Lemma 5.1.** Suppose \( \lambda \) (respectively \( \mu \)) is a splitting of the cocycle restricted to \( \mathbb{E}^2 \) (respectively \( \mathbb{F}^* \)). Assume \( \lambda(x) = \mu(x) \) for \( x \in \mathbb{E}^2 \cap \mathbb{F}^* \). For \( z \in \mathbb{E}^* \), \( x \in \mathbb{F}^* \) let \( \zeta_{\lambda, \mu}(z^2 x) := \lambda(z^2) \mu(x)c(x, z^2) \). Then \( \zeta_{\lambda, \mu} \) is a well-defined splitting of the cocycle restricted to \( \mathbb{E}^2 \mathbb{F}^* \).

Conversely if \( \zeta \) is any splitting of the cocycle restricted to \( \mathbb{E}^2 \mathbb{F}^* \) then \( \zeta = \zeta_{\lambda, \mu} \) with \( \lambda = \zeta|_{\mathbb{E}^2} \) and \( \mu = \zeta|_{\mathbb{F}^*} \).

**Proof.** Let \( \zeta \) be any splitting of the cocycle. Then \( \zeta(z^2 x) = \zeta(z^2) \zeta(x)c(x, z^2) \) by (6) and the second assertion is immediate.

Given \( \lambda \) and \( \mu \), choose any splitting \( \zeta \). Then \( \lambda \) and \( \zeta|_{\mathbb{E}^2} \) both define splittings, so \( \lambda(z^2) = \zeta(z^2) \alpha(z^2) \) for some character \( \alpha \) of \( \mathbb{E}^2 \). Similarly \( \mu(x) = \zeta(x) \beta(x) \) for some character \( \beta \) of \( \mathbb{F}^* \). Let \( \tau \) be a character of \( \mathbb{E}^* \) extending \( \alpha \) and \( \beta \); this exists since, for \( x \in \mathbb{E}^2 \cap \mathbb{F}^* \), \( \lambda(x) = \mu(x) \) implies \( \alpha(x) = \beta(x) \). Then

\[
\lambda(z^2) \mu(x)c(x, z^2) = \zeta(z^2) \alpha(z^2) \zeta(x) \beta(x)c(x, z^2) = \zeta(z^2) \alpha(z^2) \beta(x) = \zeta(z^2) \tau(z^2 x).
\]

This shows that \( \zeta_{\lambda, \mu} \) is well-defined and is a splitting of the cocycle. \( \square \)

**Lemma 5.2.** (1) Choose a character \( \alpha \) of \( \mathbb{E}^* \) satisfying

\[
\alpha(-1) = (-1, -1)_{\mathbb{F}} \gamma_{\mathbb{E}}(-1, \psi_{\mathbb{E}})
\]

and let

\[
\lambda_{\alpha}(z^2) = c(z, z) \gamma_{\mathbb{E}}(z, \psi_{\mathbb{E}}) \gamma_{\mathbb{F}}(Nz, \psi_{\mathbb{F}}) \alpha(z).
\]
Thus by (8) and (11), Part (1) is an extension of (8) to the case of a non-abelian group. 

Proof. Let \( \beta \) be a character of \( \mathbb{F}^* \) and let 

\[
\mu_\beta(x) = \gamma_\mathbb{F}(x, \psi_\mathbb{F})\beta(x).
\]

Then \( \mu_\beta \) is a splitting of the cocycle restricted to \( \mathbb{F}^* \), and every splitting of the cocycle restricted to \( \mathbb{F}^* \) is equal to \( \mu_\beta \) for some \( \beta \).

(3) Suppose \( \alpha \in \mathbb{E}^*, \beta \in \mathbb{F}^* \) satisfy

\[
(14) \quad \alpha(z) = \gamma_\mathbb{E}(z^2, \psi_\mathbb{E})\gamma_\mathbb{E}(Nz, \psi_\mathbb{E})\gamma_\mathbb{E}(z, \psi_\mathbb{E})c(z, z, \beta(z^2)) \quad (z^2 \in \mathbb{F}^*).
\]

In particular \( \alpha \) satisfies (13). For \( z \in \mathbb{E}^*, x \in \mathbb{F}^* \) define

\[
\zeta_{\alpha, \beta}(z^2 x) := \lambda_\alpha(z^2)\mu_\beta(x)c(x, z^2)
\]

\[
= \gamma_\mathbb{E}(z, \psi_\mathbb{E})\gamma_\mathbb{E}(Nz, \psi_\mathbb{E})\gamma_\mathbb{E}(x, \psi_\mathbb{E})\alpha(z)\beta(x)c(z, z)\alpha(x, z^2).
\]

Then \( \zeta \) is a well-defined splitting of the cocycle restricted to \( \mathbb{E}^* \mathbb{F}^* \). Furthermore every splitting of the cocycle restricted to \( \mathbb{E}^* \mathbb{F}^* \) is equal to \( \zeta_{\alpha, \beta} \) for some \( \alpha, \beta \) satisfying (14).

Proof. Part (1) is an extension of (8) to the case of a non-abelian group. Thus by (8) and (11),

\[
c(z^2, w^2) = c(z, z)c(w, w)c(zw, zw)\{z, w\}
\]

\[
= c(z, z)c(w, w)c(zw, zw)(Nz, Nz)_{\mathbb{F}}.
\]

Replacing \( (z, w)_{\mathbb{E}} \) by \( \gamma_\mathbb{E}(z, \psi_\mathbb{E})\gamma_\mathbb{E}(w, \psi_\mathbb{E})\gamma_\mathbb{E}(zw, \psi_\mathbb{E})^{-1} \), and similarly \( (Nz, Nz)_{\mathbb{F}} \) gives

\[
c(z^2, w^2) = \tau(z)\tau(w)\tau(zw)^{-1}
\]

with \( \tau(z) = c(z, z)\gamma_\mathbb{E}(z, \psi_\mathbb{E})\gamma_\mathbb{E}(Nz, \psi_\mathbb{E}) \). The same relation holds with \( \tau(z) \) replaced by \( \tau(z)\alpha(z) \).

We check the condition that \( \lambda_\alpha(z^2) := \tau(z)\alpha(z) \) be well-defined:

\[
\tau(-z)\alpha(-z)
\]

\[
= c(-z, -z)\gamma_\mathbb{E}(-z, \psi_\mathbb{E})\gamma_\mathbb{E}(N(-z), \psi_\mathbb{E})\alpha(-z)
\]

\[
= c(-z, -z)\gamma_\mathbb{E}(-1, \psi_\mathbb{E})\gamma_\mathbb{E}(z, \psi_\mathbb{E})(-1, z)\gamma_\mathbb{E}(Nz, \psi_\mathbb{E})\alpha(z)\alpha(-1).
\]

A simple calculation using (10) gives

\[
(16) \quad c(-z, -z) = (-1, -1)_{\mathbb{F}}(-1, Nz)_{\mathbb{F}}c(z, z)
\]

and inserting this gives

\[
\tau(-z)\alpha(-z) = (-1, -1)_{\mathbb{F}}\gamma_\mathbb{E}(-1, \psi_\mathbb{E})\alpha(-1)\tau(z)\alpha(z)
\]

\[
= \tau(z)\alpha(z) \quad \text{by (13)}.
\]
Fix $\alpha$ satisfying (13). If $\lambda$ is any splitting of the cocycle restricted to $E^*\mathbb{F}^*$ then $\lambda = \lambda_\alpha \delta$ for some character $\delta$ of $E^*\mathbb{F}^*$. Extend $\delta$ to a character $\delta^*$ of $E^*\mathbb{F}^*$. Then $\lambda_\alpha \delta = \lambda_{\alpha \delta^*}z$, and $\alpha \delta^*$ satisfies (13). This proves (1).

By (10), $c(\text{diag}(x, x), \text{diag}(y, y)) = (x, y)_\mathbb{F}$, and (2) follows as in Section 3.

For (3) apply Lemma 5.1. Inserting $\lambda_\alpha, \mu_\beta$ in the condition of the Lemma gives (14). The final assertion follows as in the proof of (1). This completes the proof.

Let $\zeta$ be any splitting of the cocycle restricted to $E^*\mathbb{F}^*$. Then $\zeta = \zeta_{\alpha, \beta}$ for some $\alpha, \beta$ satisfying (14). This implies:

**Lemma 5.3.** The map $\alpha(z) := \gamma_{\mathbb{F}}(z^2, \psi_{\mathbb{F}})\gamma_{\mathbb{F}}(Nz, \psi_{\mathbb{F}})\gamma_{\mathbb{F}}(z, \psi_{\mathbb{F}})c(z, z)$ is a character of $\{z \in E^* \mid z^2 \in \mathbb{F}^*\}$.

For completeness we also prove this directly:

\[
\alpha(zw) = \gamma_{\mathbb{F}}(z^2w^2, \psi_{\mathbb{F}})\gamma_{\mathbb{F}}(N(zw), \psi_{\mathbb{F}})\gamma_{\mathbb{F}}(zw, \psi_{\mathbb{F}})c(zw, zw) = \gamma_{\mathbb{F}}(z^2, \psi_{\mathbb{F}})\gamma_{\mathbb{F}}(w^2, \psi_{\mathbb{F}})(z^2, w^2)\gamma_{\mathbb{F}}(Nz, \psi_{\mathbb{F}})\gamma_{\mathbb{F}}(Nw, \psi_{\mathbb{F}})(Nz, Nw)_\mathbb{F} \gamma_{\mathbb{F}}(z, \psi_{\mathbb{F}})\gamma_{\mathbb{F}}(w, \psi_{\mathbb{F}})(z, w)c(zw, zw) = \alpha(z)\alpha(w)c(z, z)c(w, w)c(z^2, w^2)\{z, w\}c(zw, zw) \text{ by (11)} = \alpha(z)\alpha(w) \text{ by (8)}.
\]

**Remark 5.4.** Given $\alpha \in \hat{E}^*$ there exists $\beta \in \hat{E}^*$ satisfying (14) if and only if (13) holds.

This follows by an argument similar to that of the proof of Lemma 5.2 (1): Define $\beta(z^2) = \alpha(z)\gamma_{\mathbb{F}}(z^2, \psi_{\mathbb{F}})^{-1}\gamma_{\mathbb{F}}(Nz, \psi_{\mathbb{F}})^{-1}\gamma_{\mathbb{F}}(z, \psi_{\mathbb{F}})^{-1}c(z, z)$; this is well-defined if (13) holds, and extends to a character of $\mathbb{F}^*$.

Suppose $z \in \mathbb{E}^1$. By Hilbert’s Theorem 90, $z = w/\pi$ for some $w \in \mathbb{E}^*$. Then $z = w^2/N(w) \in E^*\mathbb{F}^*$, so $\zeta$ restricts to a splitting of the extension of $\mathbb{E}^1$. We now make explicit choices such that $\zeta$ is a $T$ or $\mu_n$ splitting as in Theorem 1.

**Lemma 5.5.** We may choose $\psi_{\mathbb{E}}$ so that

\[
\gamma_{\mathbb{E}}(x, \psi_{\mathbb{E}}) = 1 \quad (x \in \mathbb{F}^*).
\]

**Proof.** Since $\gamma_{\mathbb{E}}(\cdot, \psi_{\mathbb{E}})$ is a quadratic character of $\mathbb{F}^*$, $\gamma_{\mathbb{E}}(x, \psi_{\mathbb{E}}) = (x, y)_\mathbb{F}$ for some $y \in \mathbb{F}^*$. Suppose $\mathbb{E} = \mathbb{F}(\sqrt{\Delta})$. Then $\gamma_{\mathbb{E}}(\Delta, \psi_{\mathbb{E}}) = 1$ since $\Delta$ is a square in $\mathbb{E}^*$. Therefore $(\Delta, y)_\mathbb{F} = 1$, so $y = Nw$ for some $w \in \mathbb{E}^*$. Replacing $\psi_{\mathbb{E}}$ by $w\psi_{\mathbb{E}}$ gives (cf. [10], Appendix)

\[
\gamma_{\mathbb{E}}(x, w\psi_{\mathbb{E}}) = (x, w)_\mathbb{F}\gamma_{\mathbb{E}}(x, \psi_{\mathbb{E}}) = (x, Nw)_\mathbb{F}(x, y)_\mathbb{F} = 1.
\]

This completes the proof. \qed
Theorem 5.7. Choose a non-trivial character \( \psi_{E} \) satisfying Lemma 5.5. Choose \( \alpha, \beta \) satisfying (14) and let \( \zeta = \zeta_{\alpha, \beta} \). Then
\[
\zeta(z)^{2} = \alpha(z) \quad (z \in \mathbb{E}^{1}).
\]

Proof of Theorem 5.7. We now complete the proof of Theorem 1.

We make some explicit choices and summarize the preceding discussion. If \( (w, z)_{E, 4} \) is a character of \( \mathbb{E}^{*} \) of order 4. Choose \( z_{2} \in \mathbb{E}^{*} \) satisfying \( (z_{2}, -1)_{E, 4} = -1 \).

Theorem 5.7. Choose a non-trivial character \( \psi_{E} \) of \( \mathbb{E} \) such that \( \gamma_{E}(x, \psi_{E}) = 1 \) for all \( x \in \mathbb{E}^{*} \) (Lemma 5.5). For \( z \in \mathbb{E}^{*} \) let
\[
\alpha(z) := \begin{cases} 
\text{F = } \mathbb{R} \\
1 & (\text{if } (\text{\text{-}1}, -1)_{E} = 1) \\
(z_{1}, z)_{E} & (\text{if } (\text{\text{-}1}, -1)_{E} = -1, -1 \notin \mathbb{E}^{*} \\
(z_{2}, z)_{E, 4} & (\text{if } (\text{\text{-}1}, -1)_{E} = -1, -1 \in \mathbb{E}^{*}. 
\end{cases}
\]

Then \( \alpha \) is a character of \( \mathbb{E}^{*} \) of order \( \infty, 1, 2 \) or 4 respectively, satisfying (18).

Choose a character \( \beta \) of \( \mathbb{F}^{*} \) satisfying
\[
\beta(z^{2}) = \gamma_{E}(z^{2}, \psi_{E})^{-1} \gamma_{E}(N z, \psi_{E})^{-1} \gamma_{E}(z, \psi_{E})^{-1} c(z, z) \alpha(z) \quad (z^{2} \in \mathbb{F}^{*})
\]
(cf. Remark 5.4). In particular

\[ \beta(x^2) = (-1, x)_{\mathbb{F}} \alpha(x) \quad (x \in \mathbb{F}^*). \]

Let

\[ (21) \quad \zeta(w/\overline{w}) = \gamma_{\mathbb{F}}(w, \psi_{\mathbb{F}}) \alpha(w) \beta(Nw^{-1}) c(w, w) c(Nw^{-1}, w^2). \]

Then \( \zeta \) is a splitting of the cocycle restricted to \( \mathbb{E}^1 \). Furthermore for \( z \in \mathbb{E}^1 \)

\[ \zeta(x)^2 = \alpha(z), \]

and for \( \mathbb{F} \) non-archimedean this gives \( \zeta(z)^n = \alpha(z)^{n/2} = 1 \)

with

\[
\begin{align*}
    n &= \begin{cases} 
        2 & (-1, -1)_{\mathbb{F}} = 1 \\
        4 & (-1, -1)_{\mathbb{F}} = -1, -1 \not\in \mathbb{E}^* \\
        8 & (-1, -1)_{\mathbb{F}} = -1, -1 \in \mathbb{E}^*.
    \end{cases}
\end{align*}
\]

Remark 5.8. With \( \alpha, \beta \) as in the Theorem,

\[ \zeta(w/\overline{w}) = \zeta_{\alpha, \beta}(w/\overline{w})(-1, Nw)_{\mathbb{F}}. \]

We have dropped the term \((-1, Nw)\), which is allowed since \( w/\overline{w} \to (-1, Nw) \) is a quadratic character of \( \mathbb{E}^1 \).

Henceforth write \( \mathbb{E} = \mathbb{F}(\delta) \) with \( \Delta := \delta^2 \in \mathbb{F} \).

Remark 5.9. Condition (20) is equivalent to

\[ \beta(x^2) = (-1, x)_{\mathbb{F}} \alpha(x) \quad (x \in \mathbb{F}^*) \]

\[ \beta(\Delta) = \gamma_{\mathbb{F}}(-1, \psi_{\mathbb{F}})^{-1} \gamma_{\mathbb{E}}(\delta, \psi_{\mathbb{E}})^{-1} c(\delta, \delta) \alpha(\delta). \]

The splitting has a simple formula on \( T^2 \):

\[ \zeta(z^2) = c(z, z) \alpha(z) \quad (z \in \mathbb{E}^1), \]

which is independent of \( \beta, \psi_{\mathbb{F}} \) and \( \psi_{\mathbb{E}} \). Note that any two \( \mu_2 \) splittings of \( T \)
have the same restriction to \( T^2 \) since they differ by a quadratic character.

The map \( w/\overline{w} \to Nw \) induces an isomorphism \( T/T^2 \simeq N\mathbb{E}^*/\mathbb{F}^* \). Choose representatives \( a_1, \ldots, a_n \) of generators of \( N\mathbb{E}^*/\mathbb{F}^* \), with corresponding elements \( z_1, \ldots, z_n \in T \). Given \( \alpha \) there are two choices of each \( \beta(a_i) \) differing by sign, and these signs may may be chosen arbitrarily. The following result follows easily.

Corollary 5.10. Choose representatives \( z_1, \ldots, z_m \) of generators of \( T/T^2 \simeq N\mathbb{E}^*/\mathbb{F}^* \). Choose \( \alpha \) as in Theorem 5.7. Define

\[ \zeta(z^2) = c(z, z) \alpha(z) \quad (z \in \mathbb{E}^1) \]

and for \( 1 \leq i \leq m \) let \( \zeta(z_i) \) be either square root of \( \alpha(z_i) \). Then \( \zeta \) extends uniquely to a splitting of the cocycle as in Theorem 5.7.
In the non-archimedean case by ([4], Lemma 0.3.2) \(|T/T^2| = 2/|2|_F\), which equals 2 if the residual characteristic of \(F\) is odd.

We conclude with a few remarks about the definition of \(\zeta\).

From the definition we have for \(w \in E^*\):
\[
c(w, w) = (-Nw, x(w)x(w^2))_F
\]
and
\[
c(Nw^{-1}, w^2) = (Nw, x(w^2))_F.
\]
Note that for \(\lambda \in F^*, w \in E^*\) we have
\[
c(\lambda, w) = (\lambda, x(w))_F.
\]
Fix \(u \in E^*\) with trace\((u) = 0\), and define \(\text{Tr}_u : E^* \to F^*\) by
\[
\text{Tr}_u(z) = \begin{cases} 
\text{trace}(z) & \text{trace}(z) \neq 0 \\
u_z & \text{trace}(z) = 0.
\end{cases}
\]

Up to conjugation by \(SL(2,F)\) we may assume \(\iota(x+y\delta) = \begin{pmatrix} x & y\Delta/a \\ ya & x \end{pmatrix}\) for some \(a \in F^*\). If \(w = x+y\delta\) with \(xy \neq 0\) we have \(x(w)x(w^2) = 2xy^2a^2\).

Considering the cases with \(xy = 0\) separately gives
\[
c(w, w) = (-Nw, \text{Tr}_a\delta(w))_F \quad (z \in E^*)
\]
and
\[
\zeta(z^2) = (-1, \text{Tr}_a\delta(z))_F \alpha(z) \quad (z \in E^1).\]

For example if \(p\) is odd and \(E \neq F(\sqrt{-1})\) then \(m = 1\) (cf. Corollary 5.10), we may take \(z_1 = -1\), \(\alpha = 1\) and \(\zeta(-1) = 1\) which gives
\[
\zeta(\epsilon z^2) = c(z, z)c(\epsilon, z^2) \quad (\epsilon = \pm 1, z \in E^1).
\]

For example let \(F = \mathbb{R}\) and define \(\iota(x+iy) = \begin{pmatrix} x & y \\ -y & x \end{pmatrix}\). We take \(\alpha(z) = z\) and \(\beta(x) = \pm \sqrt{|x|}\). Then for \(z \in E^1,\)
\[
\zeta(z^2) = (-1, \text{Tr}_{-i}(z))_{\mathbb{R}}z = \text{sgn}(\text{Tr}_{-i}(z))z
\]
(independent of \(\beta\)). Note that \(a = -1\) and \(\text{Tr}_{-i}(iy) = y\ (y \in \mathbb{R}^*)\).

References


Received April 29, 1999.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MD 20742
E-mail address: jda@math.umd.edu
CAUSAL COMPACTIFICATION AND HARDY SPACES FOR SPACES OF HERMITIAN TYPE

FRANK BETTEN AND GESTUR ÓLAFSSON

Let $G/H$ be a compactly causal symmetric space with causal compactification $\Phi : G/H \to \check{S}$, where $\check{S}$ is the Bergman-Silov boundary of a tube type domain $G_1/K_1$. The Hardy space $H_2(C)$ of $G/H$ is the space of holomorphic functions on a domain $\Xi(C^o) \subset G_1/K_1$, such that $\Xi(C^o) = \{ z \in G_1/K_1 \mid \psi_m(z) \neq 0 \}$, with $\psi_m$ explicitly known. We use this to construct an isometry $I$ of the classical Hardy space $H_{cl}$ on $G_1/K_1$ into $H_2(C)$ or into a Hardy space $\tilde{H}_2(C)$ defined on a covering $\tilde{\Xi}(C^o)$ of $\Xi(C^o)$. We describe the image of $I$ in terms of the highest weight modulus occuring in the decomposition of the Hardy space.

1. Introduction.

Hardy spaces on tube type domains $T_\Omega = \mathbb{R}^n + i\Omega$, associated to a homogeneous self dual cone $\Omega \subset \mathbb{R}^n$, are important objects in analysis. The Hardy space $H_{cl}$ is by definition the space of holomorphic functions on $T_\Omega$, such that the Hardy norm

$$\|f\|_2^2 := \sup_{y \in \Omega} \int_{\mathbb{R}^n} |f(x+iy)|^2 dx$$

is finite ([SW71, FK94]). The boundary value map $\beta : H_{cl} \to L^2(\mathbb{R}^n)$ is given by

$$\beta(f)(x) = \lim_{y \to 0} f(x+iy),$$

the limit taken in the $L^2(\mathbb{R}^n)$-norm. The image of $\beta$ is described by the positivity condition

$$\text{Im}(\beta) = \mathcal{F}(L^2(\Omega)),$$

where $\mathcal{F}$ is the Fourier transform and $L^2(\Omega)$ the space of $L^2$-functions supported on $\Omega \subset \mathbb{R}^n$. The evaluation map $H_{cl} \ni f \mapsto f(w) \in \mathbb{C}$, $w \in T_\Omega$, is continuous, and thus given by an element $K_w \in H_{cl}$. The function $K(z, w) := K_w(z)$ is the Cauchy kernel associated to the tube domain $T_\Omega$. The Cauchy kernel is determined by a function of one variable $K(z, w) =$
$K(z-w)$, where $K(z)$ is the Laplace transform of the characteristic function of the cone. Finally the inverse of the boundary value map is

$$f(z) = \int_{\mathbb{R}^n} \beta(f)(y) K(z-y) \, dy.$$ 

For us the tube domain $T_\Omega$ is always the unbounded realization of a Hermitian symmetric space $G_1/K_1$ such that $H_{cl}$ is also a $G_1$-representation space ([FK94]).

The notation of Hardy spaces was generalized to compactly causal symmetric spaces $G/H$ ([HOØ91]). In this case a complex manifold $\Xi(C^o) \subset G_{C}/H_{C}$, depending on the causal structure $m \mapsto C(m)$ on $G/H$, was constructed. Three important properties of $\Xi(C^o)$ are:

1. The manifold $\Xi(C^o)$ is locally isomorphic to a tube domain $q+iC^o$, where $q$ is the tangent space of $G/H$ at $x_o = eH$ and $C^o$ the interior of $C := C(x_o)$.
2. The homogeneous space $G/H$ is a boundary component of $\Xi(C^o)$.
3. There is a semigroup $\Gamma$ containing $G$ and determined by $C$ such that $\Xi(C^o) = (\Gamma^{-1} \cdot x_o)$.

The Hardy space $H_2(C)$ is defined as in the classical case to be the space of holomorphic functions on $\Xi(C^o)$ such that the Hardy norm

$$\|f\|_{H}^2 := \sup_{\gamma \in \Gamma^o} \int_{G/H} |f(\gamma^{-1} \cdot m)|^2 \, dm$$

is finite. $H_2(C)$ is a Hilbert space with norm $\|\cdot\|_H$, and — as in the classical case — there exists an isometry $\beta : H_2(C) \rightarrow L^2(G/H)$ given by

$$\beta(f) = \lim_{\Gamma^o \gamma \to 1} \gamma \cdot f,$$

where the limit is in $L^2(G/H)$ and $\gamma \cdot f(\xi) := f(\gamma^{-1} \cdot \xi)$. The left action defines a holomorphic representation $T$ of $\Gamma$ on the Hardy space and a unitary representation $\lambda$ of $G$ on $L^2(G/H)$ such that $\beta$ is an intertwining operator for the $G$-actions.

The paralles to the classical case goes further. In particular one can describe the image of $\beta$ by a positivity condition: $\beta(H_2(C))$ is the direct sum of all the holomorphic discrete series from [ÖO91] which are $C$-admissible. Point evaluation is also continuous and thus defines a kernel, the Cauchy-Szegö kernel $K(\cdot, \cdot)$. This kernel is determined by a holomorphic $H$-invariant function $\Theta_K : \Xi(C^o) \rightarrow \mathbb{C}$ such that

$$K(\gamma_1 \cdot x_o, \gamma_2 \cdot x_o) = \Theta_K(\gamma_2 \gamma_1 \cdot x_o).$$

For the first variable fixed, the kernel extends in the second variable smoothly to the boundary of $\Xi(C^o)$. Then the inverse of the boundary
value map is given just as in the classical case

\[ f(z) = \int_{G/H} \beta(f)(x) K(z, x) \, dx = \int_{G/H} f(\dot{g}) \Theta_K(g^{-1} \gamma \cdot x_o) \, d\dot{g}, \]

where \( z = \gamma \cdot x_o \in \Xi(C^o) \) and \( \dot{g} = g \cdot x_o \in G/H \).

Let \( H_2(C) = \oplus_\delta H_\delta \) be the decomposition of \( H_2(C) \) into holomorphic discrete series. Then each of the representations \( \epsilon_\delta \) on \( H_\delta \) give rise to a spherical distribution

\[ \Theta_\delta(f) := \text{pr}_\delta(f)(x_o), \]

where \( \text{pr}_\delta \) is the orthogonal projection onto \( H_\delta \). The distribution \( \Theta_\delta \) has an analytic continuation to \( \Xi^o \), determined up to a constant by a spherical function ([ÔØ97a, ÔØ97b, ÔØ99]), with a well-known expansion formula in terms of elementary functions. On \( \Xi(C^o) \) we have then the identity \( \sum_\delta \Theta_\delta = \Theta_K \). It is still an open problem to evaluate this sum in general.

To calculate \( \Theta_K \) independently, in a series of lectures at the University of Poitiers in 1990 by B. Ørsted and one of the authors the problem of relating the Hardy spaces \( H_{c_o} \) and \( H_2(C) \) via a causal compactification \( G/H \to \tilde{S}_1 \), with \( \tilde{S}_1 \) the Bergman-Šilov boundary of \( G_1/K_1 \), was discussed. It was shown that for the Cayley type spaces those are actually isomorphic (modulo a double covering in some cases) ([ÔØ99]). In particular this result gives a formula for \( K(\cdot, \cdot) \) in terms of the well-known classical Cauchy-Szegö kernel. The above summation formula can then be interpreted as a \( G \)-equivariant decomposition of the classical Cauchy-Szegö kernel or a generalized Heine formula ([ÔØ99]).

Further results in this direction for special cases were obtained by K. Koufany and B. Ørsted ([KÔ96, KÔ97]), G.I. Ol’shanskı ([Ô95]) and V.F. Molchanov for \( SO(2, n)/SO(1, n) \) ([M97]). The Cayley type spaces were also studied via a Jordan algebra approach by M. Chadli in his thesis ([C96]).

One of the obstacles for obtaining general results in this direction was the lack of general theory for the causal compactification of compactly causal symmetric spaces. This was finally obtained in [B97]. It was shown that there is a natural causal compactification of a “central extension” of most of those spaces in the Bergman-Silov boundary of the bounded realization of a tube type domain \( T_\Omega \).

In this paper we extend the results form [B97] to the domain \( \Xi(C^o) \) and show that there is a holomorphic function \( \psi_m : G_1/K_1 \to \mathbb{C} \) such that

\[ \Xi(C^o) = \{ z \in G_1/K_1 \mid \psi_m(z) \neq 0 \}. \]

The map \( I \) from the classical Hardy space into \( H_2(C) \) is then given by \( f \mapsto f \sqrt[\psi_m]{\psi_m} \). In general this has only meaning on a \( m \)-fold covering of \( \Xi(C^o) \). In that case we get an isomorphism into a Hardy space \( \tilde{H}_{2,\text{odd}}(C) \) on this covering of \( \Xi(C^o) \). The image is a direct sum of highest weight moduls.
We describe the image in terms of the lowest $K$-type of the highest-weight modules.

We remark that similar results to ours can also be obtained with Jordan algebra methods ([Bal]).

The paper is organized as follows. In the first section we recall some structure theory for causal symmetric spaces and recollect the facts needed by us from [B97] on causal compactifications.

Section 3 is devoted to the description of the Hardy space $H^2(C)$ and here we introduce the semigroup $\Gamma$ and the domain $\Xi(C^o)$.

We construct the function $\psi_m$ in Section 4. In Section 5 we give an isometry from $L^2(\mathcal{S}_1)$ onto $L^2(G/H)$, and construct the covering $\Xi(C^o) \to \Xi(C^o)$ of $\Xi(C^o)$ corresponding to the $m$-th root of $\psi_m$.

In Section 6 we show that that $\Xi(C^o) = \{z \in G_1/K_1 \mid \psi_m(z) \neq 0\}$. This gives us the necessary tool to analyze the covering $\Xi(C^o)$ in more detail in the next section. In particular we show that we need at most a double covering for describing the Hardy spaces. The final result of this paper is Theorem 7.8 where we characterize the image of $I$.

The authors would like to thank the Institute Mittag-Leffler for their hospitality during their stay there in the Spring of 1996. The first author would like to thank Louisiana State University and the second author for their hospitality during his stay in September, 1996.

2. Causal compactifications.

In this section we collect some standard facts on symmetric spaces. We also collect some newer results on causal compactifications. We use the monograph [HÓ96] and the orginal papers [ÓØ88, Ó91, B97] as references. We call $(G,H,\tau)$ a symmetric space, when $G$ is a connected Lie group, $\tau : G \to G$ is an involutive automorphism, and $H \subset G$ a closed subgroup with

$$(G^\tau)_o \subset H \subset G^\tau,$$

where $G^\tau := \{a \in G \mid \tau(a) = a\}$ and the subscript $o$ denotes the connected component containing the identity. By abuse of notation we then also call $G/H$ a symmetric space. A symmetric Lie algebra is a triple $(\mathfrak{g},\mathfrak{h},\tau)$, where $\mathfrak{g}$ is a Lie algebra, $\tau : \mathfrak{g} \to \mathfrak{g}$ is an involutive automorphism, and $\mathfrak{h} = \{X \in \mathfrak{g} \mid \tau(X) = X\}$ the subalgebra of $\tau$-fixed elements. To every symmetric space $(G,H,\tau)$ there is associated the symmetric algebra $(\mathfrak{g},\mathfrak{h},\tau^\prime) := (\text{Lie}(G),\text{Lie}(H),d\tau)$. In the sequel we will denote the differential of $\tau$ (and similarly the differentials of all other group homomorphisms) always by the same letter. The symmetric algebra $(\mathfrak{g},\mathfrak{h},\tau)$ is irreducible if there is no nontrivial $\tau$-stable ideal of $\mathfrak{g}$ not contained in $\mathfrak{h}$. The symmetric algebra $(\mathfrak{g},\mathfrak{h},\tau)$ is called reductive respectively semisimple if $\mathfrak{g}$ is reductive respectively semisimple.
Let \((\mathfrak{g}, \mathfrak{h}, \tau)\) be a reductive symmetric algebra. Choose a Cartan involution \(\theta\) of \(\mathfrak{g}\) commuting with \(\tau\) ([H78, p. 192] or [Ó84, Lemma 2.1]). Let \(\mathfrak{q} := \{X \in \mathfrak{g} \mid \tau(X) = -X\}\), \(\mathfrak{t} := \{X \in \mathfrak{g} \mid \theta(X) = X\}\), and \(\mathfrak{p} := \{X \in \mathfrak{g} \mid \theta(X) = -X\}\), then
\[
\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{q} = \mathfrak{t} \oplus \mathfrak{p} = \mathfrak{h}_k \oplus \mathfrak{h}_p \oplus \mathfrak{q}_k \oplus \mathfrak{q}_p,
\]
where an index denotes the intersection with the corresponding subspace, i.e., \(\mathfrak{h}_p := \mathfrak{h} \cap \mathfrak{p}\), etc.

Let \(\mathfrak{g}_s := [\mathfrak{g}, \mathfrak{g}]\) denote the semisimple part of \(\mathfrak{g}\) and \(\mathfrak{z}\) the center of \(\mathfrak{g}\). If \(\mathfrak{l}\) is a subspace of \(\mathfrak{g}\) such that \(\mathfrak{l} = \mathfrak{l} \cap \mathfrak{z} \oplus \mathfrak{l} \cap \mathfrak{g}_s\), then we set \(\mathfrak{l}_z := \mathfrak{l} \cap \mathfrak{z}\), and \(\mathfrak{l}_s := \mathfrak{l} \cap \mathfrak{g}_s\). For \((G, H, \tau)\) associated to \((\mathfrak{g}, \mathfrak{h}, \tau)\), let \(G_s\) be the analytic subgroup corresponding to \(\mathfrak{g}_s\), and \(Z := \exp \mathfrak{z}\) those corresponding to \(\mathfrak{z}\). Then \(G = ZG_s\) and \(D := Z \cap G_s\) is a discrete central subgroup in \(G_s\). Moreover we have a (right) action
\[
D \times (Z \times G_s) \to Z \times G_s, \quad (d, (z, g)) \mapsto (zd, d^{-1}g),
\]
such that
\[
G \simeq (Z \times G_s)/D =: Z \times_D G_s.
\]

Let \(E\) be a finite dimensional vector space over the reals. A subset \(C \subset E\) is called a cone if \(C\) is convex and closed under multiplication by \(\mathbb{R}^+\). The closed cone \(C\) is pointed if \(C \cap -C = \{0\}\) and generating if \(C - C = E\). A pointed generating cones is called regular. We remark that a cone \(C\) is generating if and only if its interior \(C^0\) is nonempty.

Definition. Assume that \((G, H, \tau)\) is a symmetric space with associated reductive symmetric Lie algebra \((\mathfrak{g}, \mathfrak{h}, \tau)\).

1) The symmetric space is called compactly causal if there exists a \(H\)-invariant regular cone \(C\) in \(\mathfrak{q}\) such that \(C^0 \cap \mathfrak{t} \neq \emptyset\).

2) \((\mathfrak{g}, \mathfrak{h}, \tau)\) is called compactly causal if \((G, (H^\tau)_0, \tau)\) is compactly causal.

3) The symmetric space is called noncompactly causal if there exists a \(H\)-invariant regular cone \(C \subset \mathfrak{q}\) such that \(C^0 \cap \mathfrak{p} \neq \emptyset\).

4) \((G, H, \tau)\) is called of Cayley type if it is semisimple and both compactly and noncompactly causal.

5) Let \(\mathfrak{e}(\mathfrak{q}_k) := \{X \in \mathfrak{q}_k \mid [\mathfrak{q}_k, X] = \{0\}\}\). The symmetric space respectively the associated algebra are called of weakly Hermitian type if \(\mathfrak{e}(\mathfrak{q}_k) = \mathfrak{q}_k\).

Notice that in this definition we do not assume that \(G\) is noncompact. Thus every compact symmetric space is weakly Hermitian. On the other hand a compact symmetric space can only be compactly causal if it is reductive ([HÓ96]). The spaces we consider will by construction be both compactly causal and of weakly Hermitian type. We refer to [KN96] and [KNÖ97] for a general discussion of the connection between weakly Hermitian and compactly causal.
Let \((\mathfrak{g}_1, \mathfrak{h}_1, \tau_1)\) be a compactly causal irreducible symmetric algebra with \(\mathfrak{g}_1\) simple and noncompact. Let \(G_{1\mathbb{C}}\) the simply connected complex Lie group with Lie algebra \(\mathfrak{g}_{1\mathbb{C}} := \mathfrak{g}_1 \otimes \mathbb{C}\). We extend \(\tau_1\) to a complex linear involution of \(\mathfrak{g}_{1\mathbb{C}}\), and denote this involution and the corresponding involution on \(G_{1\mathbb{C}}\) again by \(\tau_1\). Then, as \(G_{1\mathbb{C}}\) is assumed simply connected, it follows that \(H_{1\mathbb{C}} := G_{1\mathbb{C}}^{\tau_1}\) is connected ([L69, p. 171]). Let \(G_1\) be the analytic subgroup of \(G_{1\mathbb{C}}\) with Lie algebra \(\mathfrak{g}_1\) and

\[
H_1 := G_1 \cap H_{1\mathbb{C}} = G_1^{\tau_1}.
\]

Let \(\theta_1\) be a Cartan involution commuting with \(\tau_1\), and let \(\mathfrak{g}_1 = \mathfrak{t}_1 \oplus \mathfrak{p}_1\) be the corresponding Cartan decomposition. Recall that for \((\mathfrak{g}_1, \mathfrak{h}_1, \tau_1)\) irreducible compactly causal the Riemannian symmetric space \(G_1/K_1\) is Hermitian symmetric ([HÖ96, Remark 3.19]) and that \((\mathfrak{g}_1, \mathfrak{h}_1, \tau_1)\) is of weakly Hermitian type ([HÖ96, Lemma 1.2.1, Lemma 1.3.5]). We will assume that \(G_1/K_1\) is a type tube domain.

Let \(\mathfrak{t}_1\) be a Cartan subalgebra of \(\mathfrak{t}_1\). Let \(\Delta(\mathfrak{g}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}})\) be the set of roots of \(\mathfrak{t}_{1\mathbb{C}}\) in \(\mathfrak{g}_{1\mathbb{C}}\). We have the two subsets

\[
\Delta(\mathfrak{t}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}}) = \{\alpha \in \Delta(\mathfrak{g}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}}) \mid \mathfrak{g}_{1\mathbb{C}, \alpha} \subset \mathfrak{t}_{1\mathbb{C}}\}
\]

of compact and

\[
\Delta(\mathfrak{p}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}}) = \{\alpha \in \Delta(\mathfrak{g}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}}) \mid \mathfrak{g}_{1\mathbb{C}, \alpha} \subset \mathfrak{p}_{1\mathbb{C}}\}
\]

of noncompact roots. We choose an ordering in \(i\mathfrak{t}_1^*\) such that the positive noncompact dominate the positive compact roots and define

\[
\mathfrak{p}_1^+ := \sum_{\alpha \in \Delta^+(\mathfrak{p}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}})} \mathfrak{g}_{1\mathbb{C}, \alpha} \quad \text{respectively} \quad \mathfrak{p}_1^- := \sum_{\alpha \in \Delta^+(\mathfrak{p}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}})} \mathfrak{g}_{1\mathbb{C}, -\alpha}.
\]

Recall that two roots \(\alpha\) and \(\beta\) are strongly orthogonal if \(\alpha \neq \pm \beta\) and \(\alpha \pm \beta \notin \Delta(\mathfrak{g}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}})\). Let \(\{\gamma_1, \ldots, \gamma_r\} \subset \Delta^+(\mathfrak{p}_{1\mathbb{C}}, \mathfrak{t}_{1\mathbb{C}})\) be a maximal system of strongly orthogonal roots. Let \(H_j \in [\mathfrak{g}_{1\mathbb{C}, \gamma_j}, \mathfrak{g}_{1\mathbb{C}, -\gamma_j}] \cap i\mathfrak{t}_1\) be such that \(\gamma_j(H_j) = 2\delta_{i,j}\). Choose \(E_{\pm j} \in \mathfrak{g}_{1\mathbb{C}, \pm \gamma_j}\) such that

\[
X_j := E_j + E_{-j}, \quad Y_j := iE_j - iE_{-j} \in \mathfrak{p}_1,
\]

and

\[
[E_j, E_{-j}] = H_j.
\]

It is known that \(a_p := \sum \mathbb{R}X_j\) is maximal abelian in \(\mathfrak{p}_1\) ([H78, p. 387]). Let

\[
X^0 := \sum X_j, \quad Y^0 := \sum Y_j, \quad \text{and} \quad Z^0 := -\frac{1}{2} \sum iH_j,
\]

then \(Z^0 \in \mathfrak{z}(\mathfrak{t}_1)\), with \(\mathfrak{z}(\mathfrak{t}_1)\) the center of \(\mathfrak{t}_1\), and \(\text{ad } X^0\) has eigenvalues 0, 2 and \(-2\). Let

\[
\eta := \text{Ad} \left(\exp \frac{\pi}{2} iX^0\right)
\]
be the involution on $G_{1\mathbb{C}}$ respectively $g_{1\mathbb{C}}$ given by conjugation with 
$\exp \frac{\pi}{2} iX^0$. Then the symmetric space $G_1/G_1^0$ is of Cayley type and

$$q_{1\eta} := g_{1\eta}^{-\eta} = \{ X \in g_1 \mid \eta(X) = -X \}$$

is the direct sum of the two eigenspaces $q_{1\eta}^{+} := \{ X \in q_{1\eta} \mid [X^0, X] = \pm 2X \}$ of 
ad $X^0$, which are also $G_1^0$-invariant. As $\theta_1(X^0) = -X^0$ we get $\theta_1(q_{1\eta}^{+}) = q_{1\eta}^{-}$.

Define $Y_{\pm} \in q_{1\eta}^{+}$ by $Y^0 = Y_+ + Y_-$. The sets $C_{\pm} := \text{Ad}(G_{1\eta})Y_{\pm}$ are regular 
cones in $q_{1\eta}^{\pm}$ and the cone $C_k := C_+ - C_- \subset q_{1\eta}$ defines the compactly causal 
structure on $G_1/G_1^0$. Let $p'$ be the sum of the $0$- and $(-2)$-eigenspaces of 
ad $X^0$. Then $P' := N_{G_1}(p')$ is a parabolic subgroup and

$$\tilde{S}_1 := G_1/P' = K_1/K_1^0$$

is the Bergman-Šilov boundary of $G_1/K_1$. The compact symmetric space $\tilde{S}_1$ is causal with causal structure defined by $C_+ \subset q_{1\eta}^{+} \cong T_eP\tilde{S}_1$. Moreover

$$G_1^0 = Z_{G_1}(X^0) \subset P',$$

and the canonical projection $\Phi_1 : G_1/G_1^0 \to \tilde{S}_1$ is causal.

Under conditions specified in [B97] there exists an involution $\sigma$, commuting with $\tau_1$, $\theta_1$, and $\eta$, such that

$$(G, H, \tau) := ((G_1^0)_o, H_1 \cap (G_1^0)_o, \tau_1 | (G_1^0)_o)$$

is a compactly causal symmetric subspace of $(G_1, G_1^0, \eta)$ and $\Phi_1$ can be used to define a causal compactification of it. (For spaces of Cayley type $(g_1, h_1, \tau_1)$ in the preceding discussion has to be replaced by $(g_1 \times g_1, h_1 \times \tau_1 \times \tau_1)$ and $\sigma(g, h) = (h, g)$.) For the general theory we will not need the explicit form of $\sigma$ but only the “axiomatic” properties of the construction that we collect in the remaining part of this section.

Define $\theta = \theta_1|_G$ then $\theta$ is a Cartan involution on $G$ commuting with $\tau$. We denote the corresponding Cartan decomposition by $g = t \oplus p$. Let $g_1 = g + q_{1\sigma}$ be the eigenspace decomposition of $\sigma$. We have $\eta|_G = \tau_1|_G$ — in fact, in many cases $\eta = \tau_1$ — such that $G \cap G_1^0 = G \cap H_1$ and also $q := q_1 \cap g = q_{1\eta} \cap g$.

**Lemma 2.1.** With notation as above the following holds:

1. The algebra $b := a_\sigma \cap q_{1\sigma}$ is maximal abelian in $p_1 \cap q_{1\sigma}$, $X^0 \in b$,
   $Y^0 \in p_1 \cap q_{1\sigma}$, and $Z^0 \in t \cap q$.
2. The cone $C_k$ is $\sigma$-invariant.
3. Let $\text{pr}_q : q_{1\eta} \to q$ be the projection with respect to the decomposition $q_{1\eta} = q \oplus q_{1\eta} \cap q_{1\sigma}$. Then $C_k \cap q = \text{pr}_q(C_k)$ and $(\text{pr}_q(C_k))^0 \cap t \neq \emptyset$.

**Proof.** (1) With $Z^0 = \frac{1}{2}[X^0, Y^0]$ the last claim follows from the preceding ones. These are prerequisites for the construction of $\sigma$ in [B97, Theorem 5.1] respectively [B97, Theorem 5.9].
(2) From
\[ Y_+ + Y_- = Y^0 = -\sigma(Y^0) = -\sigma(Y_+) - \sigma(Y_-) \]
and
\[ \sigma(Y_+) = \sigma \left( \frac{\pm 1}{2}[X^0, Y_+] \right) = \frac{\pm 1}{2}[X^0, \sigma(Y_-)] \]
we have \( \sigma(Y_+) = -Y_+ \). By the \( \sigma \)-invariance of \( G_1^n \) this implies the claim together with the definition of \( C_k \).

(3) The \( \sigma \)-invariance of \( C_k \) implies that \( C_k \cap q = \phi_q(C_k) \). That \( (\phi_q(C_k))^o \cap \mathfrak{f} \) is nonempty follows since \( Z^0 \in C_k^n \).

By the last part of the Lemma \( C := C_k \cap q \) is a \( H \)-invariant regular cone in \( q \) which defines a compactly causal structure for \( (G, H, \tau) \). It is clear that this definition makes \( (G, H, \tau) \) into a causal subspace of \( (G_1, G_1^n, \eta) \).

**Corollary 2.2.** \( C + p' = C_k + p' = C_+ \oplus p' \).

**Proof.** By definition we have
\[ C + p' \subset C + p' = C_+ \oplus p'. \]
When we write \( X \in C_+ \) in the form \((X + \sigma(X)) - \sigma(X)\), where \( X + \sigma(X) \in C_k \cap q \) and \( \sigma(X) \in \sigma(q_1^n) \subset p' \), the other inclusion is also obvious. \( \square \)

**Definition (\([\tilde{\text{O}}099]\)).** Let \( M \) and \( N \) be manifolds, with \( N \) compact, and with causal structure \( M \ni c \rightarrow C(m) \subset T_m M \) respectively \( N \ni n \rightarrow D(n) \subset T_n N \). Let \( \Phi : M \rightarrow N \) be smooth. The pair \( (N, \Phi) \) is called a causal compactification of \( M \) if \( \Phi \) is a diffeomorphism onto an open dense subset of \( N \), and \( (d\Phi)_m(C(m)) = D(\Phi(m)) \) for all \( m \in M \). Let \( L \) be a Lie group acting on \( M \) and \( N \) such that the causal structures are \( L \)-invariant. Then the causal compactification is \( L \)-equivariant if \( \Phi \) is \( L \)-equivariant.

**Theorem 2.3.** The canonical inclusion \( i : G/H \hookrightarrow G_1/G_1^n \) is causal and the map \( \Phi := \Phi_1 \circ i : G/H \hookrightarrow \hat{S}_1 \) is a \( G \)-equivariant causal compactification of \( G/H \).

**Proof.** By \([B97]\) we know that \( \Phi : G/H \rightarrow \hat{S}_1 \) is an injective \( G \)-map with open and dense image. With the canonical identifications \( T_{eH}G/H \simeq q_1 \), \( T_{eG_1^n}G/G_1^n \simeq q_1^n \), and \( T_{e\hat{S}_1}(\hat{S}_1) \simeq q_{1\eta} \), the tangent map \( d\Phi \) at the identity coset is given by first imbedding \( q \) into \( q_1^n \eta \) and then projecting onto \( q_{1\eta} \). Thus \( d\Phi_{eH}(C) = C_+ \) by Corollary 2.2. By the \( G \)-invariance of the causal structures on \( G/H \) respectively \( \hat{S}_1 \) this is all to show. \( \square \)

As all groups are subgroups of \( G_{1C} \) it is immediate to see that \( \Phi \) extends to a \( G_C \)-invariant map \( G_C/H_C \hookrightarrow G_{1C}/P'_C \), which we denote again by \( \Phi \).
Example. Let $G_1 = SU(1, 1)$. Denote by $E_{i,j}$ the matrix with entry 1 in the $i$-th row and $j$-th column and otherwise zero. We choose $E_1 = E_{1,2}$, $E_{-1} = E_{2,1}$, and $H_1 = E_{1,1} - E_{2,2}$. Then

$$X^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{and} \quad Y^0 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$  

Thus $\eta$ is given by conjugation by $iX^0$. This is easily seen to be the same involution on $SU(1, 1)$ as $g \mapsto \overline{g}$, the complex conjugation. We will also need the Cayley transform, which is given by conjugation by

$$c := \exp \frac{\pi}{4} i Y^0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$  

As is well-known, the unit disk — identifying $\mathbb{C} \simeq \mathbb{C}E_1 = p_1^\perp$ — is the Harish-Chandra realization of $G_1/K_1$, where $G_1$ operates by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d'} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SU(1, 1).$$  

The Bergman–Šilov boundary $\hat{S}_1$ is now the unit circle and the Cayley transform $z \mapsto c \cdot z$ maps the origin to $-1 \in \hat{S}_1$. We notice that $\text{Ad}(c)H_1 = X^0$. It follows in particular that $(\text{Ad}(c) (t_1 + p_1^\perp)) \cap g_1 = p'$ and $\hat{S}_1 = G_1/P' \simeq G_1 \cdot (-1)$. With $\hat{S}_1$ in the Harish-Chandra realization the map $\Phi_1$ is therefore given by $gG_1 \eta \mapsto g \cdot (-1)$.

We can choose the involution $\sigma = \theta_1$, i.e.,

$$(G, H) = \left\{ \begin{pmatrix} a & 0 \\ 0 & \alpha \end{pmatrix} \bigg| \ a \in S^1 \right\}, \{\pm \text{id}\} \right),$$

and then

$$\Phi_1 \left( \begin{pmatrix} a & 0 \\ 0 & \alpha \end{pmatrix} H \right) = -a^2.$$  

We restrict us in the following to causal compactifications constructed as in the Theorem. (Indeed, as is shown in [B97], with one natural additional assumption all causal compactifications in the Bergman–Šilov boundary are of this form.) A list of the possible compactifications can be found in [B97], cf. also below.

For short we call $(A, B, C)$ a $\mathfrak{sl}(2)$-triple if

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \mapsto A, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \mapsto B, \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \mapsto C$$

defines an isomorphism from $\mathfrak{sl}(2, \mathbb{C})$ onto the complex Lie algebra generated by $\{A, B, C\}$. 
Proposition 2.4. Let the notation be as above. Then the following holds:

1. There exist $\lambda_{j', j} \in \{0, 1\}$ such that with
   \[ E_{\pm j'} := \sum_j \lambda_{j', j} E_{\pm j}, \]
   \[ H_{j'} := [E_{j'}, E'_{-j'}] = \sum_j \lambda_{j', j} H_j \]
   and
   \[ X_{j'} := E_{j'} + E'_{-j'} = \sum_j \lambda_{j', j} X_j \]
   we have $b = \sum \mathbb{R} X_{j'}$, and the triples $\{H_{j'}, E_{j'}, E'_{-j'}\}$ are pairwise commuting $\mathfrak{sl}(2)$-triples.

2. With $c := \exp \left( \frac{\pi}{4} Y^0 \right)$, the Cayley-transformed space $a := \text{Ad}(c)ib$ is maximal abelian in $\mathfrak{k} \cap \mathfrak{q}$.

3. Let $\tau := \sum \mathbb{R} i H_j$ then we have $\tau_{-j} = a + \tau_{-j} \cap \mathfrak{q}_1$.

4. We have $\mathfrak{j} \subset \mathfrak{k}$ and, if $\mathfrak{g}$ is noncompact, then $(\mathfrak{g}, \mathfrak{h}, \tau)$, is weakly Hermitian. If $l \subset \mathfrak{h}$ is an ideal of $\mathfrak{g}$ then $l \subset \mathfrak{k}$.

5. $C^0 = \text{Ad}(H)(a \cap C^0)$ and $C = \text{Ad}(H)(a \cap C)$.

6. $a \cap C^0 = -\sum \mathbb{R}^+ i H_{j'}$ and $a \cap C = -\sum \mathbb{R} \pm i H_{j'}$.

Proof. (1) This is a direct consequence of [B97, Theorem 5.1] respectively [B97, Theorem 5.9].

(2) The inner automorphism $\text{Ad}(c)$ maps the subalgebra
   \[ g_a := \mathfrak{k} + i(p_1 \cap q_{1\sigma}) \]
   onto itself, since $Y^0 \in \mathfrak{p}_1 \cap q_{1\sigma}$ by Lemma 2.1.

We consider first the causal compactifications described by [B97, Theorem 5.1]. Here we have $\tau_1 = \eta$ and $b \subset \mathfrak{h}_1$. Then $Y^0 = [X^0, Z^0] \in \mathfrak{p}_1 \cap \mathfrak{q}_1$ is $\tau_1\eta$-fixed and consequently $\mathfrak{k} \cap \mathfrak{q} + i(\mathfrak{p}_1 \cap q_{1\sigma} \cap \mathfrak{h}_1)$, the $(-1)$-eigenspace of the involution $\tau_1\eta$ in $g_a$, is $\text{Ad}(c)$-invariant. Using $\theta_1 = \text{Ad}(\exp \pi Z^0)$, by a $\mathfrak{sl}(2)$-calculation $\text{Ad}(c)\theta_1 = \tau_1\text{Ad}(c)$. Looking at the corresponding eigenspaces in $\mathfrak{k} \cap \mathfrak{q} + i(\mathfrak{p}_1 \cap q_{1\sigma} \cap \mathfrak{h}_1)$ we get $\text{Ad}(c)i(\mathfrak{p}_1 \cap q_{1\sigma} \cap \mathfrak{h}_1) = \mathfrak{k} \cap \mathfrak{q}$. As $b \subset \mathfrak{p}_1 \cap q_{1\sigma} \cap \mathfrak{h}_1$, the claim for this case then follows.

For the compactifications given by [B97, Theorem 5.9] we have $b \subset \mathfrak{q}_1$. Similarly to the first case we get successively $Y^0 \in \mathfrak{h}_1$, the $\text{Ad}(c)$-invariance of $\mathfrak{k} \cap \mathfrak{q} + i(\mathfrak{p}_1 \cap q_{1\sigma} \cap \mathfrak{q}_1)$, and $\text{Ad}(c)\theta_1\tau_1 = \sigma\text{Ad}(c)$, where $\sigma := \eta \circ \tau_1$ as in [B97, Thm. 5.9] defined. Therefore $\text{Ad}(c)$ maps the $(+1)$-eigenspace of $\theta_1\tau_1$ onto the $(+1)$-eigenspace of $\sigma$, i.e., $\text{Ad}(c)i(\mathfrak{p}_1 \cap q_{1\sigma} \cap \mathfrak{q}_1) = \mathfrak{k} \cap \mathfrak{q}$. Now the stated result for this case follows again from Lemma 2.1.

(3) For the orthogonal sum $a_p = b \oplus b^\perp$ (with respect to the inner product $(X, Y) = -B(X, \theta(Y))$) we have by Lemma 2.1 that $b^\perp \subset \mathfrak{p}_1 \cap \mathfrak{g}$. Now $\tau_{-1} = \text{Ad}(c)\tau_{-1} = a \oplus \text{Ad}(c)b^\perp$. The second summand is contained in $\mathfrak{t}_1$, since $Z^0 \in a$. But the Cayley transform maps with $g_a = \mathfrak{k} + i(\mathfrak{p}_1 \cap q_{1\sigma})$ also
the orthocomplement \( g_a^\perp \) of \( g_a \) in \( \mathfrak{t}_1 + i\mathfrak{p}_1 \) onto itself. With \( ib^\perp \subset g_a^\perp \) we have therefore \( \text{Ad}(c) ib^\perp \subset g_a^\perp \cap \mathfrak{t}_1 = \mathfrak{t}_1 \cap q_{1\sigma} \).

(4) As \( Z^0 \in g \) by Lemma 2.1, and \( \text{ad} Z^0|_{\mathfrak{p}_1} \) is regular, it follows that \( \mathfrak{z}_g(Z^0) = \mathfrak{t} \). Hence

\[
\mathfrak{z} \subset \mathfrak{z}_g(Z^0) = \mathfrak{t}.
\]

Write \( Z^0 = Z_1 + Z_2 \), with \( Z_1 \in \mathfrak{z}(g) \), and \( Z_2 \in [g, g] \). Assume that \( Z_2 = 0 \). Then the same argument shows that \( g = \mathfrak{z}_g(Z^0) = \mathfrak{t} \). Hence \( g \) is compact. Thus, if \( g \) is not compact it follows that \( Z_2 \neq 0 \). Furthermore \( \mathfrak{z}_g(Z_2) = \mathfrak{t} \), as \( Z_1 \) commutes with \( g \). In particular it follows that \( Z_2 \in c(q_k) \) and therefore

\[
\mathfrak{z}_q(c(q_k)) \subset \mathfrak{z}_q(Z_2) = q_k.
\]

Suppose that \( \mathfrak{l} \) is an ideal in \( g \) contained in \( \mathfrak{h} \). As \( Z_2 \in \mathfrak{q} \) it follows that \( [Z_2, \mathfrak{l}] = \{0\} \). Hence

\[
\mathfrak{l} \subset \mathfrak{z}_g(Z_2) = \mathfrak{t}
\]

which shows that \( \mathfrak{l} \) is compact.

(5) Let \( X \in C^o = (C_k^o \cap g) \). Then \( G^X = \{ a \in G \mid \text{Ad}(a) X = X \} \) is compact. Therefore the elements in \( g^X \) are semisimple and with imaginary eigenvalues (see the proof of [HÓ96, Thm. 4.2.15]). Hence there is a \( h \in H \) such that \( \text{Ad}(h) X \in q_k \) ([M79, Thm. 1]). Let \( a_q \) be a maximal abelian algebra in \( q_k \) containing \( X \), then (by [M79, Lemma 7]) we find \( k \in H \cap K \) with \( \text{Ad}(k) a_q = a \). In particular \( \text{Ad}(kh) X \in a \). Therefore \( C^o \subset \text{Ad}(H)(a \cap C^o) \). As the other inclusion is obvious we get \( C^o = \text{Ad}(H)(a \cap C^o) \). Since the eigenvalues of \( \text{ad} X \) are continuous in \( X \), the same argument apply to \( X \) in the boundary of \( C \). Thus the assertion for the whole cone follows.

(6) By [ÔO99, Lemma 4.4] or [HÓ96, Chap. 4.2] we have \( \mathfrak{t}_1^- \cap C_k^o = - \sum \mathbb{R}^+ iH_j \) respectively \( \mathfrak{t}_1^- \cap C_k = - \sum \mathbb{R}^+ iH_j \). (Note that we have here a minus sign by our choice of \( Z^0 \).) The claims then follow with \( \mathfrak{t}_1^- \cap a = \sum \mathbb{R}^+ iH_j \). \( \square \)

**Remarks.**

(1) The \( \lambda_{j',j} \) are known by [B97, Thm. 5.1, Thm. 5.9] for every case. Either \( \lambda_{j',j} = \delta_{j',j} \), for \( j' = 1, \ldots , r \), or \( r = 2r' \) and \( \lambda_{j',j} = \delta_{j',j} + \delta_{j'+r',j} \), for \( j' = 1, \ldots , r' \).

(2) By the classification of causal compactifications the only \( (g, \mathfrak{h}, \tau) \) with nontrivial center are those described by [B97, Corollary 5.3].

### 3. Hardy spaces.

We will from now on allways assume that \( g \) is noncompact. Proposition 2.4(4) asserts that \( (g_a, \mathfrak{h}_a, \tau|_{g_a}) \) is a symmetric Lie algebra of Hermitian type in the sense of [ÔO88, Ô91, HÓÔ91] where we have a theory of Hardy spaces. In this section we generalize this to our reductive setting. Let \( \mathfrak{t} \) be a Cartan algebra of \( g \), with \( c(q_k) \subset a \subset \mathfrak{t} \subset \mathfrak{t} \), and \( \Delta(g_C, \mathfrak{t}_C) \) respectively \( \Delta := \Delta(g_C, a_C) \) the corresponding root systems. We may assume without loss of generality...
Indeed, if \( t_1^- \cap q_1 = \{0\} \) by Proposition 2.4 then \([t, t_1^-] = \{0\}\) and (since the constructions in the last section needed only \( t_1^- \)) we can assume \( t + t_1^- \subset t_1 \). For the remaining (four) cases this is immediate from the realization of their causal compactifications ([B97]). We assume all introduced root systems to be lexicographically ordered with respect to the flag \( c(q_k) \subset a \subset t \subset t + t_1^- \subset t_1 \). Especially we have

\[
p^\pm := \bigoplus_{\alpha \in \Delta^+(p, t_1^-)} g_{c, \pm \alpha} \subset p_1^+.
\]

Let \( \Delta_n := \Delta(p, a), \Delta_n^+ := \Delta(p^+, a), \Delta_k := \Delta(t, a) \), etc. Let \( H_\alpha \in ia \) denote the co-root determined by the two conditions

\[
H_\alpha \in [g_{c, \alpha}, g_{c, -\alpha}] \cap ia \quad \text{and} \quad \alpha(H_\alpha) = 2.
\]

Assume for the moment that \((g, h, \tau | g_\mathfrak{g})\) is irreducible. Recall that for this case in \( q \) we have the minimal closed cone \( C_{\min} \) and maximal closed cone \( C_{\max} \) given by

\[
C_{\min} = \text{conv}\text{Ad}(H)Z_2 = C_{\min, s} \subset g_\mathfrak{s}
\]

and

\[
C_{\max} = C_{\min}^* = \mathfrak{z} + C_{\max, s}
\]

where the subscript \( s \) indicates the corresponding cone in \( g_\mathfrak{s} \). We have by \([H\mathring{\text{O}}96, \text{Chap. 4.2}]\)

\[
c_{\min} := C_{\min} \cap a = -i \sum_{\alpha \in \Delta_n^+} \mathbb{R}^+ H_\alpha
\]

and

\[
c_{\max} := C_{\max} \cap a = \{H \in a | \forall \alpha \in \Delta_n^+: -i\alpha(H) \geq 0\}.
\]

**Lemma 3.1.** Denote by \( \text{pr}_z \) respectively \( \text{pr}_s \) the projection onto \( \mathfrak{z} \) respectively \( \mathfrak{g}_s \) corresponding to the decomposition \( \mathfrak{g} = \mathfrak{z} \oplus \mathfrak{g}_s \). Then the following holds:

1. \( \text{pr}_z(C) = \mathfrak{z} \).
2. \( C_s := \text{pr}_s(C) \) is a regular \( H_\mathfrak{s} \)-invariant cone in \( q_\mathfrak{s} \). In particular \( C_{\min, s} \subset \text{pr}_s(C) \subset C_{\max, s} \).

**Proof.** (1) If \( \mathfrak{z} \neq \{0\} \) the center is spanned by \( Z_1' = \sum \epsilon_j iH_j \), with \( \epsilon_1 = \ldots = \epsilon_p = 1 = -\epsilon_{p+1} = \ldots = -\epsilon_r \) ([B97, Prop. 5.5]). In this case we also have \( \lambda_{j', j} = \delta_{j', j} \). We get from Proposition 2.4(6) for elements in \( a \cap C^0 \)

\[
\sum a_j iH_j = aZ_1' + X_s,
\]

where \( a_j < 0 \) and \( X_s \in g_\mathfrak{s} \). Since both sums are orthogonal it follows \( a = \frac{1}{r} \sum \epsilon_j a_j \), which can take any value.

(2) Obviously \( C_s \) is a convex generating \( H_\mathfrak{s} \)-invariant cone in \( q_\mathfrak{s} \) because \( C \) is a convex generating \( H_\mathfrak{s} \)-invariant cone in \( q \). Assume that \( C_s \) is not
pointed. Let $q_2 := C_s \cap -C_s$. Then $g_2 := [q_2, q_2] \oplus q_2$ is a subalgebra of $g_s$. As $C$ is $\theta$-stable it follows that $g_2$ is $\theta$-stable and hence reductive. We have $[h_s, g_2] \subset g_2$ because $C$ is $H_s$-stable and because of the Jacobi identity. Let $q_2^\perp$ be the orthogonal complement to $q_2$ (with respect to the inner product $(X, Y) = -B(X, \theta(Y))$). Let $Z \in q_2$ and $Y \in q_2^\perp$. Then

$$-B([Z, Y], \theta([Z, Y])) = -B(Y, [Z, \theta(Z), \theta(Y)]) = 0$$

because $[\theta(Z), \theta(Y)] \in h_s$. It follows that $g_2$ is a $\tau$-stable ideal in $g_s$. But $(g_s, h_s, \tau|_{g_s})$ is by [B97, Prop. 5.5] irreducible, hence $q_2 = q_s$. It follows in particular that $\pm H_\alpha \in C_s$ for all $\alpha \in \Delta_n$. Choose a Lie algebra homomorphism

$$\varphi_\alpha : su(1, 1) \rightarrow g_s$$

such that

$$\varphi_\alpha \left( \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \right) = i H_\alpha$$

and

$$T_\alpha := \varphi_\alpha \left( \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \in h_s$$

([Ö88, p. 134]). Let $h_t := \exp tT_\alpha$ then a simple $SU(1,1)$-calculation shows that

$$\text{Ad}(h_t)H_\alpha + \text{Ad}(h_{-t})H_\alpha = 2 \cosh(2t)H_\alpha.$$ 

Choose $Z \in \mathfrak{z}$ such that $Z + H_\alpha \in C$. Then

$$\lim_{t \to \infty} \left( \frac{1}{\cosh(2t)} \left( \text{Ad}(h_t)(Z + H_\alpha) + \text{Ad}(h_{-t})(Z + H_\alpha) \right) \right) = 2H_\alpha \in C.$$ 

Similarly one shows that $-2H_\alpha \in C$. This contradicts the fact, that $C$ is pointed. Hence $C_s$ is also pointed. 

The cone $C_k$ is minimal in $q_{1\eta}$ and is generated by $\text{Ad}(G_1^\eta)Z^0$. A minimal extension of $C_k$ to a $G_1$-invariant cone in $g_1$ is $W_k$, the minimal cone in $g_1$ generated by $\text{Ad}(G_1)Z^0$ ([HÖ96, Chap. 4.2]). As $Z^0 \in q_\eta \cap \mathfrak{t}$ it follows that $W_k$ is $\sigma$-invariant. For a subset $D \subset W_k$ we define

$$\Gamma_1(D) := G_1 \exp iD \subset G_{1\mathbb{C}}.$$ 

It is known that the Ol'shanskiï semigroup $\Gamma_1(W_k)$ is a closed semigroup in $G_{1\mathbb{C}}$ and that

$$\Gamma_1(W_k)^o = \Gamma_1(W_k^o) \simeq G_1 \times iW_k^o,$$

where the diffeomorphism is given by $(g, iX) \mapsto g \exp iX$.

**Theorem 3.2.** For $W := W_k \cap g$ the following holds:

1. $W = W_k^c = \text{pr}_g(W_k)$, where $\text{pr}_g : g_1 \rightarrow g$ denotes the orthogonal projection.
(2) $W$ is a regular $G$-invariant cone in $\mathfrak{g}$ such that $W \cap q = \text{pr}_q(W) = C$,
where $\text{pr}_q : \mathfrak{g} \to q$ denotes the orthogonal projection, i.e., $W$ is a
$G$-invariant extension of $C$.

(3) $W^o = \text{Ad}(G)(t \cap W^o)$.

(4) Let $\Gamma(W) := G\exp iW$. Then

$$\Gamma(W) = \Gamma_1(W^o) = \Gamma_1(W_k) \cap G_{1,C}^\sigma.$$ 

Thus $\Gamma(W)$ is a closed semigroup in $G_C := G_{1,C}^\sigma$.

(5) $\Gamma(W)^o = G\exp iW^o = \Gamma(W^o)$, and

$$G \times iW^o \to \Gamma(W^o) \quad (g, iX) \mapsto g \exp iX$$

is a diffeomorphism.

Proof. (1), (2), and (3) follow in the same way as the corresponding claims
for the cones $C_k$ and $C$, using also the fact that $W_k$ is an extension of $C_k$.

(4) Obviously $\Gamma(W) \subset \Gamma_1(W^o)$, and $\Gamma(W) \subset \Gamma_1(W_k) \cap G_C$. Let $\gamma = g \exp iX \in \Gamma_1(W^o)$, $g \in G_1$ and $X \in W_k$. Then

$$\sigma(g) \exp i\sigma(X) = g \exp iX.$$ 

Because of the uniqueness of the decomposition in $\Gamma_1(W_k)$ we get $\sigma(g) = g$
and $\sigma(X) = X$. Hence $\Gamma(W) = \Gamma(W^o)$. Assume now that $\gamma = g \exp iX \in
G_C \cap \Gamma_1(W_k)$, $g \in G_1$, $X \in W_k$. Then the same argument shows that
$\sigma(g) = g$ and $\sigma(X) = X$. Hence $g \in G_{1,C}^\sigma$ and $X \in W_k^\sigma = W$. Hence
$\gamma \in \Gamma(W)$. It follows that $\Gamma(W)$ is a closed semigroup.

(5) These claims follow by (4) and the corresponding facts for $\Gamma_1(W_k)$. ∎

With these results it is straightforward to generalize the constructions of
the holomorphic discrete series for $G/H$ ([ÓØ88, ÓØ91]) and the Hardy
spaces ([HÔ96, Chap. 7]). Let

$$\rho := \frac{1}{2} \sum_{\alpha \in \Delta^+} \dim \mathfrak{g}_{C,\alpha} \alpha.$$ 

We have that $p^+$ and $p^-$ are abelian subalgebras, $\mathfrak{g}_C = p^+ \oplus \mathfrak{t}_C \oplus p^-$, and
$[\mathfrak{t}_C, p^\pm] \subset p^\pm$. With $P^\pm := \exp(p^\pm)$ we have $G \subset P^+ K_C P^- \cap H_C K_C P^+$, and
for $x \in P^+ K_C P^-$ respectively $H_C K_C P^+$ we define $p^\pm(x) \in P^\pm$, $k_C(x) \in K_C$, and
$k_H(x) \in K_C$ (only the class in $K_C \cap H_C \setminus K_C$ is well-defined) by

$$x = p^+(x)k_C(x)p^-(x)$$

respectively

$$x \in H_C k_H(x)P^+.$$
Let \((\delta, V_\delta)\) be a holomorphic irreducible representation of \(K_C\) with nonzero \(K_C^+\)-fixed vector \(\nu\) and \(\delta|K\) unitary. For \(\delta\) the contragredient representation, \(\nu^0 \in V_\delta^*\) a \(K_C^+\)-invariant vector with \(\langle \nu, \nu^0 \rangle = 1\), we define
\[
\Phi_\delta : P^+K_C H_C/H_C \to \mathbb{C}, \ gH \mapsto \langle \nu, \delta(k_H(g^{-1})^{-1})\nu^0 \rangle.
\]
This is a well-defined holomorphic function because \(\nu^0\) is \(K_C \cap H_C\)-invariant. We denote the restriction of \(\Phi_\delta\) to \(G/H\) simply by \(\Phi_\delta\). With this convention notice that \(\Phi_\delta| (G_s/H_s) = \Phi_{\delta_s}\) with \(\delta_s = \delta|(K_{s,C})\). Let \(Z_C\) be the complex torus given by \(\exp Z_C\). Define \(\chi_\delta : Z_C \to \mathbb{C}\) by
\[
\delta(z) = \chi_\delta(z) \text{id}.
\]

**Theorem 3.3 ([\OØ91]).** Let \(\mu\) be the highest weight of \(\delta\). Then the following holds:

1. The function \(\Phi_\delta\) is in \(L^2(G/H)\) if and only if \(\langle \mu + \rho, \alpha \rangle < 0\) for all \(\alpha \in \Delta_+^\delta\).
2. Suppose that \(\langle \mu + \rho, \alpha \rangle < 0\) for all \(\alpha \in \Delta_+^\delta\). Let \(H_\delta \subset L^2(G/H)\) be the \(G\)-module generated by \(\Phi_\delta\). Then \(H_\delta\) is an irreducible highest weight module.
3. Let \(C_\delta\) be the \(Z_C\)-module corresponding to \(\chi_\delta\) and let \(H_{\delta_s}\) be the \(G_s\)-module generated by \(\Phi_{\delta_s}\), then \(H_\delta \simeq C_\delta \otimes H_{\delta_s}\).
4. Let \(E \subset L^2(G/H)\) be an irreducible highest weight module. Then there exists a representation \(\delta\) such that \(E \simeq H_\delta\) and the multiplicity of \(E\) in \(L^2(G/H)\) is one.

**Proof.** (1) For \(g \in G\) choose \(z \in Z\) and \(g_s \in G_s\) such that \(g = zg_s\). Then \(\Phi_\delta(gH) = \chi_\delta(z)\Phi_{\delta_s}(g_sH_s)\), in particular
\[
|\Phi_\delta(gH)| = |\Phi_{\delta_s}(g_sH_s)|.
\]
Hence \(\Phi_\delta \in L^2(G/H)\) if \(\langle \mu + \rho, \alpha \rangle < 0\) for all \(\alpha \in \Delta_+^\delta\) by [\OØ91, Theorem 5.2] and [HØØ91, Theorem 3.3].

(2), (3) Denote the \(G_s\)-module generated by \(\Phi_{\delta_s}\) in \(L^2(G_s/H_s)\) by \(H_{\delta_s}\). Then
\[
H_\delta \simeq C_\chi_\delta \otimes H_{\delta_s}
\]
as a \(Z \times G_s\) module. The module on the right hand side is an irreducible highest weight module by [\OØ91, Theorem 5.2]. Hence \(H_\delta\) must be irreducible.

(4) Assume that \(E \subset L^2(G/H)\) is an irreducible highest weight module. Then \(E \simeq C_\chi \otimes E_s\), where \(E_s\) is the restriction of \(E\) to \(G_s\) and \(C_\chi\) is the central character of \(E\). The module \(E_s\) is an irreducible highest weight module for \(G_s\), and \(E_s \subset L^2(G_s/H_s)\). By [HØØ91, Theorem 3.3] it follows that \(E_s \simeq H_{\delta_s}\) for some irreducible \(H_{s,C} \cap K_{s,C}\)-spherical representation of \(K_{s,C}\). Let \(\delta = \chi \otimes \delta_s\). Then \(\delta\) is an irreducible \(K_C \cap H_C\)-spherical representation of \(K_C\) and \(E \simeq H_\delta\). The statement about the multiplicity follows now from [\OØ91, Theorem 7.2]. \(\square\)
These representations $\epsilon_\delta$ on $H_\delta$ form the holomorphic discrete series of $G/H$. Let $\pi$ be a unitary representation of $G$ on the Hilbert space $V$. Denote the space of smooth vectors by $V^\infty$. Denote by $\pi^\infty$ the derived representation of $\pi$ on $V^\infty$. Define the cone of negative elements by

$$C(\pi) := \{ X \in g \mid \forall u \in V^\infty : (i\pi^\infty(X)u \mid u) \leq 0 \}.$$  

Then $C(\pi)$ is a $G$-invariant cone in $g$. The representation $\pi$ is called $W$-admissible if $C(\pi) \supset W$. If $\pi$ is $W$-admissible then $\pi$ extends to a holomorphic representation of $\Gamma(W)$. (In fact, this is true for $W$ replaced by an arbitrary invariant cone.)

**Proposition 3.4.** For $W = W_k\cap g$ the representation $(\epsilon_\delta, H_\delta)$ is $W$-admissible if and only if

$$i\mu(a \cap C^0) = (i d\chi_\delta + i\mu_a)(a \cap C^0) \leq 0.$$  

*Proof.* Since $W^0 = \text{Ad}(G)(t \cap W^0)$ and $\epsilon_\delta$ is unitary, $W$-admissibility is equivalent to

\[(i \epsilon_\delta^\infty(t \cap W^0)u|u) \leq 0\]  

for all $u \in H_\delta^\infty$. Let $W(t, t)$ be the Weyl group of $\Delta(t, t)$. As $H_\delta$ is an irreducible highest weight module we have by Poincaré-Birkhoff-Witt $\mu_\lambda = \mu_+ - \sum_{\alpha \in \Delta^+(p_C, t_C)} n_\alpha \alpha$, with $n_\alpha \geq 0$ and $\mu_+ \in \text{conv}(W(t, t)\mu)$, for all occurring weights $\mu_\lambda$. Since $t \cap W \subset t_1 \cap W_k = -\sum_{\alpha \in \Delta^+(p_C, t_C)} \mathbb{R}_0^+ iH_\alpha$ (the $H_\alpha \in t_1C$ to $\alpha \in \Delta(g_1C, t_1C)$ defined as the $H_\alpha$ to $\alpha \in \Delta$) ([HÓ96, p. 102]) such that $i\alpha(t \cap W) \geq 0$ for all noncompact roots $\alpha \in \Delta^+(p_C, t_C)$, the condition $(\ast)$ remains to be checked for the vectors in $\text{conv}(W(t, t)\mu)$. Now, since $W$ is $G$-invariant, $t \cap W^0$ is $W(t, t)$-invariant, and we have to check $(\ast)$ only for $\mu$. Finally, since $\delta$ is $K_C \cap H_C$-spherical, we have $\mu|H = 0$ and $(\ast)$ reduces to the claimed condition. \hfill $\square$

**Remark.** For $g$ semisimple and $a \cap C = c_{\text{min}}$ it is known by [HÓ91] that all $(\epsilon_\delta, V_\delta)$ are admissible.

Let $\Xi(C) := \Gamma(-W)x_o \subset G_C/H_C$, where $x_o = eH_C$. As in [HÓ91, Lemma 1.3] we have $\Xi(C) \simeq G \times_H (-iC)$ and for $\Xi(C^0) := \Gamma(-W^0)x_o$ we have $\Xi(C)^0 = \Xi(C^0)$. Thus $\Xi(C^0)$ is an open complex submanifold of $G_C/H_C$. Notice that $\Gamma(-W) = \Gamma(W)^{-1}$. As $\Gamma(W)\Gamma(W^0) \subset \Gamma(W^0)$ it follows that $\Gamma(-W^0)\Xi(C) \subset \Xi(C^0)$. Thus $\Gamma(W)$ acts on functions defined on $\Xi(C)$ respectively $\Xi(C^0)$ by

$$\gamma \cdot f(x) = f(\gamma^{-1}x)$$  

and $(\gamma \cdot f)|_{G/H}$ is well-defined for $\gamma \in \Gamma(W^0)$). Define the Hardy norm on holomorphic functions on $\Xi(C^0)$ by

$$\|f\|_H := \sup_{\gamma \in \Gamma^0} \|\gamma \cdot f\|_{L^2(G/H)}.$$  

Definition. The Hardy space $H_2(C)$ on $G/H$ is
definition set of holomorphic functions $f$ in $\Xi(C^o) \to \mathbb{C}$ with $\|f\|_H < \infty$. 

Denote by $T$ the representation of $\Gamma(W)$ in $H_2(C)$ induced by the left-regular action and by $\lambda$ the left-regular representation of $G$ in $L^2(G/H)$. The proof of the following theorem then generalizes without any change (or simply using again that $G = ZG_a$).

**Theorem 3.5 ([HÖ91])**. (1) The Hardy space $H_2(C)$ is a Hilbert space with norm $\| \cdot \|_H$. 

(2) There is an isometry $\beta : H_2(C) \to L^2(G/H)$ given by 
\[
\beta(f) = \lim_{\Gamma(W)\ni \gamma \to 1} \gamma \cdot f,
\]
where the limit is in $L^2(G/H)$. 

(3) The boundary value map $\beta$ is an intertwining operator for the $G$-actions, i.e., $\beta T(g) = \lambda (g) \beta$ for $g \in G$. 

(4) The representation $T$ is a holomorphic representation of $\Gamma(W)$ in $H_2(C)$. 

(5) The image of $\beta$ in $L^2(G/H)$ is the direct sum of all the holomorphic discrete series that are $W$-admissible and each occurs with multiplicity one. 

(6) Assume that $w \in \Xi(C^o)$. Then the evaluation map $H_2(C) \ni f \mapsto f(w) \in \mathbb{C}$ is continuous. Let $K_w \in H_2(C)$ be such that $f(w) = (f \mid K_w)$, where $(\cdot \mid \cdot)$ denotes the scalar product on $H_2(C)$. Then the map $(z,w) \mapsto K(z,w) := K_w(z)$ is holomorphic in the first variable and antiholomorphic in the second variable. We have: 

a) $K(w,z) = K(z,w)$.

b) Let $\gamma \in \Gamma(-W)$, and let $z,w \in \Xi(C^o)$. Then 
\[
K(\gamma \cdot z, w) = K(z, \gamma^* \cdot w),
\]
with $(g \exp iX)^* := \exp (iX) g^{-1}$. 

c) There exists a holomorphic $H$-invariant function $\Theta_K : \Xi(C^o) \to \mathbb{C}$ such that 
\[
K(\gamma_1 \cdot x_o, \gamma_2 \cdot x_o) = \Theta_K (\gamma_2^* \gamma_1 \cdot x_o).
\]

(7) Suppose that $z = \gamma \cdot x_o \in \Xi(C^o)$. Then $w \mapsto K(z,w)$ extends to a smooth map on $\Xi(C)$, and the inverse of $\beta$ is given by 
\[
F(z) = \int_{G/H} f(x) K(z,x) \, dx = \int_{G/H} f(\hat{g}) \Theta_K (g^{-1} \gamma \cdot x_o) \, dg,
\]
where $\hat{g} = g \cdot x_o$, and $dg$ is the $G$-invariant measure on $G/H$ normalized by $dg = dh \, dg$.

We note for later use that the extension of $\Phi_{\hat{g}}$ from $G/H$ to $\Xi(C^o)$ is given by the defining formula, as $\Gamma(-W) \subset HCG C^* P^+$ ([HÖ91, Lemma 3.6]).
Example (continued). Since $g = \mathfrak{k} \cap q = a$ we have by Proposition 2.4
\[ C^o = \mathbb{R}^+ Z^0 = \mathbb{R}^+ \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = W^o \]
and further
\[ \Gamma(-W) = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \bigg| |a| \leq 1 \right\}. \]
The extension of $\Phi : G/H \to \hat{S}_1$ to a $G_{\mathbb{C}}$-invariant holomorphic map is
\[ \Phi : G_{\mathbb{C}}/H_{\mathbb{C}} \to \mathbb{C} \quad \left( \begin{array}{cc} a & 0 \\ 0 & a^{-1} \end{array} \right) H_{\mathbb{C}} \mapsto -a^2, \]
such that
\[ \Phi(\Xi(C^o)) = \Phi(\Gamma(-W^o)x_o) = \{ z \in \mathbb{C} \mid 0 < |z| < 1 \}. \]
We describe the Hardy space using this explicit realization of $\Xi(C^o)$. Since $G$ operates by rotations, the Hardy norm is given by
\[ ||f||_H = \sup_{0 < |z| < 1} \frac{1}{\pi} \int_0^\pi |f(e^{2i\varphi}z)|^2 d\varphi \]
and, using the Laurent series representation of $f$, we get
\[ H^2(C) = \left\{ f : \{ z \in \mathbb{C} \mid 0 < |z| < 1 \} \to \mathbb{C} \bigg| f(z) = \sum_{n=0}^\infty a_n z^n \right\} \]
with $\sum_{n=0}^\infty |a_n|^2 < \infty$.
We can describe this space also as sum of $W$-admissible representations. Identifying $G/H$ via $\Phi$ with $S^1$, the holomorphic discrete series representations are the characters
\[ \Phi_n : S^1 \to \mathbb{C}, \quad z \mapsto z^n, \quad n \in \mathbb{Z}. \]
All these functions have obviously a holomorphic continuation to $\Xi(C^o)$. Checking which of these representations are $W$-admissible, we get the condition $n \geq 0$, proving Theorem 3.5(5) for this case.

4. Characterization of $\Phi(G/H)$.

The Weyl group for the root system $\Delta(t_1, t_1)$ of compact roots acts by permutations on the positive noncompact roots $\Delta^+(p_1, t_1)$. Therefore
\[ \rho_{1,n} := \frac{1}{2} \sum_{\alpha \in \Delta^+(p_1, t_1)} \dim_{\mathbb{C}} g_{1,\alpha} \alpha \in i t_1^* \]
is zero on the orthogonal complement of $\mathfrak{z}(t_1)$ in $t_1$. We define $\Delta_1 := \Delta(\mathfrak{g}_{1,\mathbb{C}}, t_{1\mathbb{C}})$ and the positive roots $\Delta_1^+$ by restricting the roots $\Delta^+(\mathfrak{g}_{1,\mathbb{C}}, t_{1\mathbb{C}})$.
to $t_1^{-}$. Since $(g_1, t_1, \theta)$ is assumed to be Hermitian symmetric of tube type, and, by abuse of notation, identifying the strongly orthogonal roots $\gamma_i$ with elements of $(t_1^{-})^*$, by the Theorem of Moore ([H78, p. 528] or [H94, p. 460])

$$\Delta_1 = \left\{ \pm \frac{1}{2}(\gamma_i + \gamma_j) \mid 1 \leq i \leq j \leq r \right\} \cup \left\{ \pm \frac{1}{2}(\gamma_i - \gamma_j) \mid 1 \leq i < j \leq r \right\}. $$

Thereby the positive noncompact roots are

$$\Delta_1^{+} = \left\{ \frac{1}{2}(\gamma_i + \gamma_j) \mid 1 \leq i \leq j \leq r \right\},$$

the root spaces $g_{1C, \gamma_i}$ are one dimensional, and root spaces corresponding to roots $\frac{1}{2}(\gamma_i \pm \gamma_j)$, $i \neq j$, have all the same dimension $d$. As the restriction of a noncompact root to $t_1^{-}$ is always nonzero, we get immediately

$$\rho_{1,n} = \frac{1}{2} \left( 1 + \frac{d(r - 1)}{2} \right) (\gamma_1 + \cdots + \gamma_r).$$

For the moment we prefer to work with another Cartan algebra then $t_1$. Therefore, let $t_1^{+}$ be the orthogonal complement of $t_1^{-}$ in $t_1$, then the Cayley transform is the identity on $t_1^{+}$ and maps $t_1^{-}$ onto $a_{pC}$. For $\alpha \in \Delta(g_{1C}, t_{1C})$ we define the Cayley transformed root $\alpha^c$ by

$$\alpha^c := \alpha \circ \text{Ad}(c^{-1}) \in \Delta(g_{1C}, t_{1C}^{+} \oplus a_{pC}) =: \Delta_1^c,$$

and the positive system $\Delta_1^{c,+}$ by the transformed roots of $\Delta^+(g_{1C}, t_{1C})$.

We define

$$\rho_{+} := \text{Ad}(c^{-1})^* \rho_{1,n} = \frac{1}{2} \left( 1 + \frac{d(r - 1)}{2} \right) (\gamma_1^c + \cdots + \gamma_r^c) \in (t_{1C}^{+} + a_{pC})^*,$$

which is zero on the orthogonal complement of $CX^0$. We denote by $(\pi_m, V_m)$ the irreducible finite-dimensional representation of $G_{1C}$ with lowest weight $-m\rho_{+}$, if it exists.

We will need the Riemannian dual algebra $g_1^d$ associated to $(g_1, g, \sigma)$ for the proof of the next proposition

$$g_1^d := t_1 \cap g + i(p_1 \cap g) + p_1 \cap q_{1\sigma} + i(t_1 \cap q_{1\sigma}).$$

The involution $\sigma$ restricted to $g_1^d$ is a Cartan involution. Hence $t_1^d := t_1 \cap g + i(p_1 \cap g)$ is a maximal compactly embedded subalgebra of $g_1^d$ and the corresponding orthogonal complement is

$$p_1^d := p_1 \cap q_{1\sigma} + i(t_1 \cap q_{1\sigma}).$$

Let $\hat{a} \subset t_1 \cap q_{1\sigma} \cap g_1, (b)$ be maximal abelian, then $a^d := i\hat{a} + b$ is maximal abelian in $p_1^d$. For the root system $\Delta_1^d := \Delta(g_1^d, a^d)$ we choose the positive roots $\Delta_1^{d,+} := \Delta^+(g_1^d, a^d)$ compatible with $\Delta^+(g_1^d, b) = \Delta^+(g_1, b) = \left\{ \alpha|_b \mid \alpha \in \Delta_1^{c,+}, \alpha|_b \neq 0 \right\}$. 
Proposition 4.1. Let the notation be as above, then the following holds:

1. There exists an irreducible finite-dimensional representation \((\pi_2, V_2)\) of \(G_{1C}\) with lowest weight \(-2\rho_+\).
2. Assume that \(g_1 \neq sp(2n, \mathbb{R}), so(2, 2k + 1)\), \(n, k \geq 1\). Then there exists an irreducible finite-dimensional representation \((\pi_1, V_1)\) of \(G_{1C}\) with lowest weight \(-\rho_+\).
3. The weight space \(V_{m, -m\rho_+}\) is left pointwise fixed by the identity component of \(M'_{1C}\), where \(M'_{1C}\) is defined by the Langlands decomposition \(P' \simeq M' \times \exp(\mathbb{R}X^0) \times Q_{1n}'\). For \(m = 4\) it is fixed pointwise by \(M'_{1C}\) and, if \(g_1 \neq sp(2n, \mathbb{R}), so(2, 2k + 1)\), \(n, k \geq 1\), the same is true for \(m = 2\).
4. There is a \(m \in \{1, 2, 4, 8\}\) such that \(\pi_m\) is \(G_{1C}\)-spherical. In this case, let \((\cdot, \cdot)\) be a scalar product, invariant under the analytic subgroup to \(\mathfrak{k}_1 + i\mathfrak{p}_1\), then we can choose \(v_m \in V_{m, -m\rho_+}\) and \(\xi_m \in V_{G_{1C}}^C\) such that \((v_m|\xi_m) = 1\).

Proof. (1), (2) The first two assertions are [ÖØ99, Theorem 2.6].

(3) Let \(\bar{a}\) be the orthogonal complement of \(\mathbb{R}X^0\) in \(a_p\) and \(\Delta_0 := \{\alpha \in \Delta_1^c | \alpha(X^0) = 0\}\) then

\[
m_{1C} = t_{1C}^+ + \bar{a}_{1C} + \sum_{\alpha \in \Delta_0} g_{1C, \alpha}.
\]

Every lowest weight vector is invariant under \(t_{1C}^+ + \bar{a}\) and elements of root spaces \(g_{1C, \alpha}\) with \(-\alpha \in \Delta_1^{c, +}\). By the invariance of the weights under the Weyl group, for \(\alpha \in \Delta_1^{c, +}\) with \(\langle \alpha, \rho_+ \rangle = 0\) the weight space to \(-m\rho_+ + \alpha = s_\alpha(-m\rho_+ - \alpha)\) is trivial, hence \(m_{1C}\) operates trivial on \(V_{m, -m\rho_+}\). By the Theorem of Helgason \(\pi_m(M_{\min}), \text{with} M_{\min} := Z_{K_1}(a_p)\), leaves the weight space pointwise fixed if and only if \(\frac{\langle m\rho_+, \alpha \rangle}{\langle \alpha, \alpha \rangle} \in \mathbb{Z}^+\) for all \(\alpha \in \Delta_1^{c, +}\). For \(\alpha = \gamma_j^c\) this quotient is \(\frac{m}{2} \left(1 + \frac{d(r-1)}{2}\right)\), for \(\alpha = \frac{1}{2}(\gamma_i^c + \gamma_j^c)\) it is \(m \left(1 + \frac{d(r-1)}{2}\right)\), and for \(\alpha = \frac{1}{2}(\gamma_i^c - \gamma_j^c)\) it vanishes. This proofs the invariance under \(M_{\min}\), for \(m\) as in the assertion, by looking at the possible combinations for \(d\) and \(r\) ([H78, p. 530ff] or the table in [ÖØ99]). By \(M_{1C} = M_{\min}(\exp(m_{1C}))\) the result now follows.

(4) By Helgason’s Theorem \(\pi_m\) is \(G_{1C}\)-spherical if and only if

\[
\frac{\langle m\rho_+, \alpha \rangle}{\langle \alpha, \alpha \rangle} \in \mathbb{Z}^+, \quad \forall \alpha \in \Delta_1^{d, +}.
\]

By definition \(2\rho_{1,n}|b\) is a linear combination of elements of \(\Delta^+(p_{1C}, b)\) with positive integer coefficients. Therefore \(2\rho_+\) is a linear combination with positive integer coefficients of elements of \(\Delta^+(g_1, b)\). Now by [B88, p. 362]
the system $\Delta(g_1, b)$ is a root system in the axiomatic sense such that
\[
2 \frac{\langle 2\rho_+, \tilde{\alpha} \rangle}{\langle \tilde{\alpha}, \tilde{\alpha} \rangle} \in \mathbb{Z} \quad \text{for} \quad \tilde{\alpha} \in \Delta(g_1, b).
\]
For $\tilde{\alpha} := \alpha|_b$, with $\alpha \in \Delta^d_1$, and $\tilde{\alpha} \neq 0$ by [O87, Lemma 2.3] we have
\[
\langle \alpha, \alpha \rangle = n \langle \tilde{\alpha}, \tilde{\alpha} \rangle, \quad n \in \{1, 2, 4\},
\]
and $2\alpha \in \Delta^d_1$ if $n = 4$. Since a root $2\alpha$ gives a more severe restriction then $\alpha$ we can assume $n \in \{1, 2\}$, which implies the first assertion. The second assertion then is a consequence of the proof of Helgason’s Theorem [H84, p. 534ff]. \(\square\)

**Example.** We consider the causal compactification of $SO(2, n)/SO(1, n)$, as treated in [B97, Example 3.3]. Here $G_1 = SO(2, n + 1)$, i.e.,
\[
G_1 = \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in SL(n + 3, \mathbb{R}) \right| \begin{array}{l}
\transpose{t}AA - \transpose{t}CC = I, \\
\transpose{t}DD - \transpose{t}BB = I
\end{array} \right\},
\]
with Cartan involution $\theta(g) = \transpose{t}g^{-1}$, $X^0 = 2(E_{1,n+3} + E_{n+3,1})$, and
\[
\sigma = \text{Ad} \left( \begin{array}{cc}
\text{id}_{n+2} & 0 \\
0 & -1
\end{array} \right).
\]
Then $q_{1,\sigma}$ consists of matrices with nonzero entries only in the last row respectively column such that $b = \mathbb{R}X^0$ is maximal abelian in $p_1^d$, i.e., $a^d = b$. For the Riemannian dual algebra $g_1^d$ we get therefore, restricting the roots of $\Delta^+(g_1, a_p) = \{\gamma_1^c, \gamma_2^c, \frac{1}{2}(\gamma_1^c \pm \gamma_2^c)\}$ to $b$, the positive system $\Delta^+(g_1^d, b) = \Delta^+(g_1, b) = \{\frac{1}{2}(\gamma_1^c + \gamma_2^c)\}$. By Helgason’s Theorem, applied to $(g_1^d, b)$, the condition for $\pi_m$ to be $G$-spherical is now
\[
m \left( 1 + \frac{d(r - 1)}{2} \right) \in \mathbb{Z}^+.
\]
For our $g_1$ here $r = 2$ and $d = n - 1$ ([H78, p. 530ff] or the table in [O99]) such that $\pi_2$ is always $G$-spherical, whereas $\pi_1$ only for $n$ odd.

In a similar fashion one calculates the minimal $m$’s for the other causal compactifications as given in the following table. Here $p$ and $q$ are nonnegative integers with $p + q = n$. In the case $(\mathfrak{so}(2, q) \oplus \mathfrak{so}(p + 1), \mathfrak{so}(1, q) \oplus \mathfrak{o}(p), \mathfrak{so}(2, n + 1))$ we also assume that $p \geq 2$:
For $G_1$ we define $P_1^\pm$, $K_{1C}$, and the functions $p_1^\pm$ and $k_{1C}$ in the same manner as for $G$. Since the map $P_1^+ \times K_{1C} \times P_1^- \rightarrow G_{1C}$, $(p^+, k, p^-) \mapsto p^+ k p^-$, is a diffeomorphism onto a dense subset we can define

$$\zeta_1 : P_1^+ K_{1C} P_1^- \rightarrow p_1^+ \quad g \mapsto \exp^{-1} p_1^+ (g).$$

We will use the Harish-Chandra realization $D_1 := \zeta_1(GK_{1C}P_1^-)$ of $G_1/K_1$ in $p_1^+$ and $\zeta_1$ to state our results in a particularly nice form and note therefore:

**Proposition 4.2.** In the Harish-Chandra realization $\hat{S}_1 = \zeta_1(G_{1c})$ and the image $\Phi(G/H)$ of the causal compactification is given by $\zeta_1(Gc)$.

**Proof.** The first assertion is [KW65, Theorem 3.6], the second then follows from Theorem 2.3. \qed

To avoid a clumsy notation we will assume from now on, using the Proposition, that $\Phi(G/H) \subset p_1^+$. Define the canonical cocycle

$$J_1(g, z) = J_1^1(g, z) := k_{1C}(g \exp z)^{2\rho_{1,n}},$$
where \((\cdot)^{2p_{1,n}}\) is the central character of \(K_{1C}\) defined by \(2p_{1,n}\). More general, assume \(m \geq 1\) such that \((\pi_m, V_m)\) exists then

\[
J_{1}^{m/2}(g, z) := \kappa_{1C}(g \exp z)^{m p_{1,n}}
\]

is defined. If \((\pi_m, V_m)\) is \(G\)-spherical define

\[
\psi_m : p_1^+ \to \mathbb{C} \quad z \mapsto (\pi_m(\exp(z) c^{-1}) v_m | \xi_m).
\]

**Theorem 4.3.** Assume \(m \geq 1\) such that \((\pi_m, V_m)\) exists and is \(G\)-spherical. Then the function \(\psi_m(z)\) is holomorphic on \(p_1^+\) and has the following properties:

1. \(\Phi(G/H) = \{ z \in \mathcal{S}_1 \ | \ \psi_m(z) \neq 0 \}\).
2. For \(g \in G_{C}\), \(z \in p_1^+\), such that \(g \cdot z\) is defined, we have

\[
\psi_m(g \cdot z) = J_{1}^{m/2}(g, z) \psi_m(z).
\]

**Proof.** (1) For \(z = \zeta_1(gc) \in \mathcal{S}_1\), with \(g \in G_1\), we have

\[
\psi_m(z) = (\pi_m(p_1^+(gc)c^{-1}) v_m | \xi_m) = (\pi_m(gc p_1^+(gc)^{-1}) k_{1C}(gc)^{-1} c^{-1}) v_m | \xi_m).
\]

Since \(\text{Ad}(c)iZ^0 = \frac{1}{2}X^0\), the Cayley transformed group \(\text{Ad}(c)(K_{1C}P_{1}^-)\) is \(P_{C}'\), where we denote by \(\text{Ad}(c)\) the inner automorphism \(x \mapsto cxc^{-1}\) of \(G_{1C}\). In [B97] it is shown that \(GP'\) is the only open \((G, P')\)-double coset in \(G_{1}\). Then by [BD92, Lemme 4] we have \((\pi_m(gp)v_m | \xi_m) \neq 0\), for \(p \in P_{C}'\), if and only if \(g \in G\). (As \(M'/M_0'\) is finite, and using that every component contains an element of the form \(\exp i\lambda X^0\), the cited result is still valid with obvious modifications in its proof, even if \(v_m\) is not \(M_0'\)-fixed. Note also, that we use a lowest weight vector \(v_m\), since the nilradical \(q_{1q}\) of our parabolic algebra \(p'\) is the sum of root spaces for negative roots.)

(2) Since \(p_1^+\) is \(K_{1C}\)-invariant and mapped by \(\text{Ad}(c)\) onto \(q_{1qC}\), we have

\[
\psi_m(g \cdot z) = (\pi_m(g \exp z) p_1^-(g \exp z)^{-1} k_{1C}(g \exp z)^{-1} c^{-1}) v_m | \xi_m)
\]

\[
= (\pi_m(\exp z p_1^-(g \exp z)^{-1} k_{1C}(g \exp z)^{-1} c^{-1}) v_m | \pi_m(\theta(g)^{-1}) | \xi_m))
\]

\[
= (\pi_m(\exp z k_{1C}(g \exp z)^{-1} c^{-1} Ad(ck_{1C}(g \exp z))(p_1^-(g \exp z)^{-1}) v_m | \xi_m)
\]

\[
= (\pi_m(\exp z c^{-1} Ad(\theta(k_{1C}(g \exp z)^{-1})) v_m | \xi_m)
\]

\[
= J_{1}^{m/2}(g, z)(\pi_m(\exp z c^{-1}) v_m | \xi_m).
\]

\(\square\)

**Example.** We want to describe the causal compactification for \(G/H\) of Cayley type. Let \(G_1 = \tilde{G} \times \tilde{G}\), with \(\tilde{G} \subset \tilde{G}_{C}\) simply connected and \((\tilde{g}, \tilde{e}, \tilde{\theta})\) Hermitian symmetric of tube type. With \(\sigma(g, h) := (h, g)\) the fixpoint group \(G\) of this involution is isomorphic to \(\tilde{G}\). For \(\tilde{a}_p \subset \tilde{p}\) maximal abelian and
\[ \theta = \bar{\theta} \times \bar{\theta}, \] we have \( \mathfrak{a}_p = \bar{\mathfrak{a}}_p \times \bar{\mathfrak{a}}_p \) and \( b = \{(X, -X) \mid X \in \bar{\mathfrak{a}}_p\} \). With \( \bar{\Delta} := \Delta(\bar{\mathfrak{g}}, \bar{\mathfrak{a}}_p) \) we can define \( \bar{\rho}_+ \) and get then \( \rho_+ = (\bar{\rho}_+, -\bar{\rho}_+) \). Given \( (\bar{\pi}_m, \bar{V}_m) \), a representation of \( \bar{G}_C \) with lowest weight \(-m\bar{\rho}_+\), the representation \( \pi_m := \bar{\pi}_m \otimes \bar{\pi}_m^\ast \), where \( \bar{\pi}_m^\ast \) is the contragredient representation to \( \bar{\pi}_m \), has lowest weight \(-m\rho_+\). Moreover every such representation has the \( G \)-fixed vector \( \sum w_i \otimes w_i^\ast \), where \( \{w_i \mid i = 1, \ldots, \dim \bar{V}_m\} \) is a base of \( \bar{V}_m \) and \( \{w_i^\ast\} \) its dual. We have now \( c = (\bar{c}, \bar{c}^{-1}) \) and for \( (z, w) \in \mathfrak{p}_1^+ \)

\[
\psi_m((z, w)) = \left( \pi_m(\exp z \bar{c}^{-1}) \otimes \pi_m^\ast(\exp w \bar{c}) \cdot \bar{v}_m \otimes \bar{v}_m^\ast, \sum w_i \otimes w_i^\ast \right)
\]

\[
= \sum (\pi_m(\exp w \bar{c}^{-1}) \otimes 1 \cdot \bar{v}_m \otimes \bar{v}_m^\ast | w_i \otimes w_i^\ast) = (\pi_m(\bar{c}^{-1} \exp(z-w)\bar{c}^{-1}) v_m | v_m)
\]

Here we used \( \bar{\mathfrak{t}}^+ + \bar{\mathfrak{a}}_p \) as a Cartan algebra contrary to \( \bar{\mathfrak{t}}^+ + \bar{\mathfrak{a}} \) used in [\( \bar{\mathfrak{O}}\bar{\mathfrak{O}} 99 \)]. Therefore the lowest weight vectors are related by the Cayley transform and our \( \psi_m \) is the same function as those constructed in [\( \bar{\mathfrak{O}}\bar{\mathfrak{O}} 99 \)].

5. Some \( L^2 \)-isometries.

We know that \( \Phi(G/H) \subset \mathcal{S}_1 \) is open and dense. In this section we relate the \( L^2 \)-spaces with respect to invariant measures on these manifolds. The following results can be found in [\( \bar{\mathfrak{O}}\bar{\mathfrak{O}} 99 \):]

For \( \mathcal{S}_1 \) in the Harish-Chandra realization the quasi-invariant measure is given by

\[
\int_{\mathcal{S}_1} f(z) \, d\mu(z) = \int_{K_1/K_1 \cap Z_{G_1}(X^0)} f(k \cdot \bar{\zeta}_1(c)) \, dk.
\]

With \( P' = Z_{G_1}(X^0)Q_{1\eta}^- \) we can decompose every \( g \in G_1 \) in the form

\[
g = k_1(g)h(g)q^-(g),
\]

where \( k_1(g) \in K_1, h(g) \in Z_{G_1}(X^0), \) and \( q^-(g) \in Q_{1\eta}^- \). To \( z \in \mathcal{S}_1 \) choose \( k \in K_1 \) such that \( z = k \cdot \bar{\zeta}_1(c) \) and define

\[
J_R(g, z) := h(gk)^{2\rho_+} = \det \left( \text{Ad}(h(gk))_{q_{\bar{\eta}}} \right)^{-1},
\]

then

\[
\int_{\mathcal{S}_1} f(g \cdot z) J_R(g, z) \, d\mu(z) = \int_{\mathcal{S}_1} f(z) \, d\mu(z)
\]

and the relation

\[
k_{1c}(g \exp z)^{2\rho_{1,n}} = k_1(gk)^{2\rho_{1,n}}k^{-2\rho_{1,n}}h(gk)^{2\rho_+}
\]

holds. Note that the two first factors on the right hand side have modulus one by the compactness of \( K_1 \).
For \( k_{1C} \) we also have the cocycle property
\[
k_{1C}(gh \exp z) = k_{1C}(g \exp(h \cdot z)) k_{1C}(h \exp z),
\]
which gives
\[
J_1^{m/2}(gh, z) = J_1^{m/2}(g, h \cdot z) J_1^{m/2}(h, z)
\]
when \( J_1^{m/2} \) is defined. The following theorem now follows from the cocycle property for \( J_1(\cdot, \cdot) \).

**Theorem 5.1.** (1) The \( G \)-invariant measure on \( \Phi(G/H) \subset \tilde{S}_1 \) is up to normalization given by
\[
\int_{\tilde{S}_1} f(z)|\psi_m(z)|^{-2/m}d\mu(z).
\]
(2) A unitary representation of \( G_1 \) in \( L^2(\tilde{S}_1) \) is given by
\[
(\lambda_0(g)f)(z) := \sqrt{|J_1(g^{-1}, z)|} f(g^{-1} \cdot z).
\]
(3) Identify \( G/H \) with \( \zeta_1(\Phi(G/H)) \) and let \( \lambda \) denote the left regular representation on \( G/H \), then \( f \mapsto f|\psi_m|^{1/m} \) is a \( G \)-equivariant isometry of \( (L^2(\tilde{S}_1), \lambda_0|G) \) onto \( (L^2(G/H), \lambda) \).

If \( g_1 \neq \text{sp}(2n, \mathbb{R}), \text{so}(2,2k+1) \), the holomorphic square root \( \sqrt{J_1} = J_1^{1/2} \) is well-defined. We get then in the same way as the preceding theorem.

**Theorem 5.2.** (1) If \( g_1 \neq \text{sp}(2n, \mathbb{R}), \text{so}(2,2k+1) \), then
\[
(\lambda_1(g)f)(z) := \sqrt{J_1(g^{-1}, z)} f(g^{-1} \cdot z)
\]
is a unitary representation of \( G_1 \) in \( L^2(\tilde{S}_1) \).
(2) If further \( \pi_1 \) is \( G \)-spherical then \( f \mapsto f|\psi_1| \) is a \( G \)-equivariant isometry of \( (L^2(\tilde{S}_1), \lambda_1|G) \) onto \( (L^2(G/H), \lambda) \).

In the general case, to define the roots \( \sqrt{J_1(\cdot, \cdot)} \) and \( \sqrt[\cdot]{\psi_m} \), we have to replace \( G_1 \) and \( \Phi(G/H) \) by appropriate coverings. For
\[
\mathcal{D} := \{ z \in \tilde{D}_1 \mid \psi_m(z) \neq 0 \},
\]
which is independent of \( m \) by the last section, define a \( m \)-fold covering by
\[
\tilde{\mathcal{D}} := \{(z, x) \in \mathcal{D} \times \mathbb{C} \mid \psi_m(z) = x^m\}.
\]
We define then a holomorphic \( m \)-th root \( \sqrt[\cdot]{\psi_m} \) on \( \tilde{\mathcal{D}} \) by
\[
\sqrt[\cdot]{\psi_m} : \tilde{\mathcal{D}} \to \mathbb{C} \quad (z, x) \mapsto x.
\]
In the same manner we define a double covering \( \tilde{G}_1 \) of \( G_1 \) such that \( \sqrt{J_1(\cdot, \cdot)} \) is defined on \( \tilde{G}_1 \times \tilde{D}_1 \) and equal to 1 on \( |K_1, \tilde{K}_1| \times \{0\} \). More general, as the Ol’shanski˘ı semigroup \( \Gamma_1(W_k) \) is homeomorphic to \( G_1 \times iW_k \) and \( \Gamma_1(W_k) \subset \{ \gamma \in G_{1C} \mid \gamma^{-1}\tilde{D}_1 \subset \tilde{D}_1 \} \) (\[\text{HÔÔO91}\]), there is a double...
covering $\tilde{\Gamma}_1(W_k)$ with $\sqrt{J_1(\cdot, \cdot)}$ defined on $\tilde{\Gamma}_1(-W_k) \times \tilde{D}_1$ and holomorphic on $\tilde{\Gamma}_1(-W_k) \circ \times D_1$.

Let $\kappa : \tilde{\Gamma}_1(-W_k) \rightarrow \Gamma_1(-W_k)$ be the canonical projection, then the covering semigroup operates on $\tilde{D}_1$ by projecting first with $\kappa$. Moreover the cocycle relation

$$\sqrt{J_1(\tilde{\gamma} \tilde{\delta}, z)} = \sqrt{J_1(\tilde{\gamma}, \tilde{\delta} \cdot z)} \sqrt{J_1(\tilde{\delta}, z)},$$

with $\tilde{\gamma}, \tilde{\delta} \in \tilde{\Gamma}_1(-W_k)$ and $z \in D_1$ is still valid. Let $\tilde{\Gamma}(\pm W) := \kappa^{-1}(\Gamma(\pm W)) \subset \tilde{\Gamma}_1(\pm W_k)$ be the set lying above $\Gamma(\pm W)$ then by Theorem 4.3 we get the $\Gamma(-W)$- respectively $\tilde{\Gamma}(-W)$-invariance of $D \subset \tilde{D}_1$. We have an operation of $\tilde{\Gamma}(-W)$ respectively $\tilde{\Gamma}(-W)$ on $\tilde{D}$ by setting

$$\tilde{\gamma} \cdot (z, x) := (\tilde{\gamma} \cdot z, \sqrt{J_1(\tilde{\gamma}, z)} x).$$

Therefore we can define as usual the left action $\tilde{\lambda}$ by

$$\tilde{\lambda}(\tilde{\gamma}) f(z, x) := f(\tilde{\gamma}^{-1} \cdot (z, x))$$

for functions defined on $\tilde{D}$.

For $X \subset D$ let $\tilde{X} \subset \tilde{D}$ be the subset lying above $X$.

**Definition.** Let $X \subset D$ and $f : \tilde{X} \rightarrow \mathbb{C}$ be a function.

1. The function $f$ is odd if $f(z, \xi x) = \xi f(z, x)$ for all $m$-th roots of unity $\xi$ and $(z, x) \in \tilde{X}$.
2. The function $f$ is even if $f(z, \xi x) = f(z, x)$ for all $m$-th roots of unity $\xi$ and $(z, x) \in \tilde{X}$.

Every even function corresponds to a unique function on $X$ and, by abuse of notation, we will denote both by the the same letter. We note also that we have a $G$-invariant measure on $\tilde{G}/H$, defined as the pullback of the $G$-invariant measure on $G/H$ (and normalized by $m^{-1}$). Therefore the $L^2$-space of odd functions $L^2_{\text{odd}}(\tilde{G}/H)$ is a well-defined object.

With these definitions the proof of the last theorem can be generalized now and we get:

**Theorem 5.3.**

1. We have a unitary representation of $\tilde{G}_1$ in $L^2(\tilde{S}_1)$ given by

$$\tilde{\lambda}_1(\tilde{g}) f(z) := \sqrt{J_1(\tilde{g}^{-1}, z)} f(\tilde{g}^{-1} \cdot z).$$

2. If $\pi_m$ is $G$-spherical then $L^2(\tilde{S}_1, \tilde{\lambda}_1|_{\tilde{G}}) \rightarrow L^2(\tilde{G}/H, \tilde{\lambda}), f \mapsto f \overline{\psi}_m$, is a surjective $\tilde{G}$-equivariant isometry.
6. Characterization of $\Phi(\Xi(C^0))$.

Remember that $D_1 = \zeta_1(G_1 K_{1\mathbb{C}} P_{1\mathbb{C}})$ is the Hermitian symmetric space in the Harish-Chandra realization. Since $b \subset q_{1\sigma} \cap p_1$ is maximal abelian, we have $G_1 = G \exp(b) K_1$. This, together with Proposition 2.4(1), gives now

$$D_1 = G \cdot \left\{ \sum x_j E'_j \mid |x_j| < 1 \right\}$$

by a $\mathfrak{sl}(2)$-calculation, cf. [H78, p. 387].

**Proposition 6.1.** (1) $\Phi(\Xi(C^0)) = \{ z \in D_1 \mid \psi_m(z) \neq 0 \}$.

(2) There is a constant $c \neq 0$ such that

$$\psi_m \left( \sum x_j E'_j \right) = c \prod x_j^{mp_{1,n}(H'_j/2)} = c \prod x_j^{m \left( \frac{1 + d(r-1)}{2} \right) \sum \lambda_{j,j}}.$$  

**Proof.** First we note that $\Gamma(-W^0) Z_{G_{1\mathbb{C}}}(X^0) = G \exp(-i C^0) Z_{G_{1\mathbb{C}}}(X^0)$, which can be proven as in [HÓØ91, Prop. 1.4]. Using this, we get

$$\Phi(\Xi(C^0)) = \zeta_1(\Gamma(-W^0) \cdot c) = \zeta_1(G \exp(-i (a \cap C^0)) \cdot c),$$

since $H \subset Z_{G_{1\mathbb{C}}}(X^0) \subset \text{Stab}_{G_{1\mathbb{C}}}(c) = P_{1\mathbb{C}}$ and using Proposition 2.4(5). By a $\mathfrak{sl}(2)$-calculation

$$\exp(\lambda H_j) \cdot E_k = \begin{cases} E_k, & \text{if } j \neq k, \\ e^{2\lambda} E_k, & \text{if } j = k, \end{cases}$$

and, as $\zeta_1(c) = -\sum E_j = -\sum E'_j$, we finally get from Proposition 2.4(1) and (6) that

$$\Phi(\Xi(C^0)) = G \cdot \left\{ \sum x_j E'_j \mid 0 < |x_j| < 1 \right\} \subset D_1.$$

Conversely, to determine the set $\{ z \in D_1 \mid \psi_m(z) \neq 0 \}$ by Theorem 4.3 it is enough to consider $\{ \psi_m(\sum x_j E'_j) \mid |x_j| < 1 \}$. Using Proposition 2.4(1) we can calculate $\psi_m(\sum x_j E'_j)$ via $\mathfrak{sl}(2)$-reductions. Thus we assume for the moment $G_{1\mathbb{C}} = SL(2, \mathbb{C})$,

$$E'_j = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \ E'_{-j} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \ H'_j = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

for $j'$ fixed. Then $G_{1\mathbb{C}} \exp(\mathfrak{c} X'_j) Q_{1\mathbb{C}} \subset G_{1\mathbb{C}}$ is open and dense, defining a $r_j \in \mathbb{C}$ by $\exp(x_j E'_j) c^{-1} \in G_{1\mathbb{C}} \exp(r_j X'_j) Q_{1\mathbb{C}}$ for almost all $x_j$. To determine $r_j$, choose $(X|Y) := \text{tr}(XY^*)$ as an invariant scalar product on $\mathfrak{g}_{1\mathbb{C}}$ and $v \in q_{1\mathbb{C}}^+$ nonzero. Let $g \in G_{1\mathbb{C}}$ and $q \in Q_{1\mathbb{C}}^-$ be chosen such that $\exp(x_j E'_j) c^{-1} = g \exp(r_j X'_j) q$. Then we have

$$(\exp(x_j E'_j) c^{-1})^{-1} \sigma(\exp(x_j E'_j) c^{-1}) = q^{-1} \exp(-2x_j X'_j) \sigma(q)$$
and consequently, using \( \sigma(Q_{\eta C}) = \theta(Q_{1\eta C}) = Q_{1\eta C}^\tau \), we get

\[
(\text{Ad}((\exp(x_jE_j')c^{-1})^{-1}\sigma(\exp(x_jE_j')c^{-1}))v|v) = e^{-4r_j'(v|v)}.
\]

Inserting the appropriate matrices it follows easily \( r_j' = -\frac{1}{2} \log 2x_j' \) and further

\[
\psi_m\left(\sum x_j' E_j'\right) = \left(\prod_{j'} e^{-\kappa\rho_j'(X_j')}\right)(v_m|\xi_m) = c\prod_{j'} x_j' \frac{(1 + d(\kappa - 1))}{2} \sum_{j'} \lambda_{j'.j}.
\]

**Remark.** Using \( C = \text{Ad}(H)(a \cap C) \) together with \( a \cap C = -\sum \mathbb{R}_0^+ iH_j' \), the arguments of the proof show that \( \Xi(C) \subset D \).

### 7. Some Hardy space isometries.

The classical Hardy space for \( G \) respectively the covering group \( \tilde{G} \) in case \( g_1 = sp(2n, \mathbb{R}) \), \( so(2,2k+1) \), is defined as the Hilbert space

\[
H_{cl} := \left\{ f : D_1 \to \mathbb{C} \mid f \text{ is holomorphic and} \sup_{0<r<1} \int_{K_1} |f(rk \cdot c)|^2 \, dk < \infty \right\}
\]

with \( G_1 \)-action

\[
(\lambda(g)f)(z) := \sqrt{J_1(g^{-1},z)} f(g^{-1} \cdot z)
\]

respectively \( \tilde{G}_1 \)-action

\[
(\tilde{\lambda}(g)f)(z) := \sqrt{J_1(g^{-1},z)} f(g^{-1} \cdot z)
\]

([FK94, XIII.3]). By Theorem 5.2(1) respectively Theorem 5.3(1), the operation of taking boundary values is a \( G_1 \)- respectively \( \tilde{G}_1 \)-equivariant isometry into \( L^2(\tilde{S}_1) \).

**Theorem 7.1.** We identify \( \Xi(C^\circ) \) with \( \Phi(\Xi(C^\circ)) \). If \( g_1 \neq sp(2n, \mathbb{R}) \), \( so(2,2k+1) \) and \( \pi_1 \) is \( G \)-spherical then \( I:H_{cl} \to H_2(C) f \mapsto f\psi_1|_{\Xi(C^\circ)} \) is a \( G \)-equivariant isometry.

**Proof.** In \( H_2(C) \) the action is given by left translation, i.e., for \( f \in H_{cl} \) we get

\[
(\lambda(g)I(f))(z) = f(g^{-1} \cdot z) \psi_1(g^{-1} \cdot z)
\]

the second equality by Theorem 4.3. The map is also an isometry by Theorem 5.2(2), as taking boundary values is an isometry for the Hardy spaces. \( \square \)
For the general case we define first \( \tilde{H}_2(C) \) as covering space of \( H_2(C) \) with a \( \tilde{\Gamma}(W) \)-action by

\[
\tilde{H}_2(C) := \{ f : \tilde{\Xi}(C^o) \to \mathbb{C} \mid f \text{ is holomorphic and } \|f\|_{\tilde{H}} < \infty \},
\]

where \( \|f\|_{\tilde{H}} := \sup_{\gamma \in \tilde{\Gamma}^o} \|\tilde{\gamma} \cdot f\|_{L^2(\tilde{G}/\tilde{H})} \). For this space the analogous results as stated in Section 3 hold. One proves now in the same manner as the theorem above:

**Theorem 7.2.** If \( \pi_m \) is \( G \)-spherical then

\[
\tilde{\Gamma} : H_{cl} \to \tilde{H}_{2,odd}(C) \ f \mapsto f \sqrt[2]{\psi_m}|_{\tilde{\Xi}(C^o)}
\]

is a \( \tilde{G} \)-equivariant isometry.

Let us look closer at the different cases \( m = 1, 2, 4 \).

**m=1:** In this case both \( \psi_1 \) and \( \sqrt{\tilde{\Gamma}_1} \) exist, so all coverings split in a trivial way. In particular \( L^2_{odd}(G/H) \simeq L^2_{even}(G/H) \) and similarly also \( \tilde{H}_{2,odd}(C) \simeq \tilde{H}_{2,even}(C) \simeq H_2(C) \). By the table before Proposition 4.2 the following \((g, h, g_1)\)-triples correspond to this case:

| \((s(u(p, q) \oplus u(p, q)), su(p, q), su(n, n))\) | \(n \) even | \(n = 3 \pmod{4}\) |
| \((u(p, q), so(p, q), sp(n, R))\) | \(n = 3 \pmod{4}\) |
| \((so(2, n - 1) \oplus so(2), so(1, n - 1), so(2, n + 1))\) | \(n = 3 \pmod{4}\) |
| \((so^*(2n), so(n, C), su(n, n))\) | \(n = 3 \pmod{4}\) |
| \((sp(2n, R), sp(n, C), su(2n, 2n))\) | \(n = 3 \pmod{4}\) |
| \((so(2, 2k), so(1, 2k) \oplus o(p), so(2, n + 1))\) | \(g \neq so(2, 2k + 1), \ sp(2n, R)\) |
| \((g, h, g \times g)\) for \((g, h)\) of Cayley type | \(n \) odd |

**m=2:** Let us first assume \( g_1 \neq sp(2n, R) \), \( so(2, 2k + 1) \) such that \( \sqrt{\tilde{J}_1} \) exists as a holomorphic function on \( \Gamma_1(-W_k) \times D_1 \). Those are the \((g, h, g_1)\)-triples:

| \((s(u(p, q) \oplus u(p, q)), su(p, q), su(n, n))\) | \(n \) odd | \(n = 1 \pmod{4}\) |
| \((u(2p, 2q), sp(p, q), so^*(4n))\) | \(n = 1 \pmod{4}\) |
| \((u(p, q), so(p, q), sp(n, R))\) | \(n = 1 \pmod{4}\) |
| \((so(2, n - 1) \oplus so(2), so(1, n - 1), so(2, n + 1))\) | \(n = 1 \pmod{4}\) |
| \((\epsilon_{6(-14)} \oplus so(2), f_{4(-20)}, \epsilon_{7(-25)})\) | \(n = 1 \pmod{4}\) |
| \((so^*(2n), so^*(2n), so^*(2n), so^*(4n))\) | \(n = 1 \pmod{4}\) |
| \((so(2, 2n), so(n, C), su(n, n))\) | \(n = 1 \pmod{4}\) |
| \((so(2, q) \oplus so(p + 1), so(1, q) \oplus o(p), so(2, n + 1))\) | \(n = 1 \pmod{4}\) |

Let \( E := \sum E_j = -\zeta_1(c) \). By the equation

\[
\psi_2(g : z) = J_1(g, z) \psi_2(z)
\]
and the fact that $\Gamma(-W)$ acts transitively on $\Xi(C)$ we can define a global square root of $\psi_2$ on $\Xi(C)$ by

$$\sqrt{\psi_2(z)} := \sqrt{J_1(\gamma^{-1}, -E)},$$

with $\gamma$ chosen such that $z = \gamma^{-1} \cdot (-E)$. Clearly this square root is holomorphic on $\Xi(C^\alpha)$.

**Lemma 7.3.** For

$$\Xi(C)_+ := \left\{ \left( \gamma \cdot (-E), \sqrt{J_1(\gamma, -E)} \right) \mid \gamma \in \Gamma(-W) \right\}$$

and

$$\Xi(C^\alpha)_- := \left\{ \left( \gamma \cdot (-E), -\sqrt{J_1(\gamma, -E)} \right) \mid \gamma \in \Gamma(-W) \right\}$$

we have:

1. $\Xi(C)_+ = \Gamma(-W) \cdot (-E, 1)$ and $\Xi(C)_- = \Gamma(-W) \cdot (-E, -1)$.
2. $\Xi(C) = \Xi(C)_+ \cup \Xi(C)_-$ and the canonical projection $\kappa_2 : \Xi(C)_+ \rightarrow \Xi(C)$ is a diffeomorphism, biholomorphic on $\Xi(C^\alpha)_+$.

**Proof.** It is clear that $\Xi(C)_+, \Xi(C)_- \subset \Xi(C)$ and that $\Xi(C)_+ \cap \Xi(C)_- = \emptyset$.

Let $\xi = (z, x) \in \Xi(C)$ and $\gamma \in \Gamma(-W)$ such that $z = \gamma \cdot (-E)$ then

$$x^2 = \psi_2(z) = \psi_2(\gamma \cdot (-E)) = J_1(\gamma, -E) \psi_2(-E) = J_1(\gamma, -E).$$

Thus $\xi = (z, \sqrt{J_1(\gamma, -E)}) \in \Xi(C)_+$ or $\xi = (z, -\sqrt{J_1(\gamma, -E)}) \in \Xi(C)_-$ and in particular $\Xi(C) = \Xi(C)_+ \cup \Xi(C)_-$.

To define the inverse map to $\kappa_2$, choose for $z \in \Xi(C)$ a $\gamma \in \Gamma(-W)$ such that $z = \gamma \cdot (-E)$. Then $\kappa_2^{-1}$ is given by

$$z \mapsto (z, \sqrt{J_1(\gamma, -E)}),$$

which is a holomorphic map on $\Xi(C^\alpha) = \Gamma(-W^\alpha) \cdot (-E)$, since locally we may choose $\gamma$ using a holomorphic section. \hfill \Box

The following is now clear:

**Proposition 7.4.** (1) The restriction from $\Xi(C^\alpha)$ to $\Xi(C^\alpha)_+$ together with the canonical projection $\kappa_2 : \Xi(C^\alpha)_+ \rightarrow \Xi(C^\alpha)$ induces an isomorphism of $\Gamma(W)$-modules

$$\widetilde{H}_{2,\text{odd}}(C) \simeq \widetilde{H}_{2,\text{even}}(C) \simeq H_2(C).$$

(2) The restriction from $\widehat{G/H}$ to $G/H_+ := \widehat{G/H} \cap \Xi(C)_+$ together with the canonical projection $\kappa_2[G/H_+] : G/H_+ \rightarrow G/H$ induces an isomorphism of $G$-modules

$$L^2_{\text{odd}}(\widehat{G/H}) \simeq L^2_{\text{even}}(\widehat{G/H}) \simeq L^2(G/H).$$

(3) Taking boundary values in the $L^2$-norm intertwines the $G$-actions on the respective modules.
If \((g, h, g_1) = (so(2, 2k), so(1, 2k), so(2, 2k + 1))\) or \((g, h, g \times g)\) is one of the remaining Cayley type cases the preceding construction does not work. However, using the fact that \(g_1 = g \times g\), for the Cayley type cases it is possible to define a global square root \(\sqrt{J_1(\cdot, \cdot)}\) on \(\Gamma(-W)\). Therefore analogous results to Lemma 7.3 and Proposition 7.4 hold, cf. [ÖO99, Lemma 5.5, Lemma 6.5].

\(m=4\): For these cases \(\sqrt{J_1(\gamma, z)}\) does not exist as a globally defined map in the variable \(\gamma \in \Gamma_1(-W)\), as can be seen from the occurring \(g_1\) for the corresponding \((g, h, g_1)\)-triples:

\[
\begin{array}{|c|c|}
\hline
(g(p, q), so(p, q), sp(n, R)) & n \text{ even} \\
(so(2, n-1) \oplus so(2), so(1, n-1), so(2, n+1)) & n \text{ even} \\
(sp(n, R) \oplus sp(n, R), sp(n, R), sp(2n, R)) & n \text{ even} \\
(so(2, q) \oplus so(p+1), so(1, q) \oplus o(p), so(2, n+1)) & \\
\hline
\end{array}
\]

From the equation
\[
\psi_4(\gamma \cdot (-E)) = J_1^2(\gamma, -E) \psi_4(-E) = (J_1(\gamma, -E))^2,
\]
valid for \(\gamma \in \Gamma(-W)\), it follows that we can define a holomorphic square root of \(\psi_4\) on \(\Xi(C)\). But to define the needed function \(\sqrt{\psi_4}\) we have to go to a double covering. By the above, for \((\tilde{\gamma} \cdot (-E), x) \in \Xi(C)\), i.e., \(x^4 = \psi_4(\tilde{\gamma} \cdot (-E))\), we have \(x^2 = \pm J_1(\tilde{\gamma}, -E)\). Therefore we define in this case
\[
\Xi(C)_+ := \{(\tilde{\gamma} \cdot (-E), x) \mid \tilde{\gamma} \in \Gamma(-W), x^2 = J_1(\tilde{\gamma}, -E)\} \subset \Xi(C)
\]
and
\[
\Xi(C)_- := \{(\tilde{\gamma} \cdot (-E), x) \mid \tilde{\gamma} \in \Gamma(-W), x^2 = -J_1(\tilde{\gamma}, -E)\} \subset \Xi(C).
\]
Then \(\Gamma(-W)\) acts on \(\Xi(C)_\pm\) as subsets of \(\Xi(C) \subset D\) by
\[
\tilde{\gamma} \cdot (z, x) = (\tilde{\gamma} \cdot z, \sqrt{J_1(\tilde{\gamma}, z)} x).
\]
Note that for \(z \in \Xi(C)\) there are two elements \(\tilde{\gamma}, \tilde{\delta} \in \Gamma(-W)\) with \(z = \tilde{\gamma} \cdot (-E) = \tilde{\delta} \cdot (-E)\) and \(\sqrt{J_1(\tilde{\gamma}, -E)} = -\sqrt{J_1(\tilde{\delta}, -E)}\). Using this, the proof of Lemma 7.3 now generalizes word-by-word to give:

**Lemma 7.5.**

1. \(\Xi(C)_+ = \Gamma(-W) \cdot (-E, 1)\) and \(\Xi(C)_- = \Gamma(-W) \cdot (-E, -1)\).
2. We have
\[
\Xi(C) = \Xi(C)_+ \cup \Xi(C)_-
\]
and the canonical projection \(\kappa_4 : \Xi(C)_+ \rightarrow \Xi(C)\) is a double covering, holomorphic on \(\Xi(C)_+\).
We have on \( \widehat{G/H}_+ := \kappa^{-1}_+ (G/H) \) the space of \( L^2 \)-functions \( L^2(\widehat{G/H}_+) \) and on \( \widetilde{\Xi}(C^o)_+ = \kappa^{-1}_+ (\Xi(C^o)) \) the Hardy space \( \widetilde{H}_2(C)_+ \), as a space of holomorphic functions defined in the same way as \( \widetilde{H}_2(C) \). Using roots of unity we can define the subspaces \( L^2_{\text{odd}}(\widehat{G/H}_+) \) respectively \( \widetilde{H}^2_{\text{odd}}(C)_+ \) of odd functions.

**Proposition 7.6.** (1) The restriction from \( \widetilde{\Xi}(C^o) \) to \( \widetilde{\Xi}(C^o)_+ \) induces an isomorphism of \( \widetilde{\Gamma}(W) \)-modules
\[
\widetilde{H}^2_{\text{odd}}(C) \simeq \widetilde{H}^2_{\text{odd}}(C)_+.
\]

(2) The restriction from \( \widehat{G/H} \) to \( \widehat{G/H}_+ \) induces an isomorphism of \( G \)-modules
\[
L^2_{\text{odd}}(\widehat{G/H}) \simeq L^2_{\text{odd}}(\widehat{G/H}_+).
\]

The proof is obvious. We can in particular draw the conclusion from those considerations that we will need maximally a double covering of \( \Xi(C) \) for identifying the Hardy space.

For the characterization of the image we give first the extension of \( \Phi_\delta \) to a function on \( \Phi(\Xi(C^o)) \subset D_1 = G \cdot \{ \sum x_j E_j \} \).

**Lemma 7.7.** For \( g \in G \), define \( g = k(g) \exp(p(g)) \) by the Cartan decomposition. Let \( \{ \nu_{\mu', r} \} \) be a base of weight vectors of \( V_\delta \), with \( \nu_{\mu', r} \) of weight \( \mu' \), and define \( a_{\mu', r} \) by \( \delta(k) \nu = \sum a_{\mu', r} (k^{-1}) \nu_{\mu', r} \). Identifying \( \Xi(C^o) \) with \( \Phi(\Xi(C^o)) \), there is a constant \( c \neq 0 \) such that
\[
\Phi_\delta \left( g \cdot \sum x_j E_j \right) = c \sum a_{\mu', r}(k(g)) \prod_{j'} x_{j'}^{-\mu'(H_j'/2)} \langle \delta \left( f \left( g, \sum x_j E_j \right) \nu_{\mu', r}, \nu_0 \right) \rangle,
\]
with
\[
f \left( g, \sum x_j E_j \right) = k_H \left( \text{Ad} \left( \exp \left( -\frac{1}{2} \sum (\log x_j + \pi i) H_j' \right) p(g)^{-1} \right) \right)
\]
and \( \log \) the principal branch continued continuously to \(-\mathbb{R}^+\) from the upper half plane.

**Proof.** By Equation \( (**) \) and Proposition 2.4(1) we have
\[
g \cdot \sum x_j E_j = \zeta_1 \left( g \exp \left( \frac{1}{2} \sum (\log x_j + \pi i) H_j' \right) \right)c.
\]
Therefore
\[ \Phi_{\delta} \left( g \cdot \sum x_{j'} E_{j'}^\prime \right) = \left\langle \nu, \right. \delta \left( k_H \left( \exp \left( -\frac{1}{2} \sum (\log x_{j'} + \pi i) H_{j'} \right) g^{-1} \right) \nu^0 \right) \right\rangle \]
\[ = \delta(k(g)^{-1}) \left\langle \nu, \right. \delta \left( \exp \left( \frac{1}{2} \sum (\log x_{j'} + \pi i) H_{j'} \right) f \left( g, \sum x_{j'} E_{j'}^\prime \right)^{-1} \nu^0 \right) \right\rangle \]
\[ = c \sum_{\mu', r} a_{\mu', r} (k(g)) \prod_{j'} x_{j'}^{-\mu'(H_{j'}/2)} \left\langle \delta \left( f \left( g, \sum x_{j'} E_{j'}^\prime \right) \right) \nu_{\mu', r}, \nu^0 \right\rangle. \]

\[ \square \]

For \( f \in H_2(C) \), the inverse map to \( I \) is given by extending the holomorphic function \( f/\psi_1 \) to \( D_1 \). This extension is possible if and only if \( f/\psi_1 \) is locally bounded near \( D_1 \setminus \Xi(C^\circ) \). By Theorem 3.5(5) it is enough to check this for the generating functions \( f = \Phi_{\delta} \). Now by Theorem 4.3, Proposition 6.1, and the last lemma
\[ \left( \Phi_{\delta}/\Psi_1 \right) \left( g \cdot \sum x_{j'} E_{j'}^\prime \right) = \sum_{\mu'} b_{\mu'} \left( g, \sum x_{j'} E_{j'}^\prime \right) \prod_{j'} x_{j'}^{-(\mu' + \rho_{1,n})(H_{j'}/2)}, \]
where \( b_{\mu'} \) is locally bounded on \( G \times D_1 \).

**Theorem 7.8.** \( H_\delta \subset I(H_d) \iff (\mu + \rho_{1,n})(H_{j'}^\prime) \leq 0 \) for all \( j' \).

**Proof.** Equating \( \Phi_{\delta}(\sum x_{j'} E_{j'}^\prime) \) by Lemma 7.7, we get the necessary condition \( (\mu' + \rho_{1,n})(H_{j'}^\prime) \leq 0 \) for all \( \mu' \) with \( a_{\mu', r}(1) \neq 0 \) and \( \langle \nu_{\mu', r}, \nu^0 \rangle \neq 0 \). By [H84, p. 535ff] for the highest weight \( \mu \) of \( \delta \) both \( \nu_{\mu,1} \neq 0 \) and \( \langle \nu_{\mu,1}, \nu^0 \rangle \neq 0 \), showing that the given condition is necessary.

For sufficiency, let \( \mu' \) be a weight of \( (\delta, V_\delta) \) then \( \mu' \in \text{conv}(W(\mathfrak{t}, \mathfrak{t}) \cdot \mu) \subset \mathfrak{t}^* \).

Applying [B88, Lemma 1.1] to the Lie algebra \( \mathfrak{t}_C \) with involution \( -\sigma \), we can assume \( W(\mathfrak{t}, \mathfrak{t}) \subset W(\mathfrak{t}_1, \mathfrak{t}_1) \), where \( W(\mathfrak{t}_1, \mathfrak{t}_1) \) is the Weyl group of the root system \( \Delta(\mathfrak{t}_1, \mathfrak{t}_1) \). As we already have seen \( \rho_{1,n} \in (\mathfrak{g}(\mathfrak{t}_1)_C)^* \subset (\mathfrak{t}_1)_C^* \)

such that now \( W(\mathfrak{t}, \mathfrak{t}) \rho_{1,n} = \rho_{1,n} \) and further
\[ (w \cdot \mu + \rho_{1,n})(H_{i'}^\prime) = w \cdot (\mu + \rho_{1,n})(H_{i'}^\prime) = (\mu + \rho_{1,n})(w^{-1}H_{i'}^\prime). \]

We note that \( i(\mathfrak{a} \cap C^\circ) = \sum \mathbb{R}^+ H_{i'}^\prime \subset i\mathfrak{a} \) is the Weyl chamber for the roots \( \Delta_{1,n}^\circ \mathfrak{a} \). Indeed, since \( \gamma_i(H_{i'}) = 2\delta_{i,j} \), we deduce from the Remark following Proposition 2.4 that the co-roots for \( \Delta_{1,n}^\circ |_a \) are the vectors \( \{ \frac{1}{2}(H_{i'}^\prime + H_{j'}^\prime) \mid 1 \leq i' \leq j' \leq r(\mathfrak{t}) \} \). This means that \( i(\mathfrak{a} \cap C^\circ) \) is the set of all linear combinations of co-roots with positive coefficients, which — together with \( \langle \alpha, \beta \rangle \geq 0 \) for all \( \alpha, \beta \in \Delta_{1,n}^\circ |_a \) — implies the assertion.
Now, as $W(t_1, t_1)$ permutes the noncompact roots, we get
\[ \left( W(t, t) (\Delta^+_{t, n} | a) \right) | a \subset \Delta^+_{t, n} | a. \]
Expressing this with co-roots we get
\[ w^{-1} H_{ij} \in \sum \mathbb{R}_0^+ H_{ij} ', \]
such that by our equation above the given condition is sufficient for the boundedness of $\Phi / \Psi_1$. \Box

Of course, the same condition determines the image of $\overline{I}$ in $\overline{H}_{2, \text{odd}}(C)$. This theorem allows us now to determine the image of the classical Hardy space in $H_2(C)$ respectively $\overline{H}_{2, \text{odd}}(C)$ as described by the root inequalities of Theorem 3.3 and Proposition 3.4.

**Example.** We consider $(g, h, g_1) = (\mathfrak{sp}(n, \mathbb{R}) \oplus \mathfrak{sp}(n, \mathbb{R}), \mathfrak{sp}(n, \mathbb{R}), \mathfrak{sp}(2n, \mathbb{R}))$ as treated in [B97, Section 7.4]. Here
\[ g_1 = \left\{ \begin{array}{c} A \ B \\ \overline{A} \\ \overline{B} \end{array} \right| A, B \in M(2n \times 2n, \mathbb{C}), -\overline{A} = A, \overline{B} = B \right\} \]
and
\[ g = \left\{ \begin{array}{c} A_1 \\ B_1 \\ 0 \\ 0 \\ A_4 \\ B_4 \\ \overline{A}_1 \\ \overline{B}_1 \\ 0 \\ 0 \end{array} \right| A_i, B_i \in M(n \times n, \mathbb{C}) \right\}. \]
Further, $t_{iC}$ is spanned by
\[ H_j = \frac{1}{2} \left( \begin{array}{cccc} E_{jj} & E_{jj} & 0 & 0 \\ E_{jj} & E_{jj} & 0 & 0 \\ 0 & 0 & -E_{jj} & -E_{jj} \\ 0 & 0 & -E_{jj} & -E_{jj} \end{array} \right), \]
\[ H_{j+n} = \frac{1}{2} \left( \begin{array}{cccc} E_{jj} & -E_{jj} & 0 & 0 \\ -E_{jj} & E_{jj} & 0 & 0 \\ 0 & 0 & E_{jj} & E_{jj} \\ 0 & 0 & E_{jj} & E_{jj} \end{array} \right), \]
for $j = 1, \ldots, n$, whereas a base for $a_C$ is given by the vectors $H_j' = H_j + H_{j+n}$. Since $g_1$ is a normal Lie algebra the multiplicity of all roots in $\Delta_1$ is one. Defining $\gamma_j' := \frac{1}{2} (\gamma_j + \gamma_{j+n})$ we get by restricting the roots of $\Delta^+_1$ to $a_C$
\[ \Delta^+_1 = \left\{ \frac{1}{2} (\gamma_i' + \gamma_j') | 1 \leq i \leq j \leq n \right\} \cup \left\{ \frac{1}{2} (\gamma_i' - \gamma_j') | 1 \leq i < j \leq n \right\}, \]
and all these roots have multiplicity 2. With $\Delta^+_{a} = \left\{ \frac{1}{2} (\gamma_i' + \gamma_j') | 1 \leq i \leq j \leq n \right\}$ by Proposition 2.4(6) and the definition of $c_{\min}$ we get $a \cap C = c_{\min}$. Therefore, by the Remark following Proposition 3.4 the Hardy space $\overline{H}_2(C)$ comprises the whole holomorphic discrete series. To describe these in more
To apply Theorem 3.3(1) we calculate

\[ m = \sum m_j \gamma_j', \quad m_1 \geq m_2 \geq \ldots \geq m_n, \quad m_j \in \mathbb{Z} + \frac{1}{2}. \]

To apply Theorem 3.3(1) we calculate

\[ \rho = \frac{1}{2} \sum_{\alpha \in \Delta^+} \dim g_{\mathbb{C}, \alpha} \alpha = \sum_{1 \leq i \leq j \leq n} \frac{1}{2} (\gamma_i' + \gamma_j') + \sum_{1 \leq i < j \leq n} \frac{1}{2} (\gamma_i' - \gamma_j') \]

\[ = (n + 1) \sum_{j=1}^{n} \gamma_j' - \sum_{j=1}^{n} j \gamma_j'. \]

Inserting into \( (\mu + \rho, \alpha) < 0, \) for \( \alpha \in \Delta^+_n, \) we derive the condition

\[ H_3 \subset \bar{H}_{2, \text{odd}}(C) \iff -n \geq m_1 \geq m_2 \geq \ldots \geq m_n. \]

We show next that \( \bar{I} \) is actually surjective. In fact, with

\[ \rho_{1,n} = \frac{1}{2} \left( \frac{2n + 1}{2} \right) (\gamma_1 + \cdots + \gamma_{2n}) \]

the condition of Theorem 7.8 reduces to

\[ (\mu, \gamma_j') \leq -\frac{2n + 1}{4} \quad \text{or} \quad m_j \leq -\frac{2n + 1}{2} \]

such that, \( m_j + \frac{1}{2} \) being an integer, the embedding \( \bar{I} \) is onto. Another discussion of this example can be found in [K097].

Similar to the given example this “best possible” case, with \( H_2(C) \) respectively \( \bar{H}_{2, \text{odd}}(C) \) being the direct sum of all holomorphic discrete series representations and \( I \) respectively \( \bar{I} \) onto, occurs for the \((g, h, g_1)\)-triples

\( \text{(sp}(n, \mathbb{R}) \oplus \text{sp}(n, \mathbb{R}), \text{sp}(n, \mathbb{R}), \text{sp}(2n, \mathbb{R})) \)
\( \text{(sp}(2n, \mathbb{R}), \text{sp}(n, \mathbb{C}), \text{su}(2n, 2n)) \)
\( (g, h, g \times g) \) for \((g, h)\) of Cayley type
\( (\text{so}(2, n), \text{so}(1, n), \text{so}(2, n + 1)). \)

Note that these are exactly the cases with \( a \neq t_1 \).

**Example.** We discuss next the compactifications described by [B97, Cor. 5.3]. These are the \((g, h, g_1)\)-triples

\( (\text{su}(p, q) \oplus \text{u}(p, q)), \text{su}(p, q), \text{su}(n, n)) \), \( (\text{u}(2p, 2q), \text{sp}(p, q), \text{so}^*(4n)) \), \( (\text{u}(p, q), \text{so}(p, q), \text{sp}(n, \mathbb{R})), \) \( (\text{so}(2, n - 1) \oplus \text{so}(2), \text{so}(1, n - 1), \text{so}(2, n + 1)) \), \( (\text{e}_6(-14) \oplus \text{so}(2), \text{f}_4(-20), \text{e}_7(-25)). \)
Here $\gamma' = \gamma'$, i.e., $H'_{j'} = H_{j'}$, and $\zeta(g) = \sum \epsilon_j i H_j$, where without loss of generality $\epsilon_j = 1$ for $j \leq p$ and $\epsilon_j = -1$ for $j > p$. By the explicit form of $\sigma = \text{Ad} \left( \exp \pi i \sum \epsilon_j H_j \right)$

the action of this involution on the root spaces $g_{1C, \alpha}$ can be determined easily and gives us

\[
\Delta_k = \left\{ \pm \frac{1}{2}(\gamma_i - \gamma_j) \mid i < j \leq p \text{ or } p < i < j \right\} \text{ and } \\
\Delta_n = \left\{ \pm \frac{1}{2}(\gamma_i + \gamma_j) \mid i \leq p < j \right\}.
\]

The weight spaces for $\gamma \in \Delta$ are the same as for $\gamma \in \Delta_1$, especially all have the same dimension $d$, such that

\[
\rho = \frac{d}{4} \left[ \sum_{i<j \leq p} (\gamma_i - \gamma_j) + \sum_{p<i<j} (\gamma_i - \gamma_j) + \sum_{i<p<j} (\gamma_i + \gamma_j) \right] = \frac{d}{4} (r+1) \sum_{j=1}^{r} \gamma_j.
\]

With $\Delta_k^+ = \{ \frac{1}{2}(\gamma_i - \gamma_j) \}$ the highest weight has the form $\mu = \sum n_j \gamma_j$ with $n_1 \geq n_2 \geq \ldots \geq n_p$ and $n_{p+1} \geq \ldots \geq n_r$. Alternatively, with $\mu = d \chi_\delta + \mu_s$ we can describe the highest weight with $d \chi_\delta = \lambda(\frac{1}{2r} \sum \epsilon_j \gamma_j)$ and $\mu_s = \sum s_j \gamma_j$, where $\lambda \in \mathbb{R}$ and the $s_j$ are integers with $s_1 \geq s_2 \geq \ldots \geq s_p$ and $s_{p+1} \geq \ldots \geq s_r$. (The $s_j$ are also restricted by $\sum \epsilon_j s_j = 0$, since $\mu_s$ is zero on $g$..) As $d \chi_\delta$ integrates to a representation $\chi_\delta \in \tilde{Z}$ respectively $\tilde{\chi}_\delta \in \tilde{Z}$ we have $\lambda \in \mathbb{Z}$ respectively $\lambda \in \frac{1}{2} \mathbb{Z}$. Finally, to get a representation $\delta = \chi_\delta \otimes \delta_s$ the representations $\chi_\delta$ and $\delta_s$ must be equal on the central subgroup $D = Z \cap G_s$ respectively $\tilde{D}$ for the covering group. We call this the $(D)$-condition.

With $\rho$ as above we get from Theorem 3.3(1)

\[
n_i + n_j < -\frac{d}{2} (r+1), \quad \text{for } i \leq p < j,
\]

which is by the form of $\mu$ equivalent to $n_1 + n_{p+1} < -\frac{1}{2} (r+1)$, and from Proposition 3.4 the condition of $W$-admissibility

\[
n_j \leq 0, \quad \text{for } j = 1, \ldots, r.
\]

To summarize, the Hardy space $H_2(C)$ ($\tilde{H}_2(C)$) is determined by $(n_1, \ldots, n_r) \in \mathbb{Z}^r + (\frac{1}{2} \mathbb{Z}(1, \ldots, 1))$ with

(1) the $(D)$-condition,
(2) $0 \geq n_1 \geq \ldots \geq n_p$ and $0 \geq n_{p+1} \geq \ldots \geq n_r$,
(3) $n_1 + n_{p+1} < -\frac{d}{2} (r+1)$. 


For the covering case $\tilde{H}_{2,\text{odd}}(C)$ is the subspace spanned by those holomorphic discrete series representations which are not representations for $G$.

We remark that this Hardy space is strictly smaller than the Hardy space $H_2(C_{\text{min}})$ respectively $\tilde{H}_{2,\text{odd}}(C_{\text{min}})$. Indeed, by Theorem 3.3, Proposition 3.4, and the following Remark, it is clear that this “maximal” Hardy space is the direct sum $\bigoplus_{\delta_s, \chi \in \hat{Z}(\delta_s)} C_\chi \otimes H_{\delta_s}$, where the $\delta_s$ exhaust the holomorphic discrete series and $\hat{Z}(\delta_s)$ is the set of unitary characters which fulfill with $\delta_s$ the $(D)$-condition, i.e., $\chi \otimes \delta_s$ is a representation of $G$ respectively $\tilde{G}$. To see that this is a bigger space than $H_2(C)$ we calculate, with $\Delta^+ + n$ as above,

$$a \cap C^o_{\text{min}} = -i \sum_{a \leq p < b} \mathbb{R}^+(H_a + H_b).$$

Therefore the condition of admissibility for the minimal cone is $n_a + n_b \leq 0$, for $a \leq p < b$, or equivalently $n_1 + n_{p+1} \leq 0$ compared with $n_1 \leq 0$ and $n_{p+1} \leq 0$ for the cone $C$.

The image of the classical Hardy space is described by

$$n_j \leq -\frac{1}{2} \left( 1 + \frac{d(r - 1)}{2} \right), \quad \text{for } j = 1, \ldots, r.$$

To give a more explicit example, we specialize further to the case $G/H = S(U(1, 1) \times U(1, 1))/SU(1, 1)$, cf. [B97, Section 6.5] for the realization. Here $H_j = (E_{j,j} + E_{j+2,j+2}) \in \mathfrak{su}(2, 2)_C = \mathfrak{sl}(4, \mathbb{C})$ and

$$a = t \cap q = \mathbb{R} \left( \begin{array}{cc} i & -i \\ -i & i \end{array} \right) + \mathbb{R} \left( \begin{array}{cc} i & i \\ -i & -i \end{array} \right),$$

where the first summand is equal to $j(\mathfrak{g})$. Let $\mu \in i(t \cap q)^*$ be the highest weight for a holomorphic discrete series representation of $G/H$. Identifying $t \cap q$ with its dual, we write

$$\mu = a \left( \begin{array}{ccc} 1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & 1 \end{array} \right) + b \left( \begin{array}{ccc} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{array} \right),$$

such that the first summand gives the central character $d\chi_\delta$ and the second $\mu_s$. We have the integrability conditions $a, b \in \mathbb{Z}$, and the $(D)$-condition gives $a + b \in 2\mathbb{Z}$. Further

$$\Delta^+ = \Delta^+ = \left\{ \frac{1}{2} (\gamma_1 + \gamma_2) \right\}$$

and $\rho = \frac{1}{2} (\gamma_1 + \gamma_2)$. Thus Theorem 3.3(1) gives us $b < -1$ and of course no restriction on the “central parameter” $a$.
However we have also the condition of $W$-admissibility. Rewriting this in terms of the parameters $a$ and $b$ shows that $W$-admissibility is equivalent to $|a| \leq -b$.

For the determination of the image of the classical Hardy space in $H_2(C)$ it is straightforward from the definition that $\rho_{1,n} = H_1 + H_2$ (with our identification of $t$ and $t^\ast$). Then the determining equations $(\mu + \rho_{1,n})(H_i) \leq 0$ are equivalent to $|a| \leq -b - 1$. The image of the classical Hardy space therefore misses the two half lines $a + b = 0$ and $a - b = 0$. We remark that this differs from [KØ96, Figure 3] where the missing half lines lie on one side of the cone.

One can deal with the remaining cases as above. With notation as in the preceding examples, we summarize the results as follows:

$(g, h, g_1) = (so^\ast(2n) \oplus so^\ast(2n), so^\ast(2n), so^\ast(4n))$: The Hardy space $H_2(C) = H_2(C_{\text{min}})$ is parametrized by $0 \geq n_1 \geq \ldots \geq n_n$ with $n_i + n_j < -2n + i + j$, for $i \neq j$, and $I(H_{cl})$ is given by $n_1 \leq -n$.

$(g, h, g_1) = (so(2n), so(n, C), su(n, n))$: The Hardy space $H_2(C) = H_2(C_{\text{min}})$ is parametrized by $0 \geq n_1 \geq \ldots \geq n_n$ with $n_i + n_j < -n + \frac{i + j}{2}$, for $i \neq j$, and $I(H_{cl})$ is given by $n_1 \leq -\frac{n}{2}$.

$(g, h, g_1) = (so(2, q) \oplus so(p + 1), so(1, q) \oplus o(p), so(2, n + 1))$: $H_2(C)$ respectively $\tilde{H}_{2, \text{odd}}(C)$ is parametrized by $0 \geq n_1 \geq n_2$ with $n_1 + n_2 \leq -\frac{1}{4}(n + q - p)$ and $I(H_{cl})$ or $I(H_{cl})$ is given by $n_1 \leq -\frac{1}{4}(n + 1)$. Note that for this case the decomposition of $(g, h, \tau)$ into irreducible symmetric algebras gives not a decomposition of the causal structure.

We remark at the end that our results give now the Cauchy-Szegö kernel $K_f(\cdot, \cdot)$ for the subspace $I(H_{cl}) \subset H_2(C)$ respectively $\tilde{I}(H_{cl}) \subset \tilde{H}_{2, \text{odd}}(C)$. Indeed, with $\{\varphi_n\}$ an orthonormal base of $H_2(C)$ the kernel is generally given by

$$K(z, w) = \sum \varphi_n(z)\overline{\varphi_n(w)},$$

such that

$$K_f(z, w) = \sqrt[\ast]{\psi_m(z)}K_{cl}(z, w)\sqrt[\ast]{\overline{\psi_m(w)},}$$

where the kernel $K_{cl}$ for the classical Hardy space is known ([FK94, Prop. X.1.3]). Note also that this kernel is invariant by deck transformations of the covering $\tilde{\Xi}(C) \to \Xi(C)$. Therefore, when we have an isomorphism of Hardy spaces as in Proposition 7.4, we get the corresponding image of the kernel by simply projecting down.

References


312 FRANK BETTEN AND GESTUR ÓLAFSSON


Received December 29, 1998. The second author was supported by NSF grant DMS-9626541, DMS-0070607, and LEQSF (1996-99)-RD-A-12.

Universität Göttingen
Mathematisches Institut
Bunsenstrasse 3-5, D-37073 Göttingen
Germany
E-mail address: betten@cfgauss.uni-math.gwdg.de

Department of Mathematics
Louisiana State University
Baton Rouge, LA 70803
E-mail address: olafsson@math.lsu.edu
PRODUCT FORMULA FOR SELF-INTERSECTION NUMBERS

YANGHYUN BYUN AND SEUNGHUN YI

We provide product formulae for self-intersection numbers in various coefficients.

1. Introduction.

Given an immersion $f : M^m \rightarrow P^{2m}$, $m \geq 1$, from a closed smooth manifold $M$ to a smooth manifold $P$, there is a well-known invariant $I(f)$ called the self-intersection number of $f$. We will consider $I(f)$ in the $\mathbb{Z}_2$-coefficient or in the $\mathbb{Z}$-coefficient if $M, P$ are oriented and $m$ is even or in a $\mathbb{Z}$-module coefficient which is a quotient module of the free $\mathbb{Z}$-module on $\pi_1(P)$ (see §5). These self-intersection numbers can be used to determine whether or not $f$ is regularly homotopic to an embedding if $m \geq 3$. We will consider the problem of what happens to the intersection number if one forms the product of two given immersions.

The problem is much simpler in the case of the type of intersection number which behaves as an obstruction for two submanifolds in an ambient space to get separated from each other by a homotopy (which is not necessarily regular): Let $M_1^{m_1}, M_2^{m_2}$ be submanifolds of $P^{m_1+m_2}$ which intersects transversely and $N_1^{n_1}, N_2^{n_2} \subset Q^{m_1+n_2}$ be another such triple. Then $M_1 \times N_1$ intersects transversely $M_2 \times N_2$ in $P \times Q$ at the points in $(M_1 \cap M_2) \times (N_1 \cap N_2)$. On the other hand, assume we are given two immersions $f : M^m \rightarrow P^{2m}, g : N^n \rightarrow Q^{2n}$ which are completely regular, that is, which are proper, self-transverse and have no triple points (see §2). In general, $f \times g$ is neither self-transverse nor without triple points. For example, if $p, q \in M$ are such that $f(p) = f(q), p \neq q$, we have $(f \times g)(p, y) = (f \times g)(q, y)$ for any $y \in N$. Therefore we must first transform $f \times g$ into a completely regular immersion through a regular homotopy before we calculate the intersection number.

In fact, a special case of the problem arised in the process of deriving the product formula for surgery obstructions in 1970’s (cf. p. 55, [Mo]). It essentially concerned the case when $g$ is an embedding, $N$ is orientable and $Q$ is simply connected in the above and the answer was given by, when rewritten in our notation:

$$I(f \times g) = I(f) \chi(\nu_g),$$

313
where $\chi(\nu_g)$ denotes the integral Euler characteristic of the normal bundle $\nu_g$ of $g$ and the intersection numbers should be understood in such coefficients as introduced in §5 below. This of course coincides with our results in this paper. We are motivated by a different reason (cf. [BY]) and treat the problem in a complete generality.

The following is one of the two main results of this paper, which concerns the intersection number in the $\mathbb{Z}_2$ or the $\mathbb{Z}$-coefficient.

**Theorem A.** Let $f : M^m \to \mathbb{P}^{2m}$, $g : N^n \to \mathbb{Q}^{2n}$ be immersions where $M, N$ are closed smooth manifolds and $P, Q$, smooth manifolds. Then,

(I) for the mod 2 intersection numbers, we have

$$I(f \times g) = \chi(\nu_f)I(g) + I(f)\chi(\nu_g) \in \mathbb{Z}_2,$$

where $\chi(\cdot)$ is the Euler characteristic in the $\mathbb{Z}_2$-coefficient.

Furthermore, assume $M, N, P, Q$ are oriented and $m + n$ is even. Then, for the integral intersection numbers, we have

(II) if both $m, n$ are even,

$$I(f \times g) = 2I(f)I(g) + \chi(\nu_f)I(g) + I(f)\chi(\nu_g) \in \mathbb{Z},$$

where $\chi(\cdot)$ mean the integral Euler characteristic,

(III) and, if both $m, n$ are odd, $I(f \times g) = 0 \in \mathbb{Z}$.

In the above, $\nu_f, \nu_g$ denote the normal bundles. It must be understood that a normal bundle is given the orientation which is consistent with the orientations of the manifolds. Then the formula in (II) above is invariant under the changes of the orientations of $M, N$ and under those of $P, Q$.

In general, the mod 2 or the integral intersection number is not sophisticated enough to be an exact obstruction for the immersion in concern to be regularly homotopic to an embedding. Such an intersection number takes its value in a $\mathbb{Z}$-module which is a quotient module of the free $\mathbb{Z}$-module on the fundamental group of the codomain of the immersion in concern. Theorem B in the last section is none other than a generalization of Theorem A to this case. Even if the former unifies the equalities of the latter, it does so only by sacrificing simplicity of the coefficient in which the intersection number takes its values.

The key idea of the proofs of Theorems A and B might be best revealed by the following simple example.

Consider the case when $I(f) = 0$, $\chi(\nu_f) = 0$ and $P$ is simply connected, $m \geq 3$ (cf. [BY]): Under the condition, we may assume $f$ is an embedding and that $\nu_f$ admits a nowhere vanishing section. This enables us to construct an embedding $F : M \times I \to P$ such that $F(x, 0) = f(x)$ for any $x \in M$, using for instance the exponential map. For simplicity, assume $g$ has only one double point and $p, q \in N, p \neq q$, are such that $g(p) = g(q)$. Choose a smooth function $\varphi : N \to I$ so that $\varphi(p) = 0, \varphi(q) = 1$. Define $\Lambda_t : M \times N \to P \times Q$,
0 \leq t \leq 1$, by $\Lambda_t(x,y) = (F(x,t \varphi(y)), g(y))$ for any $(x,y) \in M \times N$. Then it is straightforward to see that $\Lambda_t$ is a regular homotopy and $\Lambda_1$ is an embedding. Note that this observation is consistent with Theorem A.

The two key steps to the proofs of Theorems A and B are to construct carefully a regular homotopy for each of the immersions $f, g$ and subsequently to use them to obtain a completely regular immersion regularly homotopic to $f \times g$ in a way similar to the above.

2. Basic notions and facts.

Throughout this section, let $f : M \to P$ be a smooth map between connected smooth manifolds.

We say $f$ is an immersion if $f$ is a proper map and $df : T_x M \to T_{f(x)} P$ is injective for each $x \in M$. Let $I$ denote the closed unit interval in the real line $\mathbb{R}$. A homotopy $f_t : M \to P$, $t \in I$, is regular if $f_t$ is an immersion for each $t \in I$.

From now on let $M$ be of dimension $m$ and $P$, of dimension $2m$.

We say an immersion $f : M \to P$ is completely regular if $f$ has no triple points and $f$ is self-transverse, that is, $f$ satisfies the following condition,

$$df_p M + df_{p'} M = T_{f(p)} P = T_{f(p')} P,$$

for any $p, p' \in M$ such that $f(p) = f(p'), p \neq p'$ (cf. \cite{[A]}). We will call $\{p, p'\}$ a double pair of $f$ and $f(p) \in P$ a double point of $f$.

Now assign a metric $d$ on $P$ which induces the topology of $P$. Then, given any immersion $f : M \to P$ and any continuous function $\delta : M \to \mathbb{R}$, $\delta(x) > 0$, $x \in M$, H. Whitney (\cite{Wh}) has shown that there is a regular homotopy $f_t$, $t \in I$, such that $f_0 = f$ and $f_1$ is a completely regular immersion and $d(f(x), f_t(x)) < \delta(x)$ for any $t \in I, x \in M$.

In the rest of this section, we assume further that $M$ is a closed manifold.

If $f : M \to P$ is a completely regular immersion, one may define the intersection number $I(f)$ of $f$ as follows: (i) For the mod 2 intersection number, one defines $I(f) \in \mathbb{Z}_2$ as the number of the double points mod 2. (ii) Assume that $M$, $P$ are oriented and $m$ is even. Then one may define the integral intersection number as follows: Let $r = f(p) = f(p'), p \neq p'$, be a double point of $f$. Let $v = (v_1, v_2, \ldots, v_m)$, $v' = (v'_1, v'_2, \ldots, v'_m)$ be sequences of tangent vectors which represent the orientation of $M$ at $p$ and $p'$, respectively. If the sequence of tangent vectors $(dfv, dfv') = (dfv_1, dfv_2, \ldots, dfv_m, dfv'_1, dfv'_2, \ldots, dfv'_m)$ represents the orientation of $P$ at $r$, write $\varepsilon_r = +1$ and, otherwise, write $\varepsilon_r = -1$. Note that $\varepsilon_r$ remains unchanged even if we interchange $p, p'$. Define $I(f) = \sum_r \varepsilon_r \in \mathbb{Z}$, where $r$ runs through all the double points of $f$.

If $f, g$ are completely regular immersions which are regularly homotopic to each other, then we have $I(f) = I(g)$: According to J. Cerf (\cite{C}), for generic regular homotopy, the double points vary continuously except at a finite set
of points at each of which a pair of double points appear or disappear. If \( m \) is even, the two has opposite values for \( \varepsilon_r \). Furthermore, since every immersion is regularly homotopic to a completely regular immersion, it follows that \( I(f) \) is well-defined for any immersion \( f \).

Now assume \( m \geq 3 \) and \( P \) is simply connected. Let \( I(f) \) denote the mod 2 intersection number if the dimension of \( M \) is odd or \( M \) is unorientable and, in the remaining case, the integral intersection number. Then \( I(f) \) vanishes if and only if the regular homotopy class of \( f \) can be represented by an embedding, which is a consequence of the Whitney trick (cf. [Mi], [Wh] and §5 of this paper).

3. A model case.

Throughout this section, let \( M^m, N^n, m, n \geq 1 \), be smooth manifolds and \( f, g \), completely regular immersions from \( M, N \), respectively, into \( P^{2m} \) and into \( Q^{2n} \), each of which has only one double point. Furthermore, we assume that both \( \nu_f, \nu_g \) admit nowhere vanishing sections. Then we will prove the following.

**Proposition 3.1.** The product \( f \times g \) is regularly homotopic to a completely regular immersion with exactly two double points. Furthermore, assume \( M, N, P, Q \) are oriented and \( m+n \) is even. Then the signs of the two double points differ from each other by multiplication by \((-1)^n\). If both \( m, n \) are even, then both of the signs for the two double points are the multiplication of the sign of the double point of \( f \) with that of \( g \).

To prove 3.1, we need the following lemma which will be proved later in this section.

**Lemma 3.2.** There is a smooth regular homotopy \( f_t : M \to P, t \in I \), such that \( f_0 = f \) and the following conditions hold:

(i) \( f_t \) is a completely regular immersion with exactly one double pair \( \{p_t, p'_t\} \) for each \( t \),

(ii) the map \( I \times \{0, 1\} \to M \) which sends \((t, 0)\) to \( p_t \) and \((t, 1)\) to \( p'_t \) is a smooth embedding,

(iii) ‘\( f_t(x) = f_s(y), (x, t) \neq (y, s) \)’ implies that ‘\( (x, y) = (p_s, p'_t) \) or \( (x, y) = (p'_s, p_t) \)’ and

(iv) \( f_t \) meets \( f_s \) transversely if \( t \neq s \).

**Proof of 3.1.** Let \( f_t, \{p_t, p'_t\} \) be as in Lemma 3.2 and also let \( g_t : N \to Q, t \in I \), be a smooth regular homotopy for \( g \) satisfying the conditions of 3.2 with double pairs \( \{q_t, q'_t\} \).

Choose a smooth function \( \varphi : M \to I \) which is constantly 1 on a neighborhood of \( \{p_t | t \in I\} \) and constantly 0 on a neighborhood of \( \{p'_t | t \in I\} \). Likewise choose a smooth function \( \psi : N \to I \) satisfying the same condition for the two sets \( \{q_t | t \in I\}, \{q'_t | t \in I\} \).
Then we define a homotopy \( \Lambda_t : M \times N \to P \times Q, t \in I, \) by
\[
\Lambda_t(x, y) = (f_{\psi(t)}(x), g_t(x))(y).
\]
Then it is straightforward to see that \( \Lambda_t \) is a smooth homotopy through
immersions such that \( \Lambda_0 = f \times g \).

We must show that \( \Lambda_1 = \Lambda \) has only two double points.

Assume \( \Lambda(x, y) = \Lambda(x', y') \) and \( (x, y) \neq (x', y') \). Then we have \( x \neq x' \) or \( y \neq y' \).

First consider the case \( x \neq x' \). Then we have from \( f_{\psi(t)}(x) = f_{\psi(t')}(x') \) that
\[
(x, x') = (p_{\psi(t)}, p'_{\psi(t)}) \text{ or } (x, x') = (p'_{\psi(t')}, p_{\psi(t)}).
\]
Assume \( (x, x') = (p_{\psi(t')}, p'_{\psi(t)}) \), it follows that \( \psi(x) = 1, \psi(x') = 0 \) and that \( g_1(y) = g_0(y') \), which means that \( (y, y') \) is \( (q_0, q_0') \) or \( (q_1, q_1') \). If \( (y, y') = (q_0, q_0') \), then \( (x, x') = (p_0, p_1') \) and, if \( (y, y') = (q_1, q_1') \), then \( (x, x') = (p_1, p_0) \).
Thus we have in this case as the double pairs for \( \Lambda \)
\[
\{(p_0, q_0), (p_1, q_1')\}, \{(p_1, q_0), (p_0, q_1)\}.
\]
Assume \( (x, x') = (p'_{\psi(t')}, p_{\psi(t)}) \) and proceed similarly as in the above. Then we obtain the same two double pairs for \( \Lambda \) as in the above.

Now assume \( y \neq y' \). Then, from \( g_{\varphi(x)}(y) = g_{\varphi(x')}(y') \), we may easily infer that \( \psi(y) \neq \psi(y') \). Then, from \( f_{\psi(y)}(x) = f_{\psi(y')}(x') \), we conclude that \( x \neq x' \). Thus this case reduces to the case where \( x \neq x' \).

We conclude that \( \{(p_0, q_0), (p_1, q_1')\}, \{(p_1, q_0), (p_0, q_1)\} \) are the only two
double pairs for \( \Lambda \).

That \( \Lambda \) is self-transverse follows from the fact that \( f_0, f_1 \) are transverse
to each other as well as \( g_0, g_1 \) together with the fact that \( \varphi, \psi \) are constant
on each of some neighborhoods of \( p_i, p'_i, q_i, q'_i, i = 0, 1 \).

Finally we prove the last statement of the proposition.

Let \( v_i = (v_{1,i}, v_{2,i}, \ldots, v_{m,i}) \), \( v'_i = (v'_{1,i}, v'_{2,i}, \ldots, v'_{m,i}) \) and \( w_t = (w_{1,t}, w_{2,t}, \ldots, w_{m,t}) \) be sequences of vectors, continuously
parameterized by \( t \in I \), representing the given orientations of \( M \) and \( N \) at
\( p_t, p'_t \) and at \( q_t, q'_t \) respectively.

Let \( \varepsilon(\omega) \) be \( 1 \) or \( -1 \) for each sequence \( \omega \) of independent \( 2(m+n) \) tangent
vectors in \( T_{(x,y)} P \times Q \), \( (x, y) \in P \times Q \), according to whether or not it
represents the orientation of \( P \times Q \), which is non other than the product
orientation.

Then \( \varepsilon(df v_0, dg w_0, df v'_t, dg w'_t) \), \( \varepsilon(df v_1, dg w_1, df v'_0, dg w'_0) \) are constant
for \( t \in I \) and, by the usual sign sign convention, we have
\[
\varepsilon(df v_0, dg w_0, df v'_0, dg w'_0) = (-1)^{2mn+n^2} \varepsilon(df v_0, dg w_0, df v'_0, dg w'_0).
\]
Note that \( 2mn + n^2 \equiv n \mod 2 \). Thus we conclude that
\[
\varepsilon(df v_0, dg w_0, df v'_0, dg w'_1) = (-1)^{n} \varepsilon(df v_1, dg w_0, df v'_0, dg w_1).
\]
Proof. Let \(\Lambda(p_0, q_0) = \Lambda(p_1, q_1)\) and the right hand side is the intersection number at \(\Lambda(p_0, q_0) = \Lambda(p_1, q_1)\). These observations also proves the last statement of the proposition. \(\square\)

The rest of this section will be devoted to the proof of Lemma 3.2.

Write \(r = f(p) = f(p')\), \(p, p' \in M, p \neq p'\). Let \(D^m_\rho\) denote the open disk in \(\mathbb{R}^m\) of radius \(\rho > 0\) centered at the origin.

To prove 3.2, we will make use of Lemma 3.3 and Lemma 3.4 below.

**Lemma 3.3.** There is a coordinate neighborhood \(\psi : V \to \mathbb{R}^{2m} = \mathbb{R}^m \times \mathbb{R}^m\) of \(r\), such that \(\psi(r) = 0\) and there are disjoint open neighborhoods \(U, U' \subset M\) of \(p, p'\) so that \(f^{-1}V = U \cup U'\) and \(\psi f U = \mathbb{R}^m \times \{0\}, \psi f U' = \{0\} \times \mathbb{R}^m\).

**Proof.** Let \(\psi_0 : V_0 \to \mathbb{R}^{2m}\) be a coordinate neighborhood of \(r\). Also let \(\varphi_0 : U_0 \to \mathbb{R}^m, \varphi_0' : U_0' \to \mathbb{R}^m\) be coordinate neighborhoods of \(p, p'\), respectively, such that \(U_0, U_0' \subset f^{-1}V_0, U_0 \cap U_0' = \emptyset\). We choose \(\psi_0, \varphi_0, \varphi_0'\) so that \(\psi_0(r) = 0\) and \(\varphi_0(p) = \varphi_0'(p') = 0\).

Consider \(h : \mathbb{R}^{2m} = \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^{2m}\), defined by

\[
h(x, y) = \psi_0 f \varphi_0^{-1}(x) + \psi_0 f \varphi_0'^{-1}(y).
\]

Then it is straightforward to see that \(dh_0 : T_0 \mathbb{R}^{2m} \to T_0 \mathbb{R}^{2m}\) is an isomorphism exploiting the fact that \(f\) is self-transverse. Therefore, there is an \(\epsilon > 0\) such that \(h\) restricts to a diffeomorphism

\[
h_1 : D^m_\epsilon \times D^m_\epsilon \to h(D^m_\epsilon \times D^m_\epsilon).
\]

Then we consider the coordinate neighborhood of \(r\),

\[
\psi_1 = h_1^{-1} \psi_0 : \psi_0^{-1}(h(D^m_\epsilon \times D^m_\epsilon)) \to D^m_\epsilon \times D^m_\epsilon.
\]

Now choose \(\delta > 0\), exploiting the fact that \(f\) is proper, so that

\[
\psi_1^{-1}(D^m_\delta \times D^m_\delta) \cap f(M - (\varphi_0^{-1} D^m_\epsilon \cup \varphi_0'^{-1} D^m_\epsilon)) = \emptyset.
\]

Then we choose \(\psi\) as the restriction \(\psi_1^{-1}(D^m_\delta \times D^m_\delta) \to D^m_\delta \times D^m_\delta\) of \(\psi_1\) followed by a diffeomorphism \(\alpha \times \alpha : D^m_\delta \times D^m_\delta \to \mathbb{R}^m \times \mathbb{R}^m\), where \(\alpha\) is a diffeomorphism. \(\square\)

Note that, for any riemannian manifold, there is the exponential map defined in terms of the geodesics, which we denote by \(\exp\). In general, the map \(\exp\) is a smooth map from an open neighborhood of the zero section in the tangent vector bundle into the manifold. On the other hand, once a riemannian metric \(\langle \cdot, \cdot \rangle\) is introduced on \(P\), we will identify the normal bundle \(N\) with the subspace

\[
\{(x, v) : x \in M, v \in T_{f(x)}P \text{ and } \langle v, w \rangle = 0 \text{ for any } w \in df_x T_x M\}
\]
of $X \times TP$. Let $\pi : \nu_f \to M$ denote the projection. Then there is a map from a neighborhood of the zero section in $\nu_f$ into $P$ which maps $(x, v) \in \nu_f$ to $\exp(v)$. We denote this map again by $\exp$ slightly abusing the notation.

Note that $\exp$ is an embedding on a neighborhood of the zero section of $\nu_f|_A$ if $f|_A : A \to P$ is an embedding for a subspace $A \subset M$.

**Lemma 3.4.** Consider $P$ with any riemannian metric. Let $V \subset P$ any open neighborhood of $r$. Then, for any open neighborhoods $U, U' \subset M$ of $p, p'$ such that $U \cap U' = \emptyset$ and $f(U \cup U') \subset V$, there is an open neighborhood $T$ of the zero section in $\nu_f$ satisfying the following conditions:

(i) $T \cap \pi^{-1}U, T \cap \pi^{-1}U' \subset \exp^{-1}V$.
(ii) $\exp$ is an embedding on each of $T \cap \pi^{-1}(M - U'), T \cap \pi^{-1}(M - U)$.

**Proof.** Let $U_1, U_1'$ be open neighborhoods of $p, p'$, respectively, such that $U_1 \subset U, U_1' \subset U'$.

Then, since $f$ embeds each of $M - U_1', M - U_1$ into $P$ (note that $f$ is proper), there are open neighborhoods $T_1, T_1'$ of the zero sections respectively in $\pi^{-1}(M - U_1')$ and in $\pi^{-1}(M - U_1)$ so that $\exp$ is an embedding. Furthermore, since $U \subset M - U_1', U' \subset M - U_1$ and $f(U \cup U') \subset V$, we may choose $T_1, T_1'$ so that $\exp(T_1 \cap \pi^{-1}U), \exp(T_1' \cap \pi^{-1}U') \subset V$.

Then we define $T$ as follows:

$$T = (T_1 \cap \pi^{-1}(M - U_1')) \cap (T_1' \cap \pi^{-1}(M - U_1)) \cup (T_1 \cap \pi^{-1}U) \cup (T_1' \cap \pi^{-1}U').$$

It is straightforward to see that $T$ satisfies all the conditions of the lemma. □

**Proof of 3.2.** Cover $P$ by a locally finite collection of coordinate neighborhoods $\psi_i : U_i \to \mathbb{R}^{2m}, i = 1, 2, \ldots$, such that (a) $r \in U_1$ and $\psi_1 : U_1 \to \mathbb{R}^{2m} = \mathbb{R}^m \times \mathbb{R}^m$ satisfies the conditions of Lemma 3.3 and (b) $\psi_1^{-1}(D_2^m \times D_2^m)$ does not intersect $U_i$ for any $i > 1$. Write $C_2 = \psi_1^{-1}(D_2^m \times D_2^m)$.

Construct a riemannian metric on $P$ by piecing together the pull-back metrics on $U_i$'s of the standard metric on $\mathbb{R}^{2m}$ using a partition of unity for $\{U_i| i = 1, 2, \ldots\}$.

Then, by the condition (b), $\psi_1 : C_2 \to D_2^m \times D_2^m$ is an isometry.

Let $T$ be an open neighborhood in $\nu_f$ of the zero section which satisfies the conditions (i), (ii) of Lemma 3.4 with respect to $C_2$ and with the open neighborhoods $U, U'$ of $p, p'$ defined by $f(U) = D_0^m \times \{0\}, f(U') = \{0\} \times D_0^m$. Write $\pi_T : T \to M$ for the restriction of the projection $\pi : \nu_f \to M$.

By assumption, there is a section $\alpha : M \to \nu_f$ such that $\alpha(x) \neq 0$ for any $x \in M$. We may assume $\alpha(x) \in T$ for any $x \in M$.

Allow ourselves a slight abuse of notation so that we may mean by $\alpha(x)$ the vector $v \in T_{f(x)}P$ for which $(x, v)$ is the value of $\alpha$ at $x$. We may choose $\alpha$ so that $\alpha(x)$'s, $x \in U$, are parallel in $C_2$ as well as $\alpha(x')$'s, $x' \in U'$, and also so that $\langle \alpha(x), \alpha(x) \rangle^{\frac{1}{2}} < 1$ for any $x \in U \cup U'$. 

The following proves Lemma 3.2.

**Claim.** The homotopy \( f_t : M \to V \) defined by the rule
\[
f_t(x) = \exp(t\alpha(x)), \quad 0 \leq t \leq 1,
\]
satisfies all the conditions of Lemma 3.2.

**Proof.** For any tangent vector \( v \in T_xP, \ x \in C_2 \), let \( v_0 \) denote the the tangent vector at \( r \in P \) parallel to \( v \) in \( C_2 \).

First of all, we observe the following: Assume \( \exp(v) = \exp(w), v, w \in T, v \neq w \). Then, from our choice of \( T \), it follows that \( (v, w) \in \pi^{-1}_T U \times \pi^{-1}_T U \) or \( (v, w) \in \pi^{-1}_T U' \times \pi^{-1}_T U' \). And exploiting the flatness of \( C_2 \), we may easily conclude that \( f\pi_T(v) = \exp(w_0), f\pi_T(w) = \exp(v_0) \).

It is clear that \( f_t \) is an immersion for each \( t \in I \) since it is the immersion \( \tau \alpha : M \to T \) followed by the local diffeomorphism \( \exp : T \to P \).

Assume \( f_t(x) = f_t(y) \) for some \( x, y \in M, x \neq y \). Then \( \exp(\tau\alpha(x)) = \exp(\tau\alpha(y)) \). Since \( \tau\alpha(x), \tau\alpha(y) \in T \), we must have: \( (x, y) \in U \times U \) or \( (x, y) \in U' \times U' \).

If \( (x, y) \in U \times U' \), we have
\[
f(x) = \exp(\tau\alpha(y)) = \exp(\tau\alpha(p')), \quad f(y) = \exp(\tau\alpha(x)) = \exp(\tau\alpha(p')).
\]
Similarly, if \( (x, y) \in U' \times U \), we have \( x = \exp(\tau\alpha(p)), y = \exp(\tau\alpha(p')) \).

Therefore, let \( p_t \in U, p'_t \in U', t \in I, \) be defined by:
\[
f(p_t) = \exp(\tau\alpha(p'))), f(p'_t) = \exp(\tau\alpha(p)).
\]
Then \( f_t \) has only one double pair, \( \{p_t, p'_t\} \), for each \( t \in I \). Note that \( p_0 = p, p'_0 = p' \).

Also note that \( f_t \) is self-transverse since \( (df_t)_{p_t} T_{p_t} M \) is parallel to \( df_t T_{p_t} M \) in \( C_2 \) and \( (df_t)_{p'_t} T_{p'_t} M \) is also parallel to \( df_t T_{p'_t} M \) in \( C_2 \). This proves that the homotopy \( f_t \), \( t \in I \), satisfies (i). It is clear that the homotopy \( f_t \), \( t \in I \), satisfies (ii) with the double pairs \( \{p_t, p'_t\}, t \in I \), since we have \( \{p_t\} \subset U, \{p'_t\} \subset U', U \cap U' = \emptyset \).

Assume \( f_t(x) = f_s(y), (x, t) \neq (y, s) \). If \( t = s \), then (iii) follows from (i). If \( t \neq s \), then from the equality \( \exp(\tau\alpha(x)) = \exp(\tau\alpha(y)) \) it follows that we must have that \( (x, y) \in U \times U' \) or \( (x, y) \in U' \times U \) and that \( f(x) = \exp(\tau\alpha(y)) \) and \( f(y) = \exp(\tau\alpha(x)) \). If \( (x, y) \in U \times U' \), then \( \alpha(x), \alpha(y) \) are parallel respectively to \( \alpha(p), \alpha(p') \) in \( C_2 \), which leads to the conclusion \( (f(x), f(y)) = (\exp(\tau\alpha(p_0)), \exp(\tau\alpha(p_0))) \). Thus we have \( (x, y) = (p_s, p'_s) \). Likewise we conclude that if \( (x, y) \in U' \times U \) then \( (x, y) = (p'_s, p_t) \). This proves that the homotopy \( f_t \), \( t \in I \), satisfies (iii).

Assume \( t \neq s \). Note that \( (df_t)_{p_t} T_{p_t} M \) is parallel to \( df_t T_{p_t} M \) in \( C_2 \) and \( (df_t)_{p'_t} T_{p'_t} M \) to \( df_t T_{p'} M \). Thus \( f_t \) meets \( f_s \) transversely at \( f_t(p_s) = f_s(p'_t) \). Likewise we may conclude that \( f_t \) meets \( f_s \) transversely at \( f_t(p'_s) = f_s(p_t) \) as well. Thus the homotopy \( f_t, t \in I \), satisfies (iv). \[\square\]
4. Proof of Theorem A.

We begin this section by recalling the following well-known fact.

**Lemma 4.1.** Let \( m \) be a positive odd integer. Then any orientable vector bundle of rank \( m \) over an orientable manifold \( M^m \) admits a nowhere vanishing section.

**Proof.** The Euler class of an oriented vector bundle of an odd rank is 2-torsion (cf. p. 98, [MS]). Since \( H^m(M; Z) \) has no torsion, this means the Euler class of the bundle vanishes. However the Euler class is the exact obstruction for an oriented vector bundle in concern to admit a nowhere vanishing section. This completes the proof. \( \square \)

Assume \( M^m \) is oriented and \( m \) is even. Let \( \xi \) be a smooth oriented vector bundle of rank \( m \) over \( M \). We will denote the total space of \( \xi \) again by \( \xi \). Then \( \xi \) itself is an oriented manifold with the orientation determined by those of the bundle \( \xi \) and \( M \). Assume \( s \) is a smooth section of \( \xi \) which meets the zero section transversely. If \( s(p) = 0 \), let \( \varepsilon(p) \) be the sign of the intersection \( p \) between the two embeddings of \( M \) into \( \xi \), that is, between the zero section and \( s \). Then the integral Euler characteristic \( \chi(\xi) \) satisfies the equality, \( \chi(\xi) = \sum_p \varepsilon(p) \), in which \( p \) runs through all the zero points of \( s \).

The proof of Theorem A is immediate from the following.

**Proposition 4.2.** Let \( f : M^m \to P^{2m}, \ g : N^n \to Q^{2n} \) be completely regular immersions with respective double points \( r_1, r_2, \ldots, r_1, r_2, \ldots, r_1, r_2, \ldots, r_1, r_2, \ldots, r_1, r_2, \ldots, r_1, r_2, \ldots, \in P, s_1, s_2, \ldots, \in Q \). Assume there are sections \( \alpha : M \to \nu_f \) and \( \beta : N \to \nu_g \) which meet the zero sections transversely respectively at \( a_1, a_2, \ldots, \in M, \{a_1, a_2, \ldots\} \cap f^{-1}\{r_1, r_2, \ldots\} = \emptyset \), and at \( b_1, b_2, \ldots, \in N, \{b_1, b_2, \ldots\} \cap g^{-1}\{s_1, s_2, \ldots\} = \emptyset \). Then,

(a) \( f \times g : M \times N \to P \times Q \) is regularly homotopic to a completely regular immersion \( \Lambda \) which has, as its double points, two for each of the ordered pairs \( (r_1, s_1) \) and one for each of \((a_1, s_1), (r_1, b_1)\), all of which are distinct among themselves.

Furthermore, assume \( m+n \) is even and \( M, N, P, Q \) are oriented. Then we have that

(b) if \( x_{i,j}, x_{i,j'} \) are the two double points of \( \Lambda \) corresponding to each of \((r_i, s_j)\), we have \( \varepsilon(x_{i,j}) = (-1)^n \varepsilon(x_{i,j'}) \) and,

(c) if in addition both \( m, n \) are even, we have \( \varepsilon(x_{i,j}) = \varepsilon(x_{i,j'}) = \varepsilon(r_i) \varepsilon(s_j) \) and if \( y_{k,l}, z_{i,l} \) denote the double points of \( \Lambda \) corresponding respectively to \((a_k, s_j), (r_i, b_l)\), we have \( \varepsilon(y_{k,l}) = \varepsilon(a_k) \varepsilon(s_j), \varepsilon(z_{i,l}) = \varepsilon(r_i) \varepsilon(b_l) \).

The proof of 4.2 will be given in the later of this section. Here we provide:

**Proof of Theorem A.** We may assume \( f, g \) are completely regular, say, with respective double points \( r_1, r_2, \ldots, r_k \in P \) and \( s_1, s_2, \ldots, s_\lambda \in Q \). Also let
\(\alpha : M \rightarrow \nu_f, \beta : N \rightarrow \nu_g\) be the sections which meet the zero section transversely, say, respectively at \(a_1, a_2, \ldots, a_\mu \in M\) and at \(b_1, b_2, \ldots, b_\nu \in N\). We may assume \(\{a_1, a_2, \ldots, a_\mu\} \cap f^{-1}\{r_1, r_2, \ldots, r_\kappa\} = \emptyset, \{b_1, b_2, \ldots, b_\nu\} \cap g^{-1}\{s_1, s_2, \ldots, s_\lambda\} = \emptyset\). Then let \(\Lambda : M \times N \rightarrow P \times Q\) be a completely regular immersion regularly homotopic to \(f \times g\) as in 4.2.

Then the statement (I) is clear since \(\Lambda\) has \((2\kappa \lambda + \mu \lambda + \kappa \nu)\) double points by (a) of 4.2.

Now assume \(M, N, P, Q\) are oriented and \(m + n\) is even.

If both \(m, n\) are odd, then by 4.1 we may assume that \(\{a_1, a_2, \ldots, a_\mu\} = \emptyset, \{b_1, b_2, \ldots, b_\nu\} = \emptyset\). Then by (a), (b) of 4.2, \(\Lambda\) has \(2\kappa \lambda\) double points, two for each \((r_i, s_j), i = 1, 2, \ldots, \kappa, j = 1, 2, \ldots, \lambda\) whose signs are opposite to each other. This proves the clause (III).

Also if both \(m, n\) are even, then by (a), (b) of 4.2, we have

\[
I(\Lambda) = 2 \sum_{i,j} \varepsilon_{r_i} \varepsilon_{s_j} + \sum_{k,j} \varepsilon(a_k) \varepsilon_{s_j} + \sum_{i,t} \varepsilon_{r_i} \varepsilon(b_t). 
\]

Thus it follows that \(I(f \times g) = I(\Lambda) = 2I(f)I(g) + \chi(\nu_f)I(g) + I(f)\chi(\nu_g) \in \mathbb{Z}\) as claimed in the clause (II).

To prove 4.2, we need the following generalization of Lemma 3.2. For more details of the proof, one must refer to the Proof of 3.2.

**Lemma 4.3.** Let \(f : M \rightarrow P, r_1, r_2, \ldots \in P, \alpha : M \rightarrow \nu_f, a_1, a_2, \ldots \in M\) be as in 4.2. Then there is a smooth regular homotopy \(f_t : M \rightarrow P, t \in I, \) such that \(f_0 = f\) and satisfying the following conditions:

(i) \(f_t\) is a completely regular immersion with exactly one double pair \(\{p_{i,t}, p'_{i,t}\}\) for each \(t \in I\) and for each \(i = 1, 2, \ldots,\),

(ii) the map \(I \times \{0, 1\} \times \{1, 2, \ldots\} \rightarrow M\) which sends \((t,0,i)\) to \(p_{i,t}\) and \((t,1,i)\) to \(p'_{i,t}\) is a smooth embedding,

(iii) \('f_t(x) = f_s(y), (x,t) \neq (y,s)' implies that \('x,y) = (p_{i,s},p'_{i,t}), or (x,y) = (p'_{i,s},p_{i,t}), for some i = 1, 2, \ldots, or x = y = a_j, for some j = 1, 2, \ldots, '"

(iv) and \(f_t\) meets \(f_s\) transversely if \(t \neq s\).

**Proof.** First choose disjoint coordinate neighborhoods \(\psi_i : V_i \rightarrow \mathbb{R}^{2m}\) of \(r_i, i = 1, 2, \ldots\) so that each of them satisfies the conditions of Lemma 3.3 with some neighborhoods \(U_i, U'_{i}\) of \(p_i, p'_i\), where \(f(p_i) = f(p'_i) = r_i, p_i \neq p'_i\). Then it is straightforward to construct a riemannian metric for which the restriction of \(\psi_i\) to \(\psi_i^{-1}(D^{2m}_m \times D^{2m}_m) \rightarrow D^{2m}_m \times D^{2m}_m\) is an isometry for each \(i\).

Write \(U_{i,1}, U'_{i,1} \subset M\) for the open neighborhoods of \(p_i, p'_i\) such that \(\psi_i f U_{i,1} = D^{2m}_i \times \{0\}, \psi_i f U'_{i,1} = \{0\} \times D^{2m}_i\). Then, by slightly generalizing Lemma 3.4 above, we may construct an open neighborhood \(T\) of the zero section in \(\nu_f\) such that \(\exp : T \rightarrow P\) embeds each of \(\pi_T^{-1}(M - \cup_i U_{i,1}), \pi_T^{-1}(M - \cup_i U'_{i,1})\) and
exp \pi^{-1}(U_{i,1} \cup U'_{i,1}) \subset \psi_i^{-1}(D^m \times D^m), \text{ where } \pi_T : T \to M \text{ is the restriction of the projection } \pi : \nu_f \to M.

We may assume the section \( \alpha \) is chosen so that \( \alpha M \subset T \). Then we define a homotopy \( f_t : M \to P, t \in I \) by \( f_t(x) = \exp(t\alpha(x)) \). It is straightforward to see that \( f_t : M \to P, t \in I, \) is a regular homotopy satisfying all the conditions of the lemma.

Proof of 4.2. Let \( f_t, t \in I, \) be as in 4.3 and choose \( g_t, t \in I, \) so that it satisfies the conditions of 4.3 with the double pairs \( \{q_{j,t}, q'_{j,t}\}, j = 1, 2, \ldots \).

Let \( \varphi : M \to I \) be a smooth function which is constantly 1 on a neighborhood of \( \cup_i \{p_{i,t} \mid t \in I\} \cup \{a_1, a_2, \ldots\} \) and constantly 0 on a neighborhood of \( \cup_i \{p'_{i,t} \mid t \in I\}. \) Likewise choose a smooth function \( \psi : N \to I \) satisfying the same condition for the two sets, \( \cup_j \{q_{j,t} \mid t \in I\} \cup \{b_1, b_2, \ldots\} \), \( \cup_j \{q'_{j,t} \mid t \in I\}. \)

Now we define a homotopy \( \Lambda_t : M \times N \to P \times Q, t \in I, \) as before, by

\[ \Lambda_t(x, y) = (f_t\varphi(y)(x), g_t\psi(x)(y)). \]

Then it is straightforward to see that \( \Lambda_t, t \in I, \) is a regular homotopy such that \( \Lambda_0 = f \times g. \)

Write \( \Lambda = \Lambda_1 \) and assume \( \Lambda(x, y) = \Lambda(x', y'), (x, y) \neq (x', y'). \)

Then, we obtain, as the double pairs of \( \Lambda, \)

\[ \{(p_{i,0}, q_{j,0}, p'_{i,1}, q'_{j,1})\}, \{(p_{i,1}, q_{j,0}), (p'_{i,0}, q_{j,1})\}, \]

for each \( i, j, \) and

\[ \{(a_k, q_{j,1}), (a_k, q'_{j,1})\}, \{(p_{i,1}, b_l), (p'_{i,1}, b_l)\}, \]

for each \( k, j \) and for each \( i, l. \) The former are the two double pairs corresponding to \( (r_i, s_j) \), which are obtained essentially by 3.1. Note that in this case we have \( x \neq x', y \neq y'. \) The latter are the double pairs respectively corresponding to \( (a_k, s_j), (r_i, b_l) \), for which we have \( x = x', y \neq y' \) or \( x \neq x', y = y'. \)

Since \( \Lambda \) is clearly self-transverse, this proves the statement (a). The clause (b) has been essentially proved by Lemma 3.2. The first part of (c) also has been proved by Lemma 3.2 and its last part is clear. \( \square \)

5. The non-simply connected case.

We begin this section with a detailed description of the intersection number which behaves as the exact obstruction for a given immersion to be regularly homotopic to an embedding even when the relevant manifolds are not simply connected. In what follows, the usual notational conventions for the paths must be understood.

Let \( f : M^m \to P^{2m} \) be a completely regular immersion between connected smooth manifolds. Assume \( M \) is closed.

First of all, we recall when the Whitney trick can be applied to cancel two double points (cf. [Mi], [Wh]). Let \( r_0, r_1 \in P \) be two double points
of \( f \) and \( \{p_0, p'_0\} = f^{-1}\{r_0\}, \{p_1, p'_1\} = f^{-1}\{r_1\} \). Assume there are paths, 
\( \alpha, \alpha' : I \to M, \) such that:

(i) \( \alpha(0) = p_0, \alpha(1) = p_1, \alpha'(0) = p'_0, \alpha'(1) = p'_1, \)

(ii) \( (f\alpha) \cdot (f\alpha')^{-1} \) is a contractible loop in \( P \) and,

(iii) for continuously parameterized orientations \( \omega_t, \omega'_t \) respectively of 
\( T_{\alpha(t)}M, T_{\alpha'(t)}M, \) the signs \( \varepsilon(df\omega_0, df\omega'_0), \varepsilon(df\omega_1, df\omega'_1) \) are opposite, 
when the signs are determined with respect to the orientations of 
\( T_{\gamma_0}P, T_{\gamma_1}P \) which are the restrictions of a continuously parameterized 
orientation of \( T_{f\alpha(t)}P, t \in I. \)

Furthermore, assume \( m \geq 3. \) Then \( \alpha, \alpha' \) can be chosen as smooth embedd-
ings and there is a smoothly embedded disk in \( P \) which meets \( fM \) on two 
points which extend the arcs \( f\alpha I, f\alpha'I \) slightly and subsequently one may use 
the Whitney trick to cancel the two double points.

Now choose a base point \( x_0 \in M \) and write \( f(x_0) = z_0 \) and fix orientations 
for \( T_{z_0}M, T_{z_0}P. \) Let \( r \in P \) be a double point of \( f \) and \( \{p, p'\} = f^{-1}\{r\}. \)
Choose paths \( \alpha, \alpha' : I \to M \) such that \( \alpha(0) = \alpha'(0) = x_0 \) and \( \alpha(1) = p, \alpha'(1) = p'. \) Then \( f\alpha(1) = f\alpha'(1) = r \) and therefore \( (f\alpha) \cdot (f\alpha')^{-1} : I \to P \) is a loop based at \( z_0. \) Write \( \gamma_r = [(f\alpha) \cdot (f\alpha')^{-1}] \in \pi_1(P). \) Also we 
decide the sign \( \varepsilon_r = \pm 1 \) as follows: Use the paths \( \alpha, \alpha' \) together with the 
orientation of \( T_{z_0}M \) to orient \( T_pM, T_{p'}M \) and use the path \( f\alpha \) together with 
the orientation of \( T_{z_0}P \) to orient \( T_rP. \) We write \( \varepsilon_r = 1 \) if the orientation 
of \( T_rP \) coincides with the one determined by the ordered pair of oriented 
subspaces \( dfT_pM, dfT_{p'}M \) and \( \varepsilon_r = -1 \) otherwise. We will consider \( \varepsilon_r\gamma_r \) in 
\( \mathbb{Z}\pi_1(P), \) the free \( \mathbb{Z} \)-module on \( \pi_1(P). \)

However, \( \varepsilon_r\gamma_r \) depends on the choice of \( \alpha, \alpha'. \) Let \( w_M : \pi_1(M) \to \{\pm 1\}, \)
\( w_P : \pi_1(P) \to \{\pm 1\} \) be the orientation characters, that is, the homomor-
phisms which respectively represent the first Stiefel-Whitney classes of \( M \) and \( P. \) Then for any \( a, a' \in \pi_1(M), \) the element

\[
w_P(f_*a)w_M(a)w_M(a')\varepsilon_r(f_*a)\gamma_r(f_*a')
\]

could have been chosen instead of \( \varepsilon_r\gamma_r \) if we chose \( \alpha, \alpha' \) differently. On the 
other hand, if we interchanged \( p, p', \) \( (-1)^m w_P(\gamma_r) \varepsilon_r\gamma_r^{-1} \) could have been 
chosen by the same process. Here the multiplication by \( w_P(\gamma_r) \) is due to 
our using the path \( f\alpha' \) to orient \( T_rP \) instead of \( f\alpha. \)

Therefore we denote by \( K_f \) the submodule of \( \mathbb{Z}\pi_1(P) \) generated by 
\[
\{b - w_P(f_*a)w_M(a)w_M(a')(f_*a)b(f_*a'), \quad b - (-1)^m w_P(b)b^{-1} | a \in \pi_1(M), b \in \pi_1(P)\}
\]

and consider the quotient module

\[
\mathbb{Z}\pi_1(P)/K_f \equiv \Gamma_f.
\]
Note that, if there is a homotopy from \( f \) to \( f' \), then there corresponds a natural isomorphism from \( \Gamma_f \) to \( \Gamma_{f'} \). Furthermore, if the homotopy fixes the base point, then \( \Gamma_f = \Gamma_{f'} \) and the isomorphism is just the identity.

Write \( \gamma_r \) to denote the class in \( \Gamma_f \) represented by \( \gamma_r \in \pi_1(P) \). Note that \( \varepsilon_r[\gamma_r] \) is a well-defined element in \( \Gamma_f \) for any double point \( r \), even if each of \( \varepsilon_r, [\gamma_r] \) in general depends on the choice of \( \alpha, \alpha' \). Then, the intersection number \( I_f(f) \) of \( f \) is defined by

\[
I_f(f) = \sum_r \varepsilon_r[\gamma_r] \in \Gamma_f
\]

where \( r \) runs through all the double points of \( f \).

It is straightforward to see that \( I_f(f) \) is invariant of the regular homotopy class of \( f \) up to the natural isomorphisms and that \( I_f(f) \) is well-defined even if \( f \) is only an immersion (see §2 and also p. 46, [Wa]). Also it is clear by construction that \( I_f(f) = 0 \) if and only if \( f \) is regularly homotopic to an embedding, assuming \( m \geq 3 \).

Let \( g : \mathbb{N}^n \to Q^{2n} \) be another immersion from a connected closed manifold \( N \) to a connected manifold \( Q \) where \( N, Q \) have base points \( y_0, w_0 = g(y_0) \) and \( T_{y_0}N, T_{w_0}Q \) are oriented. Consider \( M \times N, \pi_1(P \times Q) \) with the base points \( (x_0, y_0), (z_0, w_0) \) and with the product orientations for \( T_{(x_0, y_0)}M \times N, T_{(z_0, w_0)}P \times Q \). Then, \( I_f(f \times g) \in \Gamma_{f \times g} \) must be well-defined.

Note that there is a natural isomorphism \( \pi_1(P) \times \pi_1(Q) \to \pi_1(P \times Q) \).

We will write \( a \otimes b \) for the image of \( (a, b) \in \pi_1(P) \times \pi_1(Q) \) by this map. For simplicity we let \( w \) denote the orientation character for any manifold in concern.

**Lemma 5.1.** There is a well-defined map \( * : \Gamma_f \times \Gamma_g \to \Gamma_{f \times g} \), defined by extending the rule bilinearly

\[
*(\{a\},\{b\}) \equiv [a] * \{b\} = [a \otimes b] + (-1)^n w(b)\{a \otimes b^{-1}\},
\]

for any \( a \in \pi_1(P), b \in \pi_1(Q) \).

**Proof.** We have that

\[
((-1)^n w(a)\{a^{-1}\}) * \{b\} = (-1)^n w(a)\{a^{-1} \otimes b\} + (-1)^{n+n} w(a) w(b)\{a^{-1} \otimes b^{-1}\} = (-1)^n w(b)\{a \otimes b^{-1}\} + \{a \otimes b\} = [a] * \{b\}.
\]

Similarly we have \( [a] * (-1)^n w(b)\{b^{-1}\}) = [a] * \{b\} \). Also, for any \( c, c' \in \pi_1(M) \), \( d, d' \in \pi_1(N) \), it is straightforward to see the identities:

\[
(w(f_*c)w(c)w(c')[(f_*c)a(f_*c')]) * \{b\} = [a] * \{b\},
\]

\[
[a] * (w(g_*d)w(d)w(d')[(g_*d)b(g_*d')]) = [a] * \{b\}.
\]

□
Let \( e \) denote the identity element for any fundamental group. Even if in general the rule \(([a], [b]) \rightarrow [a \otimes b]\) does not provide a well-defined map from \( \Gamma_f \times \Gamma_g \) into \( \Gamma_{f \times g} \), we have the following, for which we omit the proof:

**Lemma 5.2.** The maps, \( \iota_1 : \Gamma_f \rightarrow \Gamma_{f \times g}, \iota_2 : \Gamma_g \rightarrow \Gamma_{f \times g} \) respectively defined by extending the rules \( \iota_1([a]) = [a \otimes e], \iota_2([b]) = [e \otimes b] \) for any \( a \in \pi_1(P), b \in \pi_1(Q) \) linearly is well-defined.

On the other hand, we need to consider the Euler characteristic of a vector bundle over a manifold in a more general context than before.

Let \( \xi \) be a smooth vector bundle of rank \( l \) over the connected manifold \( L \), which has the base point \( x_0 \) and with a fixed orientation for \( T_{x_0}L \). Assume \( \xi_{x_0} \) is oriented. Let \( s \) be a smooth section of \( \xi \) which meets the zero section transversely. Let \( p \in L \) be such that \( s(p) = 0 \). Let \( \alpha \) be a path from \( x_0 \) to \( p \). Then \( \xi_p, T_pL \) are oriented subspaces of \( T_p\xi \) regarding \( \xi \) itself as a manifold, in which the orientations are respectively determined by the orientations of \( \xi_{x_0}, T_{x_0}L \) together with the path \( \alpha \). Define the sign \( \varepsilon(p) = 1 \) if the orientation of \( T_p\xi \) determined by the ordered pair of oriented subspaces \( dsT_pL, T_pM \) coincides with the one determined by another such pair \( \xi_p, T_pL \) and \( \varepsilon(p) = -1 \) otherwise. Then \( \varepsilon(p) \) does not depend on the choice of the path \( \alpha \) if and only if \( w_1(\xi) = w_1(L) \equiv w_1(TL) \), where \( w_1 \) denotes the first Stiefel-Whitney class. If \( w_1(\xi) = w_1(L) \), \( \chi(\xi) \) is defined by \( \chi(\xi) = \sum_p \varepsilon(p) \in \mathbb{Z} \), where \( p \) runs through all the zeros of \( s \). In fact, this is the Euler characteristic in the twisted integral coefficient (cf. [B]). If \( w_1(\xi) \neq w_1(L) \), then \( \chi(\xi) \) is defined as the number of the zeros of \( s \) modulo 2. We will refer to \( \chi(\xi) \) defined in this way as the Euler characteristic in the twisted coefficient.

Before the statement of the second main result of the paper, we need to observe the following.

**Lemma 5.3.** Assume \( w_1(\nu_f) \neq w_1(M) \). Then \([e] \in \Gamma_f\) is 2-torsion and \( \iota_2(y) \in \Gamma_{f \times g} \) is also 2-torsion for any \( y \in \Gamma_g \).

**Proof.** Note that under the assumption \( f^*TP \) is not orientable. Moreover \( f^*TP \) is orientable if and only if \( w_P(f_a) = 1 \) for any \( a \in \pi_1(M) \). Therefore, there is an \( a \in \pi_1(M) \) such that \( w_P(f_a) = -1 \). Then the following observations prove the lemma: Firstly we have

\[
ed - w(f_a)w(a)w(a^{-1})(f_a)e(f_a^{-1}) = 2e
\]

is in \( K_f \) and secondly for any \( b \in \pi_1(Q) \) we have

\[
e \otimes b - w((f \otimes g)_*(a \otimes e))w(a \otimes e)w(a^{-1} \otimes e)\]

\[
\quad \cdot ((f \times g)_*(a \otimes e))(e \otimes b)((f \times g)_*(a^{-1} \otimes e))
\]

is \( 2e \otimes b \) and it is in \( K_{f \times g} \).

Therefore, if \( w_1(\nu_f) \neq w_1(M) \), there is an action by mod 2 integers on \( \iota_2(y) \) coming from the \( \mathbb{Z}\)-action and the product \( \chi(\nu_f)\iota_2(y) \) in 5.1 below...
should be understood in this sense for any \( y \in \Gamma_g \). Similarly with \( \chi(\nu_g)\iota_1(x) \) for any \( x \in \Gamma_f \).

In the following, we understand the fiber \((\nu_f)_{x_0}\) is given the consistent orientation in the sense that the orientation of \( T_{x_0}P \) determined by the ordered pair of oriented subspaces \((\nu_f)_{x_0}\), \( dfT_{x_0}M \) coincides with the fixed orientation, where \( \iota \) is the natural bundle monomorphism given by a choice of Riemannian metric on \( P \). Similarly with \((\nu_g)_{y_0}\).

Then we have the following product formula which unifies the equalities in Theorem A.

**Theorem B.** Assume \( M^m, N^n, P^{2m}, Q^{2n} \) are connected closed smooth manifolds and assume further that \( M, N \) are closed. Let \( f : M \to P, g : N \to Q \) be immersions. Let \( x_0, y_0 \) be the respective base points of \( M, N \) and assume further that \( f(x_0), w_0 = g(y_0) \), those of \( P, Q \). Assume \( T_{x_0}M, T_{z_0}P \) and \( T_{y_0}N, T_{w_0}Q \) are oriented and \( T_{x_0,y_0}M \times N, T_{z_0,w_0}P \times Q \) are given the product orientations. Then, we have

\[
I_{\Gamma}(f \times g) = (-1)^{mn}(I_{\Gamma}(f) * I_{\Gamma}(g) + \chi(\nu_g)\iota_1(I_{\Gamma}(f)) + \chi(\nu_f)\iota_2(I_{\Gamma}(g))) \in \Gamma_{f \times g},
\]

where \( \chi(\cdot) \) denotes the Euler characteristic in the twisted coefficients.

**Proof.** We retain the notations and contexts of the statement of 4.2 and its proof. Note that

\[
I_{\Gamma}(f) = \sum_i \varepsilon_{r_i}[\gamma_{r_i}] \in \Gamma_f, \quad I_{\Gamma}(g) = \sum_j \varepsilon_{s_j}[\gamma_{s_j}] \in \Gamma_g
\]

where each of \( \gamma_{r_i} \in \pi_1(P) \), \( \gamma_{s_j} \in \pi_1(Q) \) and \( \varepsilon_{r_i}, \varepsilon_{s_j} \), are determined by a choice of an ordered pair of paths in \( M \) or in \( N \) for each of the double points.

Note that we have as the double points of \( \Lambda \), for each \( i, j \),

\[
x_{i,j} = (f_1(p_i, 0), g_1(q_j, 0)) = (f(p_i, 0), g(q_j, 0)),
\]

\[
x_{i,j}' = (f(p_i, 0), g(q_j, 0)) = (f'(p_i, 0), g'(q_j, 0)),
\]

and also for each \( j, k \) and for each \( i, l \),

\[
y_{k,j} = (f(a_k), g_1(q_j, 1)) = (f(a_k), g(q_j, 1)),
\]

\[
z_{i,l} = (f_1(p_i, 1), g(b_l)) = (f_1(p_i, 1), g(b_l)).
\]

By definition we have

\[
I_{\Gamma}(\Lambda) = \sum_{i,j} \varepsilon_{x_{i,j}}[\gamma_{x_{i,j}}] + \varepsilon_{x'_{i,j}}[\gamma_{x'_{i,j}}] + \sum_{k,j} \varepsilon_{y_{k,j}}[\gamma_{y_{k,j}}] + \sum_{i,l} \varepsilon_{z_{i,l}}[\gamma_{z_{i,l}}]
\]

where the expression in the right hand side depends essentially on a choice of an ordered pair of paths in \( M \times N \) for each of the double points of \( \Lambda \).
We may assume that \( \varphi(x_0) = 0, \psi(y_0) = 0 \). Then note that the homotopy \( \Lambda_u : M \times N \rightarrow P \times Q, u \in I \), which is defined by \( \Lambda_u(x,y) = (f_u \varphi(y)(x), g_u \psi(x)(y)) \), preserves the base point. The following completes the proof of the theorem.

**Claim.** We may arrange so that

\[
\gamma_{x_{i,j}} = \gamma_{r_i} \otimes \gamma_{s_j}, \quad \gamma_{x'_{i,j}} = \gamma_{r_i} \otimes \gamma_{s_j}^{-1},
\]

and

\[
\varepsilon_{x_{i,j}} = (-1)^{mn+1} \varepsilon_{r_i} \varepsilon_{s_j}, \quad \varepsilon_{x'_{i,j}} = (-1)^{mn} \varepsilon_{r_i} \varepsilon_{s_j},
\]

\[
\varepsilon_{y_{i,j}} = (-1)^{mn} \varepsilon(a_k) \varepsilon_{s_j}, \quad \varepsilon_{z_{i,j}} = (-1)^{mn} \varepsilon(b_l) \varepsilon_{r_i}.
\]

**Proof.** For any paths, \( \alpha : I \rightarrow M, \beta : I \rightarrow N \), write \( \alpha \otimes \beta : I \rightarrow M \times N \) to denote the path defined by \( \alpha \otimes \beta(t) = (\alpha(t), \beta(t)), t \in I \).

Let \( \alpha, \alpha' : I \rightarrow M \) be the paths such that \( \alpha(0) = \alpha'(0) = x_0 \) and \( \alpha(1) = p_{i,0}, \alpha'(1) = p'_{i,0} \) and the loop \( a = (f \alpha) \cdot (f \alpha')^{-1} \) represents the class \( \gamma_{r_i} \in \pi_1(P) \). Similarly let \( \beta, \beta' : I \rightarrow N \) be the paths such that the loop \( b = (g \beta) \cdot (g \beta')^{-1} \) represents the class \( \gamma_{s_j} \in \pi_1(Q) \).

First we consider the double point

\[
x_{i,j} = (f_0(p_{i,0}), g_0(q_{j,0})) = (f_0(p'_{i,1}), g_0(q'_{j,1})).
\]

Let \( \delta'_u : I \rightarrow M, u \in I \), be defined by \( \delta'_u(t) = p'_{i,u} \) and \( \varepsilon'_u : I \rightarrow N, u \in I \), by \( \varepsilon'_u(t) = q'_{j,u} \).

Then let \( \gamma_{x_{i,j}} \in \pi_1(P \times Q) \) be the class represented by the loop \( (\Lambda(\alpha \otimes \beta)) \cdot (\Lambda((\alpha \otimes \delta'_u) \otimes (\beta \cdot \varepsilon'_u)))^{-1} \).

Note that

\[
\Lambda_u(p_{i,0}, q_{j,0}) = (f_u(p_{i,0}), g_u(q_{j,0})) = (f_0(p'_{i,u}), g_0(q'_{j,u})) = \Lambda_u(p'_{i,u}, q'_{j,u}),
\]

for any \( u \in I \). Therefore, \( (\Lambda_u(\alpha \otimes \beta)) \cdot (\Lambda_u((\alpha \otimes \delta'_u) \otimes (\beta \cdot \varepsilon'_u))^{-1} \) is well-defined for any \( u \in I \). This gives a loop homotopy from \( (f \alpha \otimes g \beta) \cdot (f \alpha' \otimes g \beta')^{-1} = ((f \alpha) \cdot (f \alpha')^{-1}) \otimes ((g \beta) \cdot (g \beta')^{-1}) \) to the loop which represents \( \gamma_{x_{i,j}} \), which proves that \( \gamma_{x_{i,j}} \) can be chosen as \( \gamma_{r_i} \otimes \gamma_{s_j} \).

To prove that \( \gamma_{x'_{i,j}} \) can be chosen as \( \gamma_{r_i} \otimes \gamma_{s_j}^{-1} \), we introduce the paths \( \delta_u : I \rightarrow M, \delta_u(t) = p_{i,u} \) and \( \varepsilon_u : I \rightarrow N, \varepsilon_u(t) = q_{j,u} \), for each \( u \in I \). Then consider the homotopy \( (\Lambda_u((\alpha \cdot \delta_u) \otimes (\beta')) \cdot (\Lambda_u((\alpha' \otimes \beta) \cdot \varepsilon_u))^{-1} \), which provides a loop homotopy from \( ((f \alpha) \cdot (f \alpha')^{-1}) \otimes ((g \beta) \cdot (g \beta')^{-1}) \) to a loop which we let represent \( \gamma_{x'_{i,j}} \in \pi_1(P \times Q) \).

To prove that \( \gamma_{y_{i,j}} \) can be chosen as \( e \otimes \gamma_{s_j} \), we choose a path \( \tilde{\alpha} : I \rightarrow M \) such that \( \tilde{\alpha}(0) = x_0, \tilde{\alpha}(1) = a_k \) and consider the homotopy \( (\Lambda_u(\tilde{\alpha} \otimes (\beta \cdot \delta_u))) \cdot (\Lambda_u((\tilde{\alpha} \otimes (\beta') \cdot \varepsilon_u))^{-1}, u \in I. \) This gives a homotopy from \( ((f \tilde{\alpha}) \cdot (f \tilde{\alpha}')^{-1}) \otimes ((g \beta) \cdot (g \beta')^{-1}) \) to a loop which we let represent the class \( \gamma_{y_{i,j}} \). Similarly to prove that \( \gamma_{z_{i,j}} \) can be chosen as \( \gamma_{r_i} \otimes e \).
It remains to prove the equalities concerning the signs of the double points with respect to the above choices of ordered pairs of paths in \( M \times N \) for the double pairs of \( \Lambda \).

First consider \( \varepsilon_{x_{i,j}} \). By definition it is determined as follows: Note that \( T_{x_{i,j}} P \times Q \) is oriented by the orientation on \( T_{(x_0,y_0)} P \times Q \) and the path \( \Lambda(\alpha \otimes \beta) \). Also \( T_{(p_0,q_0,0)} M \times N \), \( T_{(p_0',q_0',1)} M \times N \) are oriented by the orientation of \( T_{(x_0,y_0)} M \times N \) and the paths \( \alpha \otimes \beta \), \( (\alpha \cdot \delta'_1) \otimes (\beta \cdot \epsilon'_1) \). Then we compare the orientation of \( T_{x_{i,j}} P \times Q \) with the orientation given by the ordered pair of subspaces \( d\mathcal{A} T_{(p_0,q_0,0)} M \times N \), \( d\mathcal{A} T_{(p_0',q_0',1)} M \times N \).

Note that, for each \( u \in I \), in particular including \( u = 0 \), the paths \( \Lambda_u(\alpha \otimes \beta) \), \( \Lambda_u((\alpha' \cdot \delta'_u) \otimes (\beta' \cdot \epsilon'_u)) \), also determines a sign for \( \Lambda_u(p_0,0,q_0,0) \). Since this assignment of signs is continuous, the sign \( \varepsilon_{x_{i,j}} \) is the same as the one given by the following data: i) The orientation of \( T_{(f(p_0,q_0),g(q_0,0))} P \times Q \) determined by the path \( f \alpha \otimes g \beta \) and ii) the orientations of \( T_{(p_0,q_0,0)} M \times N \), \( T_{(p_0',q_0',1)} M \times N \) whose orientations are respectively determined by the paths \( \alpha \otimes \beta \), \( \alpha' \otimes \beta' \). From this it is immediate that \( \varepsilon_{x_{i,j}} = (-1)^{mn} \varepsilon_{r_i} \varepsilon_{s_j} \).

Similar considerations establish the other claimed equalities. In particular, the multiplication by \( w(\gamma_{s_j}) \) in the equality \( \varepsilon'_{x_{i,j}} = (-1)^{mn+n} w(\gamma_{s_j}) \varepsilon_{r_i} \varepsilon_{s_j} \) is due to the fact that we use the path \( f \alpha \otimes f \beta' \) to orient \( T_{(r_i,s_j)} P \times Q \) instead of the path \( f \alpha \otimes f \beta \). \( \square \)

**Remark.** Theorem A is a corollary of Theorem B only modulo Lemma 4.1.

---

**References**


Received March 1, 1999. This work was supported by grant no. 2000-2-10100-002-3 from the Basic Research Program of the Korea Science & Engineering Foundation.

DEPARTMENT OF MATHEMATICS
HANYANG UNIVERSITY
SUNGDONG-gu SEOUL 133-791
KOREA
E-mail address: yhbyun@fermat.hanyang.ac.kr

LIBERAL ARTS AND SCIENCE (MATHEMATICS)
YOUNGDONG UNIVERSITY
YOUNGDONG, CHUNGBUK, 370-701
KOREA
E-mail address: seunghun@youngdong.ac.kr
STABLE RANK AND REAL RANK OF GRAPH C*-ALGEBRAS

J.A. Jeong, G.H. Park, and D.Y. Shin

Dedicated to Professor Sa Ge Lee on his 60th birthday

For a row finite directed graph \( E \), Kumjian, Pask, and Raeburn proved that there exists a universal \( C^* \)-algebra \( C^*(E) \) generated by a Cuntz-Krieger \( E \)-family. In this paper we consider two density problems of invertible elements in graph \( C^* \)-algebras \( C^*(E) \), and it is proved that \( C^*(E) \) has stable rank one, that is, the set of all invertible elements is dense in \( C^*(E) \) (or in its unitization when \( C^*(E) \) is nonunital) if and only if no loop of \( E \) has an exit. We also prove that for a locally finite directed graph \( E \) with no sinks if the graph \( C^* \)-algebra \( C^*(E) \) has real rank zero \( RR(C^*(E)) = 0 \), that is, the set of invertible self-adjoint elements is dense in the set of all self-adjoint elements of \( C^*(E) \) then \( E \) satisfies a condition \((K)\) on loop structure of a graph, and that the converse is also true for \( C^*(E) \) with finitely many ideals. In particular, for a Cuntz-Krieger algebra \( O_A \), \( RR(O_A) = 0 \) if and only if \( A \) satisfies Cuntz’s condition \((II)\).

1. Introduction.

Given an \( n \times n \) \( \{0,1\} \)-matrix \( A \) with no zero row or column, a family of \( n \) partial isometries \( S_i \) satisfying the relation

\[
S_i^* S_i = \sum_{j=1}^{n} A(i,j) S_j S_j^*
\]

is called a Cuntz-Krieger \( A \)-family. In [CK], under a condition \((I)\) on the matrix \( A \), it is proved that any two such families generate isomorphic \( C^* \)-algebras, thus the Cuntz-Krieger algebra \( O_A \) is well-defined. Furthermore when \( A \) satisfies condition \((II)\) which is stronger than \((I)\) the ideal structure of \( O_A \) was analysed by Cuntz in [C].

As a generalization of Cuntz-Krieger algebras one may consider a \( C^* \)-algebra generated by a family of partial isometries satisfying the relation \((*)\) for some infinite \( \{0,1\} \)-matrix \( A \), provided every row of \( A \) contains only finitely many 1’s, and this has been done in [KPRR] and [KPR] with
directed graphs. For any row finite directed graph $E$ with countable vertices $\{v \mid v \in E^0\}$ and edges $\{e \mid e \in E^1\}$, the associated graph $C^*$-algebra $C^*(E)$ is defined to be a universal $C^*$-algebra generated by a family of partial isometries $\{s_e \mid e \in E^1\}$ and a family of mutually orthogonal projections $\{p_v \mid v \in E^0\}$ subject to the relations:

$$s_e^* s_e = p_{r(e)}, \quad p_v = \sum_{s(f) = v} s_f s_f^*,$$

where $r(e)$ (respectively, $s(e)$) denotes the range (respectively, source) vertex of the edge $e$. If $\{A(e, f)\}$ is the edge matrix of $E$ then these relations give a generalized form of $(*)$, that is, $s_e^* s_e = \sum_{s(f) = r(e)} A(e, f)s_f s_f^*$.

If $E$ has no sinks then there is a locally compact $r$-discrete groupoid $G_E$ associated with $E$ whose unit space $G_E^0$ is identified with the infinite path space of $E$. Furthermore it is shown in [KPRR], Theorem 4.2 that the groupoid $C^*$-algebra $C^*(G_E)$ is isomorphic to $C^*(E)$, and hence those useful results on groupoid $C^*$-algebras in [Rn1] and [Rn2] could be used to analyse the structure of $C^*(E)$. One important theorem in [KPRR] is about the ideal structure of graph $C^*$-algebras; there is an inclusion preserving one-to-one map of saturated hereditary vertex subsets of $E$ into the ideals of $C^*(E)$ and moreover if $E$ satisfies a condition (K) then the map is also bijective.

A graph-theoretic condition (L) analogous to Cuntz-Krieger’s condition (I) was given in [KPR], where it was shown that if $E$ is a locally finite directed graph with no sinks and satisfies (L) then a $C^*$-algebra generated by a Cuntz-Krieger $E$-family of non-zero elements is isomorphic to $C^*(E)$. One interesting result among others in [KPR] is that $C^*(E)$ is AF if and only if $E$ has no loops. It is also shown in [D] that every AF-algebra arises as the $C^*$-algebra of a locally finite pointed directed graph in the sense of [KPRR]. Recall that every AF algebra $A$ has stable rank one ($sr(A) = 1$); the set of invertible elements is dense in $A$ (or $\tilde{A}$ if $A$ is nonunital). In Section 3, we give a necessary and sufficient graph-theoretic condition on $E$ for the graph algebra $C^*(E)$ to have stable rank one; $sr(C^*(E)) = 1$ if and only if no loop of $E$ has an exit.

We see from [KPR] that if $E$ is a cofinal graph with no sinks and satisfies (L) then the universal $C^*$-algebra $C^*(E)$ is simple and it is either AF or purely infinite. It is also well-known that all AF algebras and purely infinite simple $C^*$-algebras have real rank zero, that is, every self-adjoint element can be arbitrarily closely approximated by invertible self-adjoint elements (or in the unitized algebra for a nonunital $C^*$-algebra). So it would be interesting to know when a non-simple graph $C^*$-algebra can have real rank zero, and we prove in Section 4 that for a locally finite directed graph $E$ with no sinks if the graph algebra $C^*(E)$ has real rank zero ($RR(C^*(E)) = 0$) then the graph must satisfy condition (K). Conversely we also show that
for any locally finite graph $E$ with no sinks if $E$ satisfies condition (K) and $C^*(E)$ has finitely many ideals then $RR(C^*(E)) = 0$. In particular, if $E$ is a locally finite graph with no sinks and has finitely many vertices then $RR(C^*(E)) = 0$ if and only if $E$ satisfies condition (K). Therefore, for a Cuntz-Krieger algebra $O_A$ associated with a $\{0,1\}$-matrix $A$ satisfying (I), $RR(O_A) = 0$ if and only if $A$ satisfies condition (II) since $A$ can be viewed as a vertex matrix of a finite graph $E$ which has no sinks and satisfies (L) and that the finite graph $E$ satisfies condition (K) is equivalent to that its vertex matrix $A$ satisfies condition (II).

2. Preliminaries.

We recall some definitions and notations from [KPR] and [KPRR] on directed graphs, graph $C^*$-algebras, and groupoids associated with graphs. A directed graph $E = (E^0, E^1, r, s)$ consists of countable sets $E^0$ of vertices and $E^1$ of edges, and the range, source maps $r, s : E^1 \to E^0$. $E$ is row finite (locally finite) if for each vertex $v \in E^0$, $s^{-1}(v)$ is (both $r^{-1}(v)$ and $s^{-1}(v)$ are) finite. We call a locally finite graph $E$ finite if $E^0$ is finite. If $e_1, \ldots, e_n$ ($n \geq 2$) are edges with $r(e_i) = s(e_{i+1})$, $1 \leq i \leq n - 1$, then we can form a (finite) path $\alpha = (e_1, \ldots, e_n)$ of length $|\alpha| = n$, and extend the maps $r, s$ by $r(\alpha) = r(e_n), s(\alpha) = s(e_1)$.

Let $E^n$ be the set of all finite paths of length $n$ and

$$E^n := \bigcup_{n \geq 0} E^n, \quad r(v) = s(v) = v \text{ for } v \in E^0,$$

$$E^\infty := \{(\alpha_i)_{i=1}^\infty | \alpha_i \in E^1, r(\alpha_i) = s(\alpha_{i+1})\}.$$

A vertex $v \in E^0$ with $s^{-1}(v) = \emptyset$ is called a sink.

Given a row finite directed graph $E$, a Cuntz-Krieger $E$-family consists of a set $\{P_v | v \in E^0\}$ of mutually orthogonal projections and a set $\{S_e | e \in E^1\}$ of partial isometries satisfying the relations

$$S_e^* S_e = P_{r(e)}, \quad e \in E^1, \text{ and } P_v = \sum_{s(e) = v} S_e S_e^*, \quad v \in s(E^1).$$

From these relations, one can show that every non-zero word in $S_e, P_v$ and $S_f^*$ is a partial isometry of the form $S_\alpha S_\beta^*$ for some $\alpha, \beta \in E^*$ with $r(\alpha) = r(\beta)$ ([KPR], Lemma 1.1).

**Theorem 2.1** ([KPR, Theorem 1.2]). For a row finite directed graph $E = (E^0, E^1)$, there exists a $C^*$-algebra $C^*(E)$ generated by a Cuntz-Krieger $E$-family $\{s_e, p_v | v \in E^0, e \in E^1\}$ of non-zero elements such that for any Cuntz-Krieger $E$-family $\{S_e, P_v | v \in E^0, e \in E^1\}$ of partial isometries acting on a Hilbert space $\mathcal{H}$, there is a representation $\pi : C^*(E) \to B(\mathcal{H})$ such that

$$\pi(s_e) = S_e, \quad \pi(p_v) = P_v$$

for all $e \in E^1, v \in E^0$. 
A finite path \( \alpha \) with \(|\alpha| > 0 \) is called a loop at \( v \) if \( s(\alpha) = r(\alpha) = v \). If the vertices \( \{r(\alpha_i)\} 1 \leq i \leq |\alpha| \) are distinct, the loop \( \alpha \) is simple.

\( E \) is said to satisfy a condition (L) if every loop in \( E \) has an exit, and a condition (K) if for any vertex \( v \) on a loop there exist at least two distinct loops \( \alpha, \beta \) based at \( v \), that is, \( r(\alpha) = r(\beta) = s(\alpha) = s(\beta) = v \), \( r(\alpha_i) \neq v \) for \( 1 \leq i < |\alpha| \), and \( r(\beta_j) \neq v \) for \( 1 \leq j < |\beta| \). Note that the condition (K) is stronger than (L) and if \( E \) has no loops then the two conditions are trivially satisfied.

If \( E \) has no sinks then \( E^\infty \neq \emptyset \) and we have the following groupoid associated with \( E \)

\[
G_E = \{(x, k, y) \in E^\infty \times \mathbb{Z} \times E^\infty \mid x_i = y_{i+k} \text{ for sufficiently large } i \}
\]

\[
(x, k, y)^{-1} := (y, -k, x),
\]

\[
(x, k, y) \cdot (y, l, z) := (x, k + l, z).
\]

Then the range and source maps \( r, s : G_E \to G_E^0 \) are given by

\[
r(x, k, y) = x, \ s(x, k, y) = y.
\]

\( G_E \) is a locally compact \( r \)-discrete groupoid with respect to a suitable topology and \( G_E^0 \) is identified with \( E^\infty \). Furthermore the groupoid algebra \( C^*(G_E) \) is isomorphic to the graph \( C^* \)-algebra \( C^*(E) \) by Theorem 4.2 of [KPR].

### 3. Stable rank of \( C^*(E) \).

Recall that a \( C^* \)-algebra \( A \) has stable rank one \( (sr(A) = 1) \) if the set \( A^{-1} \) of all invertible elements is dense in \( A \) (in \( \tilde{A} \) if \( A \) is non-unital). One can show that every \( C^* \)-algebra \( A \) with \( sr(A) = 1 \) is stably finite, and so there is no infinite projection in \( A \). If two \( C^* \)-algebras \( A \) and \( B \) are strong Morita equivalent, in particular if they are stably isomorphic, then \( sr(A) = 1 \) if and only if \( sr(B) = 1 \) ([BP2], [Rf]).

**Lemma 3.1** ([BP2, Proposition 6.4]). Let \( I \) be an ideal of a \( C^* \)-algebra \( A \). Then \( sr(A) = 1 \) if and only if \( sr(I) = sr(A/I) = 1 \) and every invertible element lifts (that is, \( (A/I)^{-1} = A^{-1}/I \)).

We say that a subgraph \( H \) of \( E \) has no exit if \( e \in E^1 \), \( s(e) \in H^0 \) implies \( e \in H^1 \).

**Lemma 3.2** ([KPR, Proposition 2.1]). If \( H \) is a subgraph of a directed graph \( E \) with no exit then

\[
I := \operatorname{span}\{s_\alpha s_\beta^* \mid \alpha, \beta \in E^*, \ r(\alpha) = r(\beta) \in H^0 \}
\]

is a closed ideal of \( C^*(E) \) strong Morita equivalent to the hereditary \( C^* \)-subalgebra \( B := \operatorname{span}\{s_\alpha s_\beta^* \mid \alpha, \beta \in H^* \} \).
We call a vertex \( v \) cofinal if for any infinite path \( x = (x_1, x_2, \ldots) \in E^\infty \) there is a finite path \( \alpha \in E^* \) with \( s(\alpha) = v \) and \( r(\alpha) = s(x_n) \) for some \( n \) ([KPRR]). A directed graph \( E \) is said to be cofinal if every vertex is cofinal.

**Theorem 3.3.** Let \( E = (E^0, E^1, r, s) \) be a row finite directed graph. Then \( E \) has no loop with an exit if and only if \( \text{sr}(C^*(E)) = 1 \).

**Proof.** If \( E \) has no loops then \( C^*(E) \) is AF and so \( \text{sr}(C^*(E)) = 1 \). Assume that \( E \) has loops and every loop has no exit. Let \( H \) be the subgraph of \( E \) consisting of all the loops. Since \( H \) has no exit, by Lemma 3.2,

\[
I = \text{span}\{s_\beta s_\gamma^* | \beta, \gamma \in E^*, \ r(\beta) = r(\gamma) \in H^0\}
\]

is an ideal of \( C^*(E) \) which is strong Morita equivalent to the hereditary subalgebra \( B = \text{span}\{s_\beta s_\gamma^* | \beta, \gamma \in H^*\} \). Let \( \alpha \) be a simple loop in \( E \), then \( v = s(\alpha) \) is cofinal in the subgraph \( H_\alpha \), consisting only of \( \alpha \), and \( H_\alpha \) has no sinks. Thus \( C^*(H_\alpha) \cong C^*(G_{H_\alpha}) \) ([KPRR], Theorem 4.2). Let \( N = \{x \in H_\alpha^\infty \mid s(x) = v\} \), and \( G_{H_\alpha}^N \) be the reduction of \( G_{H_\alpha} \) to \( N \). Then by [KPRR], Theorem 3.1, \( C^*(G_{H_\alpha}^N) \) is isomorphic to the full corner of \( C^*(G_{H_\alpha}) \), so they are strong Morita equivalent. Since \( N \) consists of only one path, say \( x \), and \( G_{H_\alpha}^N = \{(x, kn, x) | k \in \mathbb{Z}\} \cong \mathbb{Z} \), \( C^*(H_\alpha) \) is strong Morita equivalent to the group \( C^*(\mathbb{T}) \cong C^*(\mathbb{T}) \). Since \( C^*(\mathbb{T}) \) has stable rank 1, it follows that \( sr(C^*(H_\alpha)) = 1 \), and so \( sr(B_\alpha) = 1 \), where \( B_\alpha := \text{span}\{s_\beta s_\gamma^* | \beta, \gamma \in H_\alpha^*\} \), because \( B_\alpha \) is a quotient algebra of \( C^*(H_\alpha) \). Thus \( sr(I_\alpha) = sr(B_\alpha) = 1 \), where

\[
I_\alpha := \text{span}\{s_\beta s_\gamma^* | \beta, \gamma \in E^*, \ r(\beta) = r(\gamma) \in H_\alpha^0\}.
\]

Therefore \( sr(I) = 1 \) since \( I \) is the direct sum of the ideals \( I_\alpha \).

Now, let \( D \) be the \( C^* \)-subalgebra of \( C^*(E) \) generated by

\[
\{s_e \mid e \in E^1 \setminus H^1\} \cup \{p_v \mid v \in E^0\},
\]

which is a Cuntz-Krieger \( G \)-family for the subgraph \( G = (E^0, E^1 \setminus H^1) \) of \( E \). Thus by Theorem 2.1 there is a \( * \)-homomorphism from \( C^*(G) \) onto \( D \). Since \( G \) has no loops at all, \( C^*(G) \) is an AF algebra having stable rank one, so we have \( sr(D) = 1 \) by Lemma 3.1.

It is clear that under the canonical projection \( \pi : C^*(E) \to C^*(E)/I \) the subalgebra \( D \) of \( C^*(E) \) maps onto \( C^*(E)/I \) and hence the stable rank of \( C^*(E)/I \) is one as a homomorphic image of an algebra of stable rank one. Also, every invertible element in the AF algebra \( \pi(D)(= C^*(\mathbb{E})/I) \) is connected to the unit, whence it lifts to an invertible element in \( C^*(E) \). Then by Lemma 3.1, \( sr(C^*(E)) = 1 \).

Conversely, suppose that \( E \) has a simple loop \( \alpha = (\alpha_1, \ldots, \alpha_n) \) with an exit at \( v = s(\alpha) \). It is easy to see that the projection \( p_v \) is infinite, so the algebra \( C^*(E) \) is not stably finite, whence \( sr(C^*(E)) \neq 1 \).
Lemma 3.4. If $V$ is the set of all sinks in $E$ then
\[ I := \text{span} \{ s_\alpha s_\beta^* | \alpha, \beta \in E^*, r(\alpha) = r(\beta) = v \text{ for some } v \in V \} \]
is a closed two-sided ideal of $C^*(E)$. With $E^*(v) = \{ \alpha \in E^* | r(\alpha) = v \}$, we have
\[ I \cong \bigoplus_{v \in V} K(\ell^2(E^*(v))). \]

Proof. For each $v \in V$, let
\[ I_v := \text{span} \{ s_\alpha s_\beta^* | \alpha, \beta \in E^*, r(\alpha) = r(\beta) = v \}. \]
Then by Corollary 2.2 of [KPR], $I_v$ is a closed ideal of $C^*(E)$ and isomorphic to $K(\ell^2(E^*(v)))$. If $\beta, \gamma \in E^*$, with $r(\beta) = v_1, r(\gamma) = v_2$, then $s_\beta^*s_\gamma = 0$ when $i \neq j$, whence the ideals are mutually orthogonal.

If a (locally finite) directed graph $E$ has sinks then it might not contain any infinite paths so that we can not directly apply results on groupoid $C^*$-algebras since the groupoid $\mathcal{G}_E$ associated with $E$ was invented to have its unit space consisting of infinite paths in $E$. In case $E$ has no sinks, in [KPRR], an isomorphism of lattice of saturated hereditary subsets $V$ of $E^0$ into the lattice of ideals $I(V)$ in $C^*(E)(\cong C^*(\mathcal{G}_E))$ was established and it is shown that the quotient algebra $C^*(E)/I(V)$ is isomorphic to the graph algebra $C^*(G)$ for a certain subgraph $G$ of $E$. The proof applies the results on ideal structure of groupoid algebras obtained in [Rn1, Rn2]. See Section 4 for this isomorphism. In the following we show a similar assertion when $V$ is the set of all sinks in $E$. For this, we need to recall that a vertex subset $H$ of $E^0$ is saturated if whenever $v \in E^0$ emits only edges $e$ with $r(e) \in H$, we have $v \in H$. The smallest saturated vertex subset containing $V$ is called the saturation of $V$.

Theorem 3.5. Let $E = (E^0, E^1, r, s)$ be a locally finite directed graph with the set $V$ of sinks. Then there is a subgraph $G = (E^0 \setminus H, \{ e \in E^1 | r(e) \notin H \})$ of $E$ with no sinks such that $C^*(E)/I(V)$ is isomorphic to $C^*(G)$, where $H$ is the saturation of $V$ and $I(V) = \text{span} \{ s_\alpha s_\beta^* | \alpha, \beta \in E^*, r(\alpha) = r(\beta) \in V \}$.

Proof. Note that the ideal $I(= I(V))$ contains the projections $p_v$, for $v \in V$.
If $e \in E^1, r(e) = v$ for some $v \in V$ then $s_e \in I$ because $s_e = s_e s_e^* s_e = s_e p_v \in I$. For an edge $e \in E^1$ with $r(e) \notin V$ we have
\[ s_e = s_e p_{r(e)} = \sum_{s(f) = r(e)} s_e s_f s_f^* p_{r(e)} \in I \]
whenever the vertex $r(e)$ emits only edges $f$ with $s_f \in I$. If $r(e)$ emits an edge $f$ with $s_f \notin I$ then $s_f s_f^* \notin I$ ($s_f = s_f s_f^* s_f$). From $s_e s_e = p_{r(e)} \geq s_f s_f^*$, we see that $s_e s_e \notin I$, so $s_e \notin I$. Thus
\[ s_e \in I \iff \text{ either } r(e) \in V \text{ or } r(e) \text{ emits only edges } f \text{ with } s_f \in I. \]
Now let \( \pi : C^*(E) \to C^*(E)/I \) be the canonical surjective homomorphism. Then \( \pi(C^*(E)) \) is generated by \( \pi(s_f), \ s_f \notin I \). Let \( G \) be the subgraph of \( E \) obtained from \( E \) by deleting the vertices \( w \) with \( p_w \in I \) and edges \( f \) with \( s_f \in I \), that is,

\[
(\ast\ast) \quad w \in G^0 \iff p_w \notin I, \ e \in G^1 \iff s_e \notin I.
\]

Then \( \pi(C^*(E)) \) is generated by \( \pi(s_f), \ f \in G^1 \). Let \( w \in G^0 \). Then \( w \notin V \) and hence \( w \) emits edges \( e_1, \ldots, e_m \) in \( E \). If \( w \) is a sink in \( G \) then \( s_{e_i} \in I, \ i = 1, \ldots, m \), and so \( p_w = \sum_i s_{e_i} s_{e_i}^* \in I \), a contradiction. Therefore the subgraph \( G \) has no sinks.

Let \( \pi(s_f) \neq 0 \), then \( f \) appears in \( G \) by \( (\ast\ast) \). If the vertex \( w = r(f) \) emits edges \( e_1, \ldots, e_k, \ldots, e_m \) in \( E \) such that \( s_{e_1}, \ldots, s_{e_k} \notin I \), and \( s_{e_{k+1}}, \ldots, s_{e_m} \in I \) then

\[
\pi(s_f^*) \pi(s_f) = \pi \left( \sum_{s(e) = r(f) = w} s_e s_e^* \right) = \sum_{i=1}^k \pi(s_{e_i}) \pi(s_{e_i})^* = \sum_{s(g) = r(f)} \pi(s_g) \pi(s_g)^*,
\]

which means that the partial isometries \( \{\pi(s_f) \mid f \in G^1\} \) is a Cuntz-Krieger \( G \)-family in \( \pi(C^*(E)) = C^*(E)/I \). Therefore there exists a homomorphism \( \phi : C^*(G) \to C^*(E)/I \) such that

\[
\phi(t_f) = \pi(s_f), \ f \in G^1 \text{ and } \phi(q_w) = \pi(p_w), \ w \in G^0,
\]

where \( \{t_f, q_w\} \) is a Cuntz-Krieger \( G \)-family generating \( C^*(G) \). On the other hand, one can form a Cuntz-Krieger \( E \)-family in \( C^*(G) \) by adding \( t_e = 0 \) for \( e \in E^1 \setminus G^1 \), and \( q_v = 0 \) for \( v \in E^0 \setminus G^0 \) to the family \( \{t_f, q_w\} \). Then we have a homomorphism \( \rho : C^*(E) \to C^*(G) \) such that

\[
\rho(s_e) = t_e, \ \rho(p_v) = q_v, \ e \in E^1, \ v \in E^0.
\]

Clearly, \( I \subset \text{Ker}(\rho) \). Now let \( x = \sum \lambda_{\alpha, \beta} s_{\alpha} s_{\beta}^* \in \text{Ker}(\rho) \). Then

\[
\pi \left( \sum \lambda_{\alpha, \beta} s_{\alpha} s_{\beta}^* \right) = \phi \left( \sum \lambda_{\alpha, \beta} t_{\alpha} t_{\beta}^* \right) = \phi \circ \rho(x) = 0.
\]

Thus \( x \in \text{Ker}(\pi) = I \). Therefore \( \text{Ker}(\rho) = I \) and the map \( \rho \) induces an isomorphism from \( C^*(E)/I \) onto \( C^*(G) \).

Recall that a \( C^* \)-algebra \( A \) is said to be purely infinite if every non-zero hereditary \( C^* \)-subalgebra of \( A \) has an infinite projection.

If an \( r \)-discrete groupoid \( G \) is essentially free and locally contracting then \( C^*(G) \) is purely infinite ([A], Proposition 2.4). From Lemma 3.4 of [KPR], we see that the groupoid \( G_E \) associated with a locally finite graph \( E \) with no sinks is essentially free if and only if \( E \) satisfies condition (L). It is also
known from the same paper that if every vertex connects to a loop with an exit then $G_E$ is locally contracting, so that $C^*(E)(\cong C^*(G_E))$ is purely infinite. Moreover there is a dichotomy for simple graph $C^*$-algebras.

**Proposition 3.6 ([KPR, Corollary 3.11]).** Let $E$ be a locally finite graph which has no sinks, is cofinal, and satisfies condition (L). Then $C^*(E)$ is simple, and

(i) if $E$ has no loops, then $C^*(E)$ is AF; 
(ii) if $E$ has a loop, then $C^*(E)$ is purely infinite.

**Proposition 3.7.** Let $E$ be a locally finite directed graph. If $E$ is cofinal then either $sr(C^*(G)) = 1$ or it is purely infinite simple.

**Proof.** If $E$ has no loop with an exit then $sr(C^*(E)) = 1$ by Theorem 3.3. Suppose $E$ has a loop with an exit. Since $E$ is cofinal, $E$ cannot have a sink. If $E$ has precisely one loop then $E$ satisfies (L) and so $C^*(E)$ is purely infinite simple by the previous proposition. Let $E$ have two distinct loops, $\alpha$, $\beta$. If $\gamma$ is a loop of $E$ then consider the infinite path $x = \alpha \alpha \cdots \alpha = (x_1, x_2, \ldots)$ assuming $\gamma \neq \alpha$. Since $E$ is cofinal the vertex $v = s(\gamma)$ connects to $x$ by a finite path, and this shows that the loop $\gamma$ has an exit. Therefore $E$ satisfies (L) and $C^*(E)$ is purely infinite simple by Proposition 3.6.

From the proof of the above proposition, we see that for a cofinal graph $E$ with no sinks $C^*(E)$ is simple unless $E$ has precisely one loop and the loop has no exit.

### 4. Real rank of $C^*(E)$.

Recall that a unital $C^*$-algebra $A$ is said to have real rank zero ($RR(A) = 0$) if every self-adjoint element can be arbitrarily closely approximated by invertible self-adjoint elements, that is, $A_{sa}^{-1}$ is dense in $A_{sa}$. For a nonunital $C^*$-algebra $A$, we say that $A$ has real rank zero if $\tilde{A}$ has real rank zero ([BP1]). Then $RR(A) = 0$ if and only if $RR(A \otimes K) = 0$. Also it is well-known that $RR(A) = 0$ is equivalent to that $A$ satisfies a condition (FS), that is, the set of self-adjoint elements with finite spectra is dense in $A_{sa}$, so $RR(A) = 0$ implies that $A$ contains fairly many projections so that the linear span of its projections is dense in $A$. Graph $C^*$-algebras $C^*(E)$ are basically generated by their partial isometries, and thus they would have plenty of projections and one might expect that most of them have real rank zero. In fact, if $C^*(E)$ is simple then it is either AF or purely infinite simple and in both cases it is well-known that these algebras have real rank zero; for real rank of a purely infinite simple $C^*$-algebra, see [Z].

In this section, we first find a necessary condition for a graph $C^*$-algebra $C^*(E)$ to have real rank zero. We need to review the ideal theory of a graph $C^*$-algebra $C^*(E)$ for a directed graph $E$ with no sinks. Recall that $C^*(E)$
can be identified with its infinite path space groupoid model $C^*(\mathcal{G})$ and $C^*(\mathcal{G}) \cong C^*_r(\mathcal{G})$ since the groupoid associated with a locally finite directed graph $E$ is amenable ([KPRR], Corollary 5.3). A subset $H$ of the vertex set $E^0$ is hereditary if $v \in H$ and $w \in E^0$ with $s(\alpha) = v$, $r(\alpha) = w$ for some $\alpha \in E^*$ then $w \in H$.

For a hereditary and saturated vertex set $H \subset E^0$, let

$$U(H) = \{x \in E^\infty | r(x_n) \in H \text{ for some } n \}.$$  

Then $U(H)$ is an open invariant subset of $E^\infty$ (which is identified with the unit space $\mathcal{G}^0$ of the groupoid $\mathcal{G}$ associated with the graph $E$). The map $H \mapsto U(H)$ is an isomorphism between the lattices of saturated hereditary subsets of $E^0$ and open invariant subsets $\mathcal{O}(\mathcal{G})$ of $E^\infty$ ([KPRR], Lemma 6.5). On the other hand, for each open invariant subspace $U \subset E^\infty(= \mathcal{G}^0)$, the space

$$C_c(\mathcal{G}^U) := \{f \in C_c(\mathcal{G}) : \text{supp } f \subset \mathcal{G}^U \}$$

is an ideal of $C_c(\mathcal{G})$, hence its closure is an ideal $I(U)$ of $C^*(\mathcal{G})$. We see from [Rn1], Proposition 4.5 that the correspondence $U \mapsto I(U)$ is a one-to-one order preserving map between $\mathcal{O}(\mathcal{G})$ and the lattice of ideals $\mathcal{J}(C^*(\mathcal{G}))$ of $C^*(\mathcal{G})$. Thus $H \mapsto I(U(H))$ is an order preserving isomorphism from the lattice of hereditary saturated vertex subsets into $\mathcal{J}(C^*(\mathcal{G}))$. It is proved in the proof of [KPRR], Theorem 6.6 that the ideals $I(H)$ and $I(U(H))$ coincide, where

$$I(H) := \overline{\text{span}} \{1_{Z(\alpha,\beta)} | \alpha, \beta \in E^*, r(\alpha) = r(\beta) \in H \},$$

and $1_{Z(\alpha,\beta)}$ is the characteristic function on the compact open subset $Z(\alpha, \beta)$ of the groupoid $\mathcal{G}$.

The isomorphism from $C^*(\mathcal{G})$ onto $C^*(E)$ obtained in [KPRR] maps the functions $1_{Z(\alpha,\beta)} (\alpha, \beta \in E^*, \ r(\alpha) = r(\beta) \in H)$ onto $s_\alpha s_\beta^*$. Therefore we have

$$I(H) = \overline{\text{span}} \{s_\alpha s_\beta^* | \alpha, \beta \in E^*, r(\alpha) = r(\beta) \in H \}.$$  

Furthermore the following is known.

**Theorem 4.1** ([KPRR, Theorem 6.6], or [P, Theorem 2.2]). Let $E$ be a locally finite directed graph with no sinks. Then the map $H \mapsto I(H)$ described above is injective, and the quotient algebra $C^*(E)/I(H)$ is isomorphic to $C^*(F)$ of the directed graph $F := (E^0 \setminus H, \{e \mid r(e) \notin H \})$. The ideal $I(H)$ is strong Morita equivalent to $C^*(K)$ of the directed graph $K := (H, \{e \mid s(e) \in H \})$. Moreover, if $E$ satisfies the condition (K) then the map $H \mapsto I(H)$ is surjective.

**Theorem 4.2** ([BP1]). Let $A$ be a $C^*$-algebra and $I$ be an ideal of $A$.

(a) If $RR(A) = 0$ then $RR(I) = RR(A/I) = 0$.

Suppose $RR(I) = RR(A/I) = 0$. Then we have the following.
(b) $RR(A) = 0$ if and only if every projection in $A/I$ lifts to a projection in $A$. In particular if $K_1(I) = 0$ then every projection lifts.

(c) If $B$ is a $C^*$-subalgebra of $A$ with $RR(B) = 0$ and $A = B + I$ then $RR(A) = 0$.

Now, we can prove our first theorem on real rank of graph $C^*$-algebras.

**Theorem 4.3.** Let $E$ be a locally finite directed graph with no sinks. If $RR(C^*(E)) = 0$ then $E$ satisfies condition (K).

**Proof.** Suppose there is a simple loop $\alpha$ with no exit in $E$. Then the subgraph $H_\alpha$ consisting of $\alpha$ has no exit and generates an ideal $I$ stably isomorphic to $C(\mathbb{T})$, that is, $I \otimes K \cong C(\mathbb{T}) \otimes K$, as in the proof of Theorem 3.3. Since $RR(C(\mathbb{T})) \neq 0$ it follows that $RR(C^*(E)) \neq 0$ by Theorem 4.2(a), a contradiction, which shows that $E$ satisfies condition (L).

To prove condition (K), let $v$ be a vertex such that there is only one loop at $v$. Let $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ be the loop and let $V$ be the set of vertices $w \in V$ such that $w = r(e)$ for an exit $e$ of $\beta$ and $H$ be the smallest hereditary and saturated vertex set containing $V$. Then $V \neq \emptyset$ because $E$ satisfies (L). Moreover, $H$ is a proper subset of $E^0$ since vertices on the loop $\beta$ are not elements in $H$. Thus there exists a proper ideal $I(H)$ in $C^*(E)$, and the quotient algebra $C^*(E)/I(H)$ is isomorphic to $C^*(F)$ of the directed graph $F = (E^0 \setminus H, \{ e \mid r(e) \notin H \})$. Hence $F$ has a loop $\beta$ with no exit in $F$ and by the argument in the first paragraph of the proof $RR(C^*(F)) \neq 0$. Therefore $RR(C^*(E)) \neq 0$ by Theorem 4.2(a).

**Corollary 4.4.** Let $E$ be a locally finite directed graph with no sinks. If $sr(C^*(E)) = 1$ and $RR(C^*(E)) = 0$ then $C^*(E)$ is AF.

**Proof.** By Theorem 3.3 and Theorem 4.3, $E$ has no loops, and the assertion follows from Theorem 2.4 in [KPR].

**Proposition 4.5.** Let $E$ be a locally finite directed graph with no sinks. Then $C^*(E)$ is simple if and only if $E$ is cofinal and satisfies (K).

**Proof.** Suppose $E$ is cofinal and satisfies condition (K) then $C^*(E)$ is simple by the proof of [KPRR], Corollary 6.8.

Since the converse has not been proved there in the same proof, we provide one for reader’s convenience. To prove the converse, suppose $E$ is not cofinal. Then there exist an infinite path $x$ and a vertex $v$ which cannot connect to $x$ by a finite path. Let $H_1$ be the set of all vertices $v$ which can be connected from $v$, that is, there is a finite path $\alpha \in E^*$ with $s(\alpha) = v, r(\alpha) = w$. Then $H_1$ is the smallest hereditary vertex set containing $v$. Let $H$ be the set of all vertices $w$ satisfying that for any path $\alpha \in E^* \cup E^\infty$ with $s(\alpha) = w$, if $\alpha \in E^*$ then there is another path $\beta \in E^*$ such that $s(\beta) = r(\alpha)$ and $r(\beta) \in H_1$, if $\alpha \in E^\infty$ then $r(\alpha_j) \in H_1$ for some $j$. Then clearly $v \in H_1 \subset H$. We show that $H$ is a saturated hereditary vertex set which does not contain
vertices on the infinite path \( x \). Suppose a vertex \( w \) emits edges \( e_1, \ldots, e_n \) and \( r(e_i) \in H \) for all \( i \). If \( \alpha \) is a path with \( s(\alpha) = w \) then \( \alpha_1 = e_j \) for some \( j \) and \( \alpha = e_j \gamma \) for some path with \( s(\gamma) = r(e_j) \in H \). Since \( \gamma \) is a path with \( s(\gamma) = r(e_j) \in H \), if \( \gamma \in E^* \) then we can find a path \( \beta \in E^* \) such that \( s(\beta) = r(\gamma) \) and \( r(\beta) \in H_1 \). If \( \gamma \in E^\infty \) then \( r(\gamma_i) \in H_1 \) for some \( i \), and hence \( r(\alpha_{i+1}) \in H_1 \). Thus \( w \in H \), and \( H \) is saturated. Now let \( u \) be a vertex connected by a finite path \( \beta \) from some vertex \( w \in H \), that is, \( s(\beta) = w, r(\beta) = u \). Then for any path \( \alpha \) with \( s(\alpha) = u \), the path \( \beta \alpha \) starts from \( w \), and it is easy to see that \( u \in H \), and \( H \) is hereditary. Obviously the infinite path \( x \) does not meet any vertex in \( H_1 \), hence \( H \) is a proper saturated hereditary subset of \( E^0 \). Therefore \( C^*(E) \) is not simple by Theorem 4.1.

Now suppose \( E \) is cofinal but does not satisfy condition (K). Since for a cofinal graph two conditions (K) and (L) are equivalent, \( E \) has a loop with no exit. We have already seen from the proof of Theorem 3.3 that such a loop generates an ideal strong Morita equivalent to \( C(\mathbb{T}) \). Thus \( C^*(E) \) can not be simple.

We prove the converse of Theorem 4.3 when \( C^*(E) \) has finitely many ideals.

**Theorem 4.6.** Let \( E \) be a locally finite directed graph with no sinks which satisfies condition (K). If \( C^*(E) \) has only finitely many ideals then \( RR(C^*(E)) = 0 \). In particular, if \( E \) is a finite graph then \( RR(C^*(E)) = 0 \).

**Proof.** Let \( n \) be the number of non-zero ideals in \( C^*(E) \). We prove our assertion by induction on \( n \).

For \( n = 1 \), \( C^*(E) \) is simple and \( RR(C^*(E)) = 0 \) since \( C^*(E) \) is either AF or purely infinite simple.

Let \( n > 1 \). Let \( I(H) \) be a maximal ideal of \( C^*(E) \) for some hereditary saturated vertex subset \( H \) of \( E^0 \). By Theorem 4.1 and induction hypothesis, \( I(H) \) and the simple \( C^* \)-algebra \( C^*(E)/I(H) \) have real rank zero. We show that \( C^*(E) = I(H) + B \) for some \( C^* \)-subalgebra \( B \) isomorphic to \( C^*(\tilde{F}) \) for a directed subgraph \( \tilde{F} \) (possibly with sinks) of \( E \) such that \( RR(C^*(\tilde{F})) = 0 \) and then apply Theorem 4.2(c). According to Theorem 4.1, \( C^*(E)/I(H) \cong C^*(F) \), where \( F = (E^0 \setminus H, \{e \mid r(e) \notin H\}) \). Let

\[ V := \{v \in H \mid v = r(e) \text{ for some edge } e \in E^1 \text{ with } s(e) \in F^0 = E^0 \setminus H\}. \]

If \( V = \emptyset \), then \( C^*(E) \cong I(H) \oplus C^*(F) \), and therefore \( RR(C^*(E)) = 0 \) since two direct summands have real rank zero by induction hypothesis. If \( V \neq \emptyset \) we set

\[ \tilde{F} = (F^0 \cup V, F^1 \cup \{f \in E^1 \mid r(f) \in V, s(f) \in F^0\}). \]
Then $V$ is the set of all sinks of $\tilde{F}$. By Theorem 3.5, $C^*(\tilde{F})/I(V)$ is isomorphic to the simple $C^*$-algebra $C^*(F)$, where  

$$I(V) = \overline{\text{span}}\{s_\alpha s_\beta^* | \alpha, \beta \in \tilde{F}^*, r(\alpha) = r(\beta) \in V\}.$$  

Thus $RR(C^*(\tilde{F})/I(V)) = RR(C^*(F)) = 0$. The ideal  

$$I(V) \cong \bigoplus_{v \in V} K(\ell^2(E^*(v)))$$  

also has real rank zero. Furthermore since $K_1(I(V)) = 0$, by Theorem 4.2(b), $RR(C^*(\tilde{F})) = 0$. Let $B$ be the $C^*$-subalgebra of $C^*(E)$ generated by the family of nonzero elements $\{p_v, s_f | v \in (\tilde{F})^0, f \in (\tilde{F})^1\}$. Then this is a Cuntz-Krieger $\tilde{F}$-family and hence $B$ is a quotient of $C^*(\tilde{F})$. Thus $RR(B) = 0$. Now, it is not hard to see that $C^*(E) = B + I(H)$, and this completes the proof.

Let $A$ be a $\{0, 1\}$-matrix with no zero row or column. Then $A$ can be viewed as a vertex matrix of a finite graph $E$ with no sinks. If $A$ satisfies Cuntz-Krieger’s condition (I) in [CK] then it clearly follows that $E$ satisfies (L) (or, equivalently condition (I) introduced for graphs in [KPR]) from their definitions. By Proposition 4.1 of [KPRR], the graph algebra $C^*(E)$ is also generated by a Cuntz-Krieger $A$-family of partial isometries, hence the Cuntz-Krieger algebra $O_A$ is isomorphic to the graph algebra $C^*(E)$. On the other hand, the graph algebra $C^*(E)$ is known to be isomorphic to the Cuntz-Krieger algebra $O_B$ associated with the edge matrix $B$ of $E$. Therefore those three algebras are all isomorphic. Furthermore by Theorem 4.3, 4.6, and Lemma 6.1 of [KPRR], we have the following corollary.

**Corollary 4.7.** Let $A$ be a $\{0, 1\}$-matrix with no zero row or column. Suppose $A$ satisfies Cuntz-Krieger’s condition (I) and let $E$ be the finite graph having $A$ as its vertex matrix. Then the following are equivalent:

(i) $RR(O_A) = 0$,

(ii) $A$ satisfies Cuntz’s condition (II),

(iii) $E$ satisfies condition (K).

---

**References**


Received May 30, 1999 and revised October 6, 1999. The first author was partially supported by GARC-KOSEF, BSRI and the third author was partially supported by BSRI-97-1420.

BK 21, MATHEMATICAL SCIENCES DIVISION
SEUL NATIONAL UNIVERSITY
SEUL, 151-742
KOREA
E-mail address: jajeong@math.snu.ac.kr

DEPARTMENT OF MATHEMATICS
HANSHIN UNIVERSITY
OSAN, 447-791
KOREA
E-mail address: ghpark@hanshin.ac.kr

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SEOUL
SEOUL, 130-743
KOREA
E-mail address: dyshin@uoscc.uos.ac.kr
REPRESENTATION TYPE OF COMMUTATIVE NOETHERIAN RINGS I: LOCAL WILDERNESS

LEE KLINGLER AND LAWRENCE S. LEVY

This is the first of a series of four papers describing the finitely generated modules over all commutative noetherian rings that do not have wild representation type (with a possible exception involving characteristic 2). This first paper identifies the wild rings, in the complete local case. The second paper describes the finitely generated modules over the remaining complete local rings. The last two papers extend these results by dropping the “complete local” hypothesis.

1. Introduction.

The goal of this project is to describe all finitely generated modules — including the indecomposables and direct-sum behavior — over as wide as possible a class of commutative noetherian rings, thus extending Steinitz’s well-known 1911 theorem on modules over Dedekind domains [S]. Steinitz’s theorem has been particularly resistant to generalization. In fact, the only noetherian integral domains (other than Dedekind domains) for which such a module structure theorem was known prior to completion of this project seem to be the Dedekind-like domains studied in [L2]. However, for rings that are not domains, some other results exist [L1, L2, NR, NRSB].

We call a ring \( \Lambda \) finitely generated tame if we can describe all isomorphism classes of finitely generated \( \Lambda \)-modules. In all cases for which we cannot obtain such a description (with a possible exception involving characteristic 2), the obstruction is wild representation type; or more precisely, finite-length wildness.

Informally, a commutative ring \( \Lambda \) is finite-length wild if it has a residue field \( k \) such that any description of all isomorphism classes of \( \Lambda \)-modules of finite length would have to contain a description of all isomorphism classes of finite-dimensional modules over all finite-dimensional \( k \)-algebras. The seeming hopelessness of this task is what is behind the name “wild representation type.” (The precise definition of finite-length wild is given in Subsection 2.2.) The notion of tame versus wild representation type has been important in the study of finite-dimensional algebras for more than 20 years but seems to be relatively new in commutative noetherian rings. So,
we follow our formal Definition 2.2 by some introductory remarks on the subject [Remarks 2.3].

One of the surprising facts about finite dimensional algebras is that the majority of algebras over algebraically closed fields are wild. The well-known tame-wild theorem of Drozd and Crawley-Boevey [CB, Theorem B] states that, if \( k \) is algebraically closed, then every finite dimensional \( k \)-algebra is either tame or wild, but never both.

One of the main objectives of the present series of papers is to obtain a similar tame-wild dichotomy for commutative noetherian rings. It is possible to give a precise definition of “tame” in the context of finite dimensional algebras over algebraically closed fields. (See, for example, [CB, Definition 6.5].) In the present series, however, “finitely generated tame” has the informal meaning stated in the second paragraph of this introduction.

For readability, this project is divided into a series of four papers.

**Paper I: Wildness, complete local case.** Let \((\Lambda, \mathfrak{m}, k)\) be a complete local commutative noetherian ring, and \(\mu_\Lambda(\mathfrak{m})\) the minimal number of generators of the \(\Lambda\)-module \(\mathfrak{m}\). We give the spirit of our Main Wildness Theorem [2.10] without the many definitions that are needed for its precise statement.

If \(\mu_\Lambda(\mathfrak{m}) \geq 3\), then \(\Lambda\) is always finite-length wild. On the other hand, if \(\mu_\Lambda(\mathfrak{m}) = 1\), then \(\Lambda\) is a principal ideal ring, and its tameness was well-known long before the word “tame” came into use. Thus \(\mu_\Lambda(\mathfrak{m}) = 2\) is the dividing line between tameness and wildness; moreover, the vast majority of rings with \(\mu_\Lambda(\mathfrak{m}) = 2\) are wild. In order to make this dividing line precise, we need to define several types of rings, which we do in Section 2.

We call \(\Lambda\) an \textit{artinian triad} if \(\mu_\Lambda(\mathfrak{m}) = 3\) and \(\mathfrak{m}^2 = 0\). These are clearly the “smallest” rings such that \(\mu_\Lambda(\mathfrak{m}) = 3\), and they are known to be finite-length wild. We also define a special kind of local artinian ring \(\Lambda\) of composition length 5, with \(\mu_\Lambda(\mathfrak{m}) = 2\), and call it a \textit{Drozd ring}. We prove that these are finite-length wild in Section 4.

Section 3 is devoted to the proof of our Ring-theoretic Dichotomy Theorem [3.1], which states that every complete local ring either (i) maps onto an artinian triad or a Drozd ring, or (ii) is a homomorphic image of a type of ring of Krull dimension 1 that we call \textit{Dedekind-like} or is an exceptional type of artinian ring that we call a \textit{Klein ring}. “Dedekind-like rings” are reduced rings, satisfy \(\mu_\Lambda(\mathfrak{m}) \leq 2\), and are very close to their normalization (in their total quotient ring), which is either a DVR (discrete valuation ring) or the direct sum of two DVRs. “Klein rings” have composition length 4 and satisfy \(\mu_\Lambda(\mathfrak{m}) = 2\), a special case being the group algebra of the Klein 4-group over a field of characteristic 2. (See Section 2 for precise definitions.)

Our Ring-theoretic Dichotomy Theorem [3.1] is a piece of pure commutative algebra; neither it nor its proof make any use of the notions of tame
or wild. In order to use this ring-theoretic dichotomy to prove our tame-wild dichotomy, we need to prove that rings of type (ii) (in the previous paragraph) are tame. We postpone this to paper II.

Extending earlier terminology of Ringel [R], we think of artinian triads and Drozd rings as “minimal wild rings,” and Dedekind-like rings and Klein rings as “maximal tame rings.” Thus, anticipating the results of paper II, our dichotomy theorem implies that every Λ either maps onto a minimal wild ring, and hence is finite-length wild, or is a homomorphic image of a maximal tame ring, and hence is finitely-generated tame. (The possible exception involving characteristic 2 is described in our discussion of paper II below.)

Paper I ends with three short sections containing examples and miscellaneous results:

Section 5 gives an example of a Klein ring that is not an algebra over a field, and studies when Klein rings are, and when they are not, homomorphic images of Dedekind-like rings.

Section 6 gives a second, more constructive definition of Drozd rings than the abstract definition in Subsection 2.4. This section also shows that all of the ramified complete local orders studied in integral representation theory are finite-length wild.

Section 7 shows that all complete local orders of infinite lattice type are finite-length wild.

Remarks on earlier work. Let Λ be a commutative noetherian ring that is a finitely generated algebra over an algebraically closed field. In this context, our tame-wild dichotomy for Λ-modules of finite length was obtained by Drozd [D], and the solution improved by Ringel [R]. The actual module structure in the tame case for modules of finite length — in fact, for finitely generated modules — had been given by earlier results of others ([NR], corrected in [NRSB]). This collection of results was the source of our interest in the subject.

As Drozd and Ringel were only interested in Λ-modules of finite length, they were able to assume, without loss of generality, that Λ is a complete local ring. Since our rings do not have to be algebras over fields and do not have to be local rings, our results on finitely generated modules apply to rings of algebraic integers, the rings that originally interested Steinitz. Our results also appear to be new in the situation that Λ is an artinian ring that is not an algebra over a field, and in the case of algebras over non-algebraically-closed fields. (On the other hand, Ringel’s paper does contain noncommutative results which we do not handle.)

As mentioned above, the original Nazarova-Roiter results in [NR] deal with finitely generated modules over the local rings that they investigated, a
special case of our Dedekind-like rings. But since Drozd’s paper, the “finite-length wild versus finitely-generated tame” aspect of the problem seems to have been mostly neglected.

In a later paper [D2], Drozd states a tame-wild dichotomy over a class of rings that can be noncommutative and need not be algebras over a field. However, most of the statements and all of the proofs of theorems in that paper assume that the ring is an algebra over an algebraically closed field. Thus, in the commutative case, the results proved in that paper do not go beyond those in his earlier paper. Moreover, in the commutative case, his unproved statement of which rings are tame does not take the simple, explicit form that is given by our definition of split and unsplit Dedekind-like rings and Klein rings. (See our §2.)

**Paper II: Tameness, complete local case [KL2]**. This long paper gives a detailed description of all finitely generated modules over Klein rings and over complete local Dedekind-like rings, with the possible exception described in (1.0.1). Except for this possible exception, this completes our finitely-generated tame versus finite-length wild dichotomy for complete local rings Λ.

(1.0.1) *The possible exception*. There are three types of Dedekind-like rings: split, unsplit, and DVRs (see Definition 2.5 below). Let Λ be a complete, unsplit Dedekind-like ring, with maximal ideal m and residue field k, and let Γ be the normalization of Λ. By definition of “unsplit Dedekind-like,” m is also the unique maximal ideal of Γ, and Γ/m is a quadratic field extension of k = Λ/m. When the quadratic field extension Γ/m of k is inseparable, our theory breaks down, and we do not know whether Λ is tame or wild or neither.

Note, however, that *this exception cannot occur for rings of algebraic integers* (whose residue fields are finite) *nor for rings of geometric origin* (whose residue fields are algebraically closed).

**Paper III: Global Wildness [KL3]**. In this short paper we remove the “complete local” hypothesis from the main wildness theorem of Paper I. We may assume, without loss of generality, that we are given a commutative, noetherian, indecomposable ring Ω. Our main result is that, if Ω is not finite-length wild, then Ω is either a homomorphic image of a “global Dedekind-like ring” Λ (that is, a reduced ring of Krull dimension 1, all of whose completions ˆΛ_m at maximal ideals are the local Dedekind-like rings, as previously defined), or Ω is a Klein ring.

As in paper I, we postpone the statements and proofs of the corresponding tameness results to the next paper in the series.

**Paper IV: Global Tameness [KL4]**. Let Λ be a global Dedekind-like ring, as defined in the discussion of paper III above. We describe the structure of
all finitely-generated $\Lambda$-modules, provided that no completion $\hat{\Lambda}_m$ is one of
the possible exceptions involving characteristic 2, described in (1.0.1).

Here it is not sufficient to describe the indecomposable $\Lambda$-modules. Since
$\text{mod-}\Lambda$, the category of all finitely generated $\Lambda$-modules, is not a Krull-
Schmidt category, we also need to describe the direct-sum relations of finitely
generated $\Lambda$-modules. Indeed, this and the local-global relations in $\text{mod-}\Lambda$
occupy most of this rather long paper.

One additional and possibly surprising complication is that — unlike rings
of number-theoretic or geometric origin — it is not necessarily true that all
but finitely many completions $\Lambda_m$ of our non-local Dedekind-like ring $\Lambda$ be
DVRs. In fact, it is possible for $\Lambda$ to have infinitely many maximal ideals
and none of its completions be DVRs.

An interesting application of our structure theory is that every ring of the
form $\mathbb{Z}[\sqrt{n}]$, with $n$ a square-free integer, is among the tame rings whose
finitely generated modules we describe. (Note that $\mathbb{Z}[\sqrt{n}]$ is not always the
full ring of algebraic integers in $\mathbb{Q}[\sqrt{n}]$.)

2. Definitions, main theorem.

In this section we define finite-length wildness and the types of rings that
appear in the statement of our Main Wildness Theorem [2.10], and give
some examples. Then we state the theorem itself and give some additional
eamples.

Notation 2.1. Throughout this paper, $\Lambda$ denotes a commutative noether-
ian ring. We say that $(\Lambda, m, k)$ is a local ring if $\Lambda$ is a noetherian ring with
unique maximal ideal $m$ and residue field $k$. We say that $(\Lambda, m, k)$ is complete
if it is $m$-adically complete.

We consistently write functions on the left except when they represent
matrix multiplication. It is typographically simpler to display a row of a
matrix than a column. Therefore our matrices normally act via right multi-
plication, and we write the corresponding functions as right operators. This
occurs throughout this series of papers, and we include a reminder when it
happens.

We let $\mu_\Lambda(M)$ denote the minimal number of generators required by a
$\Lambda$-module $M$.

Definition 2.2 (Finite-length wildness). Let $\text{fdmod-}k(X,Y)$ denote the
category of finite dimensional right modules over the free noncommutative
$k$-algebra in two indeterminates $X$ and $Y$.

We say that $\Lambda$ is finite-length wild (with respect to $k$) if $k$ is a residue
field of $\Lambda$, and there is a full subcategory $W$ of the category of $\Lambda$-modules
of finite length and an additive functor $\Phi : W \to \text{fdmod-}k(X,Y)$ such that
$\Phi$ is a representation equivalence; that is, $\Phi$ is: dense (onto all isomorphism
classes), faithful ($\Phi(M) \cong \Phi(N) \iff M \cong N$), and full (a surjection
on homomorphism groups). Thus, encoded in \( W \) we find not only all of the isomorphism classes in \( \text{fdmod-}k\langle X,Y \rangle \), but (after reducing the homomorphism groups in \( W \) modulo suitable kernels) all of the hom-structure of \( \text{fdmod-}k\langle X,Y \rangle \) as well.

**Remarks 2.3** (Introduction to wildness).

(i) (Meaning of wildness) Every module over any 2-generator \( k \)-algebra \( A \) is also a \( k\langle X,Y \rangle \)-module. Therefore \( \text{fdmod-}k\langle X,Y \rangle \) contains \( \text{fdmod-}A \) as a subcategory. Moreover, a well-known simple but clever trick of Brenner [\( B, \text{Theorem 3} \)] shows that \( \text{fdmod-}k\langle X,Y \rangle \) contains a copy of \( \text{fdmod-}A \), for every finitely generated \( k \)-algebra \( A \). Thus, any classification of all isomorphism classes in \( \text{fdmod-}k\langle X,Y \rangle \) would contain a classification of all isomorphism classes of finite dimensional \( A \)-modules for every finitely generated \( k \)-algebra \( A \).

Suppose, now, that our commutative ring \( \Lambda \) is finite-length wild, and let \( M \) and \( N \) be finite-dimensional modules over some 2-generator algebra \( A \). We wish to know whether \( M \cong N \). By density of the functor \( \Phi \) in our definition of finite-length wild, we have \( M \cong \Phi(M') \) and \( N \cong \Phi(N') \) for \( \Lambda \)-modules \( M', N' \in W \). Moreover, by faithfulness of \( \Phi \) we have that \( M' \cong N' \) as \( \Lambda \)-modules if and only if \( M \cong N \) as \( k\langle X,Y \rangle \)-modules; or equivalently, as \( A \)-modules. Moreover, in this illustration of the meaning of wildness, we can drop the requirement that \( A \) be 2-generated, by using Brenner’s trick.

Thus we see that, if \( \Lambda \) is finite length wild, then any explicit description of all isomorphism classes of \( \Lambda \)-modules of finite length would have to contain a description of all isomorphism classes of finite-dimensional \( A \)-modules for every finite dimensional \( k \)-algebra \( A \), as mentioned in the informal definition of wildness given in the introduction to this paper.

(ii) (Wildness of \( \Lambda \) versus its localizations and completions) Let \( m \) be a maximal ideal of \( \Lambda \), and suppose that the \( m \)-localization \( \Lambda_m \) or \( m \)-adic completion \( \hat{\Lambda}_m \) is finite-length wild. Then \( \Lambda \) is also finite-length wild. The reason for this is that every \( \Lambda_m \)-module of finite length and every \( \Lambda_m \)-module of finite length is also a \( \Lambda \)-module of finite length.

Therefore, in order to prove that \( \Lambda \) is finite-length wild, we can assume that \( (\Lambda, m, k) \) is a complete local ring, the situation considered in this paper.

(iii) (Strict wildness) A \( k \)-algebra \( A \) is sometimes called strictly wild if there is a full exact imbedding \( \Psi: \text{fdmod-}k\langle X,Y \rangle \rightarrow \text{fdmod-}A \). This means that \( \Psi \) is an isomorphism on hom groups, takes exact sequences in \( \text{fdmod-}k\langle X,Y \rangle \) to exact sequences in \( \text{fdmod-}A \), and is an imbedding of categories.

**Definition 2.4** (Artinian triad, Drozd ring). We call the local ring \( (\Lambda, m, k) \) an artinian triad if \( \mu_\Lambda(m) = 3 \) and \( \bar{m}^2 = 0 \). (Since \( \bar{m}^2 = 0 \), every artinian triad is indeed an artinian ring.) In the case of finite dimensional algebras
over the field $k = \Lambda/m$, this ring is $\Lambda = k[X,Y,Z]/(X,Y,Z)^2$, where $X$, $Y$, and $Z$ are indeterminates.

We call the local ring $(\Lambda,m,k)$ a Drozd ring if $\mu_\Lambda(m) = \mu_\Lambda(m^2) = 2$, $m^3 = 0$, and there is an element $x \in m - m^2$ such that $x^2 = 0$. (Since $m^3 = 0$, Drozd rings are artinian.) In the case of finite dimensional algebras over the field $k = \Lambda/m$, this ring is the 5-dimensional $k$-algebra with $k$-basis $1, x, y, xy, y^2$ and all other monomials equal to zero. An example of a Drozd ring that is not a algebra over a field is the ring $A_p$ in Example 6.1.

We note:

(2.4.1) The composition length of every Drozd ring is 5.

To see this note that the dimensions of $\Lambda/m$, $m/m^2$, and $m^2$ as $k$-vector spaces are 1, 2, 2 respectively.

The above, abstract definition of Drozd rings is tailored to the needs of our dichotomy results, but leaves one wondering what these rings really look like. It turns out that every Drozd ring is a subring of an artinian principal ideal ring, and we use this fact to give a general, explicit construction of Drozd rings in Theorem 6.5.

**Definition 2.5 (Dedekind-like ring).** Let $(\Lambda,m,k)$ be a local ring. We call $\Lambda$ a Dedekind-like ring if $\Lambda$ is reduced (no nonzero nilpotent elements) and its normalization $\Gamma$ (in the total quotient ring of $\Lambda$) has the following properties: $\Gamma$ is a direct sum of principal ideal domains (necessarily semi-local), $m = \text{rad}(\Gamma)$ (the Jacobson radical of $\Gamma$), and $\mu_\Lambda(\Gamma) \leq 2$.

We do not consider fields to be principal ideal domains. Therefore: $\Gamma$ and $\Lambda$ have Krull dimension 1.

This abstract definition is tailored to the needs of the proof of our dichotomy theorem. For a more constructive definition and examples, see Notation 2.13 and the lemmas and examples that follow it.

Note that the reduced ring $\Gamma/m$ is a vector space over $k = \Lambda/m$ of dimension at most 2. Therefore exactly one of the following three possibilities holds, and we attach the indicated name to $\Lambda$.

(2.5.1) We call the Dedekind-like ring $\Lambda$:

(i) **Split** if $\Gamma/m \cong k \times k$ as rings.
(ii) **Unsplit** if $\Gamma/m$ is a 2-dimensional field extension of $k$. (We usually call this field $F$.)
(iii) A **DVR (discrete valuation ring)** if $\Gamma/m = k$; that is, $\Lambda = \Gamma$.

Since $\Gamma/m$ has $k$-dimension at most 2 and $m = \text{rad}(\Gamma)$, $\Gamma$ must be either an integral domain or the direct sum of two integral domains. We attach names to the corresponding possibilities as follows.

(2.5.2) We call the split Dedekind-like ring $\Lambda$:

(i) **Strictly split** if $\Gamma$ is the direct sum of two integral domains (necessarily DVRs).
(ii) **Nonstrictly split** if $\Gamma$ is an integral domain (necessarily a PID with exactly two maximal ideals).

We note: *If the split Dedekind-like ring $\Lambda$ is complete, then $\Lambda$ is strictly split.* To see this, note first that $\Lambda \Gamma$ is $m$-adically complete, since $\Lambda \Gamma$ is finitely generated ([N, Theorem 17.8]). Since the ideal $m$ of $\Lambda$ also equals $\text{rad}(\Gamma)$, we see that the semilocal ring $\Gamma$ is also $m$-adically complete, and therefore a direct sum of local rings. The fact that $\Lambda$ is strictly split now follows from (2.5.2).

If $\Lambda$ is unsplit then $\Gamma$ is a DVR, because $\Gamma/m = \Gamma/\text{rad}(\Gamma)$ is a field.

**Remark 2.6** (Consistency with terminology of [L2]). All localizations and completions at maximal ideals of the rings called “Dedekind-like” in [L2] are split Dedekind-like, in our present terminology [by Lemma 2.14 below], or DVRs. Thus, for local rings, our present terminology generalizes that of [L2]. When we consider the non-local situation in [KL3] and [KL4], our terminology will again generalize that in [L2].

In order to prove that Dedekind-like rings that are not DVRs lie on the tame-wild dividing line mentioned in the introduction, we show:

**Lemma 2.7.** If $(\Lambda, m, k)$ is Dedekind-like but not a DVR, then $\mu_{\Lambda}(m) = 2$.

**Proof.** By Nakayama’s Lemma it suffices to show that the $k$-vector space $m/m^2$ has dimension 2. Since $m$ is also an ideal of the principal ideal ring $\Gamma$, we have $m/m^2 \cong \Gamma/m$ as $\Gamma$-modules, and hence as $\Lambda$-modules, and therefore as $k$-vector spaces. This last dimension equals 2 except if $\Lambda$ is a DVR, by (2.5.1).

**Definition 2.8** (Klein ring). We call a local ring $(\Lambda, m, k)$ a **Klein ring** if $\mu_{\Lambda}(m) = 2$, $\mu_{\Lambda}(m^2) = 1$, $m^3 = 0$, and $x^2 = 0$ ($\forall x \in m$). (Since $m^3 = 0$, Klein rings are artinian and hence definitely not reduced.)

The group algebra of the Klein 4-group over a field of characteristic 2 is an example of a Klein ring. We also note:

**Lemma 2.9.** If $\Lambda$ is a Klein ring with residue field $k$, then $k$ has characteristic 2, and $\Lambda$ has characteristic 2 or 4.

**Proof.** Let $m = (x, y)$. Then $0 = (x + y)^2 = x^2 + 2xy + y^2 = 2xy$. Since $\mu_{\Lambda}(m^2) = 1$ and $x^2 = y^2 = 0$, we cannot have $xy = 0$. Therefore $2xy = 0$ implies that $2 \in m$. Since $\Lambda$ is a Klein ring, this implies that $4 = 2^2 = 0$ in $\Lambda$, completing the proof.

For an example in which characteristic 4 actually occurs, see Example 5.4.

Having now defined all of the ingredients needed to state our main theorem, we now state the theorem itself.
**Theorem 2.10** (Main Wildness Theorem). Let $(\Lambda, m, k)$ be a complete local ring. Then exactly one of the following holds.

(i) $\Lambda$ maps onto an artinian triad or a Drozd ring, in which case $\Lambda$ is finite-length wild.

(ii) $\Lambda$ is either a Klein ring or a homomorphic image of a strictly split or unsplit Dedekind-like ring.

**Proof.** Our Ring-theoretic Dichotomy Theorem 3.1 states that either (ii) holds, or else $\Lambda$ maps onto an artinian triad or onto a Drozd ring. Since finite-length wildness clearly carries up from homomorphic images, it therefore suffices to prove that artinian triads and Drozd rings are finite-length wild.

Artinian triads are wild by a theorem of Warfield [GLW, Lemma 3]; and Drozd rings are wild by Theorem 4.9. □

**Remarks 2.11.** (i) As previously mentioned, we finish the complete local case of our tame-wild project in the second paper of this series [KL2] by describing the structure of all finitely generated $\Lambda$-modules when $\Lambda$ is Dedekind-like or a Klein ring and the possible exception (1.0.1) does not occur.

(ii) There exist Klein rings that are not homomorphic images of Dedekind-like rings (Theorem 5.2). Therefore, we cannot simplify Theorem 2.10 by deleting the phrase “Klein ring” from statement (ii). On the other hand, some Klein rings are homomorphic images of Dedekind-like rings (again, Theorem 5.2).

(iii) The reason that DVRs can be omitted from the statement of Theorem 2.10(ii) is that every DVR is a homomorphic image of a strictly split Dedekind-like ring [Lemma 2.19].

(iv) Note that Theorem 2.10 does not state that $\Lambda$ cannot be both tame and wild. For finite-dimensional algebras over an algebraically closed field, the terms “tame” and “wild” are known to be mutually exclusive [CB, Theorem B]. In the context of commutative noetherian rings, the question remains open.

**Corollary 2.12.** Every noetherian ring $\Lambda$ (not necessarily local) of Krull dimension greater than one is finite-length wild.

**Proof.** Since $\Lambda$ has dimension greater than 1, $\Lambda$ has a maximal ideal $m$ such that the $m$-localization $\Lambda_m$ has dimension greater than 1, and hence [N, 17.12] the $m$-adic completion $\hat{\Lambda}_m$ has dimension greater than 1. Therefore we may assume that $\Lambda$ itself is a complete local ring of dimension greater than 1 [Remarks 2.3(ii)].

Dedekind-like rings have Krull dimension 1, and Klein rings are artinian, hence have Krull dimension 0. Therefore, by Theorem 2.10, $\Lambda$ is finite-length wild. □
It is often more informative to view Dedekind-like rings in terms of their conductor square. The last few results in this section deal with this point of view.

**Notation 2.13** (Conductor Square). Consider the following commutative diagram of ring homomorphisms, where \( m \) is an ideal of both rings \( \Lambda \) and \( \Gamma \) (a "conductor" ideal).

\[
\begin{array}{ccc}
\Lambda & \subseteq & \Gamma \\
\downarrow \rho & & \downarrow \rho \\
k & \subseteq & \Gamma / m
\end{array}
\]

(2.13.1)

We say that \( \Lambda \) is the pullback of this conductor square if \( \Lambda = \{ x \in \Gamma \mid \rho(x) \in k \} \).

The next three simple lemmas make clear exactly what ingredients are needed to construct arbitrary Dedekind-like rings.

**Lemma 2.14** (Split Dedekind-like). Let \( \Gamma \) be a principal ideal ring with exactly two maximal ideals and no artinian ring direct summands. Let \( m = \text{rad}(\Gamma) \) (the Jacobson radical of \( \Gamma \)), and suppose that \( \Gamma / m \cong k \times k \), the direct product of two copies of some field \( k \). View \( k \) as a subfield of \( k \times k \) via the diagonal map \( x \rightarrow (x,x) \), and let \( \Lambda \) be the pullback of square (2.13.1). Then \( (\Lambda, m, k) \) is a split Dedekind-like ring with normalization \( \Gamma \), and every split Dedekind-like ring is isomorphic to such a \( \Lambda \).

**Proof.** The only part of this proof that requires explicit mention is perhaps to note that we have \( \mu_{\Lambda}(\Gamma) \leq 2 \) because the \( k \)-dimension of \( \Gamma \) is 2; and therefore the integrally closed reduced ring \( \Gamma \) is the normalization of \( \Lambda \). □

The next lemma is particularly useful in when the local ring \( \Lambda \) is complete, since in that situation all split Dedekind-like rings are strictly split.

**Lemma 2.15** (Strictly Split Dedekind-like). Let \((V_1, m_1, k)\) and \((V_2, m_2, k)\) be DVRs with a common residue field \( k \). For each \( i \) let \( f_i : V_i \rightarrow k \) be a surjective ring homomorphism, and let

\[
\Lambda = \{ [v_1, v_2] \in V_1 \times V_2 \mid f_1(v_1) = f_2(v_2) \}.
\]

(2.15.1)

Then \( (\Lambda, (m_1 \times m_2), k) \) is a strictly split Dedekind-like ring with normalization \( \Gamma = V_1 \times V_2 \). Moreover, every strictly split Dedekind-like ring is isomorphic to such a ring \( \Lambda \).

**Proof.** To verify the first assertion, build conductor square (2.13.1) as follows. Let \( \bar{\Gamma} = k \times k \) and view \( k \) as a subring of \( \bar{\Gamma} \) via the diagonal map \( x \rightarrow (x,x) \). Let \( m = m_1 \times m_2 \) and \( \rho = f_1 \times f_2 \). Then we can identify \( \rho(\Gamma) \) with \( \Gamma / m \). Now use Lemma 2.14 □

The proof of the next lemma is essentially the same as that of Lemma 2.14.
Lemma 2.16 (Unsplit Dedekind-like). Let $\Gamma$ be a DVR. Let $m = \text{rad}(\Gamma)$, suppose that $\Gamma/m$ is a 2-dimensional field extension of some subfield $k$, and let $\Lambda$ be the pullback of square (2.13.1). Then $(\Lambda, m, k)$ is an unsplit Dedekind-like ring with normalization $\Gamma$, and every unsplit Dedekind-like ring is isomorphic to such a $\Lambda$.

Although the class of Dedekind-like rings is very small, it contains many interesting examples; we give a few below. For more examples, and tameness properties of Dedekind-like rings, see [KL2].

Examples 2.17 (Split).

(i) The simplest example of a strictly split Dedekind-like ring is formed by taking $V_1$ and $V_2$ to be the $p$-localization or $p$-adic completion of the integers, for some prime number $p$, in Lemma 2.15. Note that this is not an algebra over a field.

For the minor modification that is an algebra over any field $k$, use Lemma 2.15 and take $V_1$ and $V_2$ to be $k[[X]]$ (formal power series).

(ii) The ring $k[[X, Y]]/(XY)$ is strictly split Dedekind-like. To see this, note that every element can be represented by an ordered pair of power series in one variable, where both power series have the same constant term. Thus this is isomorphic to the example constructed in the second paragraph of item (i) above.

(iii) Let $\Lambda = \hat{\mathbb{Z}}_p[[X]]/(pX)$ where $p$ is any prime number and $\hat{\mathbb{Z}}_p$ denotes the $p$-adic integers. Then $\Lambda$ is a strictly split Dedekind-like ring. One can view this as the number-theoretic analog of the ring $k[[X, Y]]/(XY)$ in item (ii).

To prove this, let $V_1 = \hat{\mathbb{Z}}_p[[X]]/(X) \cong \hat{\mathbb{Z}}_p$ and $V_2 = \hat{\mathbb{Z}}_p[[X]]/(p) \cong \mathbb{Z}_p[[X]]$, both of which are complete DVRs. The natural map $\nu: \Lambda \to V_1 \oplus V_2$ is clearly one-to-one, and it is not difficult to verify that $\nu$ maps $\Lambda$ onto the split Dedekind-like ring described in (2.15.1) and Lemma 2.15, with $k = \mathbb{Z}_p = \mathbb{Z}/(p)$ and $f_1$ and $f_2$ the natural homomorphisms.

(iv) (Nonstrictly split). Let $\Gamma$ be any principal ideal domain that has exactly two maximal ideals $m_1$ and $m_2$ with isomorphic residue fields $k$. (For example start with the ring of algebraic integers in any quadratic extension of the rationals, and localize away all but two maximal ideals that lie over some common rational prime.) Then let $\Lambda$ be the pullback of diagram (2.13.1) in which $\rho$ is a surjective ring homomorphism $\Gamma \to k \times k$ with kernel $m = m_1 \cap m_2$. $\Lambda$ is split Dedekind-like by Lemma 2.14, and is not strictly split since it is an integral domain.

We remind the reader that nonstrictly split Dedekind-like rings can never be complete local rings.

Examples 2.18. The easiest example of an unsplit Dedekind-like ring is the ring $\Lambda = \mathbb{R} + X\mathbb{C}[[X]]$ of power series whose constant terms are real
numbers and whose other coefficients are complex numbers. To see this, use Lemma 2.16 with \( \Gamma = \mathbb{C}[[X]] \), \( \bar{\Gamma} = \mathbb{C} \), and \( k = \mathbb{R} \).

For an example that is not an algebra over a field, let \( \Gamma' \) be the ring of all integers in any number field such that some residue field \( F \) of \( \Gamma' \) is a 2-dimensional extension of some subfield \( k \). Then let \( \Gamma \) be the localization or completion of \( \Gamma' \) with respect to the kernel of \( \Gamma' \to F \). Finish the definition by letting \( \Lambda \) be the pullback of diagram (2.13.1) and using Lemma 2.16.

Finally, for future reference, we record three useful lemmas.

**Lemma 2.19.** Let \( \Omega \) be either a DVR or an artinian local principal ideal ring. Then \( \Omega \) is a homomorphic image of a strictly split Dedekind-like ring \( \Lambda \). Moreover, if \( \Omega \) is a complete DVR or is artinian, then \( \Lambda \) can be chosen to be complete.

*Proof.* Let \((V, n, k)\) be a DVR, and \( f: V \to k \) the natural homomorphism. Then let
\[
\Lambda = \{ [x, y] \in V \times V \mid f(x) = f(y) \}.
\]
The ring \( \Lambda \) is strictly split Dedekind-like, by Lemma 2.15, and maps onto \( V \). Moreover, if \( V \) is complete, then \( \Lambda \) is also complete.

Let \( A \) be an artinian local principal ideal ring. A theorem of Hungerford [Hu] states that some complete DVR \( V \) maps onto \( A \). Since \( \Lambda \) maps onto \( V \), the proof is complete. \( \square \)

**Lemma 2.20.** Let \((\Lambda, m, k)\) be a strictly split Dedekind-like ring. Then \( m \) is generated by two elements \( x, y \) such that \( xy = 0 \).

*Proof.* View \( \Lambda \) in the notation of Lemma 2.15. Then take generators \( p_1 \) and \( p_2 \) of the maximal ideals of \( V_1 \) and \( V_2 \) respectively. Finally, note that the elements \( x = [p_1, 0] \) and \( x = [0, p_2] \) of \( V_1 \times V_2 \) are elements of \( \Lambda \) and, in fact, generate \( m \). \( \square \)

**Lemma 2.21.** Let \((\Lambda, m, k)\) be an unsplit or split Dedekind-like ring with normalization \( \Gamma \). Then the \( m \)-adic completion \((\hat{\Lambda}, \hat{m}, k)\) is, respectively, an unsplit or (necessarily strictly) split Dedekind-like ring with normalization \( \hat{\Gamma} \), and \( \text{rad} \hat{\Gamma} = \hat{m} \).

*Proof.* Since \( m \) is an ideal of both rings \( \Lambda \) and \( \Gamma \), the \( m \)-adic completion of \( \Gamma \) as a \( \Lambda \)-module coincides with the \( m \)-adic completion of \( \Gamma \) as a \( \Gamma \)-module. Therefore we can denote either by \( \hat{\Gamma} \). Moreover, since \( \Gamma \) is a finitely generated \( \Lambda \)-module, its \( m \)-adic completion as a \( \Lambda \)-module can be identified with \( \hat{\Lambda} \otimes_\Lambda \Gamma \); that is, \( \hat{\Gamma} = \hat{\Lambda} \otimes_\Lambda \Gamma \).

Viewing \( \hat{\Gamma} \) as the \( m \)-adic completion as a \( \Gamma \)-module, we see that \( \hat{\Gamma} \) is a DVR or the direct sum of two DVRs according as \( \Lambda \) is, respectively, unsplit or split.
Now tensor pullback diagram (2.13.1) for \( \Lambda \) by \( \hat{\Lambda} \) over \( \Lambda \), remembering that tensoring a pullback diagram by a flat module yields a pullback diagram (use the idea in e.g., [K, Proposition 2.10] or [L2, Lemma 6.1]). We claim that the resulting pullback diagram shows that \( \Lambda \) is the appropriate kind of Dedekind-like ring. In view of what has already been proved, we only need to note two more facts: (i) The finite-length modules in the bottom row of the original pullback diagram are unchanged by completion; and (ii) \( \ker(\hat{\rho}) = \text{rad}(\hat{\Gamma}) \). To prove this last fact, call the kernel \( K \). For every \( x \in K \), the element \( 1 - x \) is invertible since \( \hat{m} \) is the radical of the local ring \( \text{rad} \hat{\Lambda} \), and therefore \( K \subseteq \text{rad}(\hat{\Gamma}) \). The opposite inclusion holds since \( \hat{\rho}(\hat{\Gamma}) \) has radical zero.

□

3. Dichotomy theorem.

The purpose of this section is to prove the following result.

**Theorem 3.1** (Ring-theoretic Dichotomy). Let \((\Lambda, m, k)\) be a complete, local ring. Then exactly one of the following two possibilities holds.

(i) \( \Lambda \) has an artinian triad or a Drozd ring as homomorphic image.

(ii) \( \Lambda \) is a Klein ring or a homomorphic image of a complete, strictly split or unsplit Dedekind-like ring.

The remainder of this section is devoted to the proof of this theorem, which we accomplish by means of a series of propositions.

**Remark 3.2.** We begin by noting that, if \( \mu_{\Lambda}(m) \geq 3 \), then clearly \( \Lambda \) maps onto an artinian triad. On the other hand, if \( \mu_{\Lambda}(m) \leq 1 \), then by a theorem of Hungerford [Hu], \( \Lambda \) is a homomorphic image of a complete DVR \((\Gamma, m, k)\), and hence a homomorphic image of a complete, strictly split Dedekind-like ring (Lemma 2.19). Therefore, for the remainder of this section we can assume that \( \mu_{\Lambda}(m) = 2 \).

Thus, we can write \( m = (x, y) \), so that \( m^2 = (x^2, xy, y^2) \), and hence \( \mu_{\Lambda}(m^2) \leq 3 \). If in fact \( \mu_{\Lambda}(m^2) = 3 \), then \( \Lambda/((x^2) + m^3) \) is a Drozd ring. Therefore, for the remainder of this section, we can also assume that \( \mu_{\Lambda}(m^2) \leq 2 \).

The argument continues as we attempt to use generators \( x \) and \( y \) of \( m \) and the relations satisfied by \( x^2, xy, \) and \( y^2 \) to construct a strictly split or unsplit Dedekind-like ring that maps onto \( \Lambda \).

**Lemma 3.3.** Let \((\Lambda, m, k)\) be a complete, local ring, and suppose that \( \mu_{\Lambda}(m) = 2 \) and \( \mu_{\Lambda}(m^2) \leq 2 \). If \( x \) and \( y \) are generators of the maximal ideal \( m \), and if \( xy \in m^3 \), then we can find generators \( u \) and \( v \) of \( m \) such that \( uv = 0 \).

**Proof.** Note that for each integer \( n \geq 2 \) we have \( m^n = (x^n, x^{n-1}y, \ldots, xy^{n-1}, y^n) \). But \( xy \in m^3 \), so it follows that \( x^{n-1}y, \ldots, xy^{n-1} \in m^{n+1} \), and hence by Nakayama’s Lemma we have \( m^n = (x^n, y^n) \).
The strategy is to construct sequences $u_1, u_2, u_3, \ldots$ and $v_1, v_2, v_3, \ldots$ of elements of $\Lambda$, such that
\begin{equation}
\mathfrak{m} = (u_n, v_n), \quad u_n v_n \in \mathfrak{m}^{n+2}, \quad \text{and} \quad u_n - u_{n+1}, \ v_n - v_{n+1} \in \mathfrak{m}^{n+1}
\end{equation}
for each index $n$. It then follows immediately that the sequences are Cauchy sequences and, if we let $u = \lim_{n \to \infty} (u_n)$ and $v = \lim_{n \to \infty} (v_n)$, that $\mathfrak{m} = (u, v)$ and $uv = 0$.

The construction of the sequences is by induction on $n$. For $n = 1$, set $u_1 = x$ and $v_1 = y$. Now suppose, inductively, that we have selected elements $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ of $\Lambda$ satisfying (3.3.1).

As noted above, we have $\mathfrak{m}^{n+2} = (u_n^{n+2}, v_n^{n+2})$. Thus we can write
\begin{equation}
u_n v_n = au_n^{n+2} + bv_n^{n+2}
\end{equation}
for some $a, b \in \Lambda$. Rearranging (3.3.2) yields
\begin{equation}
u_n (v_n - au_n^{n+1}) = bv_n^{n+2}.
\end{equation}

Therefore setting $v_{n+1} = v_n - au_n^{n+1}$ gives a new pair of generators $\mathfrak{m} = (u_n, v_{n+1})$ such that $u_n - v_{n+1} \in \mathfrak{m}^{n+1}$. Now (3.3.3) can be rewritten as
\begin{equation}
u_n v_{n+1} = b(v_{n+1} + au_n^{n+1})^{n+2}.
\end{equation}

Expanding the right-hand side and rearranging yields
\begin{equation}(u_n - bv_{n+1}^{n+1})v_{n+1} \in \mathfrak{m}^{n+3}.
\end{equation}

Therefore setting $u_{n+1} = u_n - bv_{n+1}^{n+1}$ completes the induction, and hence the proof of the lemma.

Under the conditions of this technical lemma, we are ready to construct a complete, strictly split Dedekind-like ring that maps onto $\Lambda$.

**Proposition 3.4.** Let $(\Lambda, \mathfrak{m}, k)$ be a complete local ring, such that $\mu_{\Lambda}(\mathfrak{m}) = 2$, and suppose that $x$ and $y$ are a pair of generators of $\mathfrak{m}$ such that $xy \in \mathfrak{m}^3$. Then $\Lambda$ is a homomorphic image of a complete, strictly split Dedekind-like ring.

**Proof.** By Lemma 3.3, we can assume that $xy = 0$.

By the Structure Theorem of Complete Local Rings [N, Theorem 31.1 and preceding paragraph], there is a surjective homomorphism $\phi : V[[X, Y]] \to \Lambda$ such that $\phi(X) = x$ and $\phi(Y) = y$, where $V[[X, Y]]$ is a formal power series ring, and either $V = k$ or $V$ is a complete DVR of characteristic 0 with residue field $k$. Moreover, in this latter case, $k$ has characteristic $p \neq 0$, and the maximal ideal of $V$ is generated by $p$.

If $V$ is a field, let us set $R = V[[X, Y]]$, a complete, two-dimensional, regular local ring.

If $V$ is not a field, we note that $V[[X, Y]]$ is a complete, local, three-dimensional domain with maximal ideal $(p, X, Y)$, so that $V[[X, Y]]$ is a
regular local ring (by [AM, Theorem 11.22]) and therefore a unique factorization domain. Now \( \phi(p) \in \mathfrak{m} \) since \( k \) has characteristic \( p \), and hence there are elements \( A, B \in V[[X, Y]] \) such that \( \phi(p - (AX + BY)) = 0 \). Since \( p \) is irreducible in \( V \), clearly \( p - (AX + BY) \) is irreducible in \( V[[X, Y]] \), and hence \( (p - (AX + BY)) \) is a prime ideal. We set \( R = V[[X, Y]]/(p - (AX + BY)) \), a complete, local, two-dimensional domain by [AM, Corollary 11.18]. Moreover, the maximal ideal of \( R \) is now generated by two elements (the images of \( X \) and \( Y \)), so that \( R \) is in fact a regular local ring in this case also (again by [AM, Theorem 11.22]). Changing notation, we let \( \phi \) denote the induced map from \( R \) onto \( \Lambda \), and we denote by \( X, Y \) the cosets of the original elements \( X, Y \), respectively, in \( R \).

Thus, in either case, we have a surjective homomorphism \( \phi : R \to \Lambda \) such that \( \phi(X) = x \) and \( \phi(Y) = y \), where \( R \) is a complete, two-dimensional, regular local ring, \((X, Y)\) is the maximal ideal of \( R \), and \( R/(X, Y) \cong k \).

Now \( \phi(XY) = xy = 0 \), so that \( \phi \) induces a surjective homomorphism from \( R/(XY) \) onto \( \Lambda \). But the regular local ring \( R \) is a unique factorization domain, and \( X \) and \( Y \) are non-associate primes (being generators of the two-generated maximal ideal of \( R \)). Therefore, \((XY) = (X) \cap (Y)\) as ideals of \( R \). We set \( V_1 = R/(X) \) and \( V_2 = R/(Y) \). Each of \( V_1 \) and \( V_2 \) is a complete, local domain of Krull dimension one (since \( X \) and \( Y \) are irreducible), and each has a principal maximal ideal (since we mod out by the other generator), so that both \( V_1 \) and \( V_2 \) are complete DVRs.

Finally, by [CR1, 2.12], there is a pullback diagram with surjective maps:

\[
\begin{array}{ccc}
R/(XY) = R/((X) \cap (Y)) & \longrightarrow & V_2 = R/(Y) \\
\downarrow \quad & \quad \downarrow f_2 \\
V_1 = R/(X) & \longrightarrow & k = R/((X) + (Y)).
\end{array}
\]

That is, \( R/(XY) \) is isomorphic to the pullback

\[
\Omega = \{[s, t] \in V_1 \times V_2 \mid f_1(s) = f_2(t)\}
\]
of the pair of (natural) maps \( f_1 \) and \( f_2 \). Since \( V_1 \) and \( V_2 \) are DVRs and \( k \) is a field, the ring \( R/(XY) \cong \Omega \) is strictly split Dedekind-like by Lemma 2.15. Since \( \phi : R \to \Lambda \) is a surjective ring homomorphism whose kernel contains \((XY)\), our proof is complete.

Given the result of Proposition 3.4, we can assume that, for every pair of generators \( x \) and \( y \) of the maximal ideal \( \mathfrak{m} \) of \( \Lambda \), \( xy \notin \mathfrak{m}^3 \). We also recall that we assume that \( \mu_\Lambda(\mathfrak{m}^2) \leq 2 \). The next step is to reduce to the case where \( \mu_\Lambda(\mathfrak{m}^2) = 2 \).

**Proposition 3.5.** Let \((\Lambda, m, k)\) be a complete, local ring, such that \( \mu_\Lambda(m) = 2 \), and suppose that \( xy \notin m^3 \) for every pair of generators \( x \) and \( y \) of \( m \). Then either \( \mu_\Lambda(m^2) \geq 2 \), or \( \Lambda \) is a Klein ring.
Proof. Since we assume that \( xy \notin m^3 \) for every pair of generators \( x \) and \( y \) of \( m \), clearly \( \mu_\Lambda(m^2) \geq 1 \). Thus, we can assume that \( \mu_\Lambda(m^2) = 1 \), and it suffices to show that \( \Lambda \) is a Klein ring.

Fix a pair of generators \( x \) and \( y \) of \( m \). By assumption, \( \mu_\Lambda(m^2) = 1 \) and \( xy \notin m^3 \), so that in fact \( m^2 = (xy) \). Thus, \( x^2 \in (xy) \), say \( x^2 = axy \) for some element \( a \in \Lambda \). But then \( x(x-ay) = 0 \), and if \( a \) were a unit in \( \Lambda \), then we could replace \( y \) by \( x - ay \) to get a pair of generators of \( m \) whose product is 0, contrary to assumption. Therefore, \( a \in m \), and hence \( x^2 \in m^3 \). Similarly, \( y^2 \in m^3 \). But now \( m^3 = (x^3, x^2 y, xy^2, y^3) \subseteq m^4 \) implies, by Nakayama’s Lemma, that \( m^3 = 0 \). In particular, \( x^2 = y^2 = 0 \).

To complete the proof that \( \Lambda \) is a Klein ring, we must show that \( \Lambda \) is a homomorphic image of a complete unsplit Dedekind-like ring onto \( \Lambda \).

Proposition 3.6. Let \((\Lambda, m, k)\) be a complete local ring such that \( \mu_\Lambda(m) = \mu_\Lambda(m^2) = 2 \). Suppose that \( xy \notin m^3 \) for every pair of generators \( x \) and \( y \) of \( m \), and \( x^2 \notin m^3 \) for every element \( x \in m - m^2 \). Then \( \Lambda \) is a homomorphic image of a complete unsplit Dedekind-like ring.

Proof. Fix generators \( x \) and \( y \) of \( m \). By assumption, \( \mu_\Lambda(m^2) = 2 \), and \( xy \notin m^3 \). Thus, we can choose \( xy \) as one of the two generators of \( m^2 \). Without loss of generality we can assume that \( m^2 = (x^2, xy) \). Hence we can write

\[
(3.6.1) \quad y^2 + sxy + tx^2 = 0
\]

for some elements \( s, t \in \Lambda \). Let \( \overline{s} \) and \( \overline{t} \) denote the images of \( s \) and \( t \), respectively, in \( k = \Lambda/m \).

We claim that the polynomial \( Z^2 + \overline{s}Z + \overline{t} \) is irreducible in the polynomial ring \( k[Z] \). Suppose it were reducible, say \( Z^2 + \overline{s}Z + \overline{t} = (Z + \overline{u})(Z + \overline{v}) \), where \( \overline{u} \) and \( \overline{v} \) are the images in \( k \) of the elements \( u, v \in \Lambda \), respectively. This would imply that \( y^2 + sxy + tx^2 - (y + ux)(y + vx) \in m^3 \). Then by (3.6.1) we would have \( (y + ux)(y + vx) \in m^3 \). On the one hand, if \( u - v \in \Lambda - m \), then \( y + ux \) and \( y + vx \) would be generators of \( m \) (because their difference would be a unit times \( x \)), but their product is in \( m^3 \), contrary to the hypotheses.
of the proposition. On the other hand, if \( u - v \in m \), then it would follow that \( (y + ux)^2 \in m^3 \). But \( x \) and \( y \) are linearly independent modulo \( m^2 \), since they generate \( m \) and \( \mu_A(m) = 2 \). Therefore \( y + ux \in m - m^2 \), and this again contradicts the hypotheses of the proposition. Therefore the claim holds.

We continue as in the proof of Proposition 3.4. Using the Structure Theorem of Complete Local Rings [N, Theorem 31.1], there is a surjective homomorphism \( \phi : V[[X,Y]] \to \Lambda \) such that \( \phi(X) = x \) and \( \phi(Y) = y \), where \( V[[X,Y]] \) is a formal power series ring, and either \( V = k \) or \( V \) is a complete DVR of characteristic 0 with residue field \( k \), and where, in this latter case, \( k \) has characteristic \( p \neq 0 \) and the maximal ideal of \( V \) is generated by \( p \).

If \( V \) is a field, we set \( R = V[[X,Y]] \). If \( V \) is not a field, we set \( R = V[[X,Y]]/(p - (AX + BY)) \), where \( A, B \in V[[X,Y]] \) are elements such that \( \phi(p - (AX + BY)) = 0 \). In the latter case we change notation: Denote again by \( \phi \) the induced map from \( R \) onto \( \Lambda \), and by \( X \) and \( Y \) the cosets in \( R \) of \( X \) and \( Y \), respectively. Then, exactly as in the proof of Proposition 3.4, \( \phi : R \to \Lambda \) is a surjective homomorphism such that \( \phi(X) = x \) and \( \phi(Y) = y \), where \( R \) is a complete two-dimensional regular local ring and \( m_R = (X,Y) \) is the maximal ideal of \( R \), and of course \( \phi \) induces an isomorphism between the residue fields of \( R \) and \( \Lambda \). Note that (3.6.1) still holds since \( \Lambda \) has not changed.

Now select elements \( S, T \in R \) such that \( \phi(S) = s \) and \( \phi(T) = t \). We claim that \( Y^2 + SXY + TX^2 \) is an irreducible element in the unique factorization domain \( R \). Suppose, by way of contradiction, that \( Y^2 + SXY + TX^2 = CD \) in \( R \), where neither \( C \) nor \( D \) is a unit. We know that \( \mu_R(m_R^2) = 3 \) (by [AM, Theorem 11.22]), and \( m_R^2 = (X^2, XY, Y^2) \), so that \( CD = Y^2 + SXY + TX^2 \in m_R^2 - m_R^3 \). Note that \( \phi(C)\phi(D) = \phi(Y^2 + SXY + TX^2) = 0 \) by (3.6.1). But \( \phi(m_R) = m \), and both maximal ideals require two generators, from which we get a contradiction as follows. If \( C \) and \( D \) are linearly independent modulo \( m_R^2 \), then they generate \( m_R \), so that \( \phi(C) \) and \( \phi(D) \) are generators of \( m \) such that \( \phi(C)\phi(D) = 0 \), contrary to the hypotheses of the proposition. If \( C \) and \( D \) are linearly dependent modulo \( m_R^2 \), then they are associates, from which it follows that \( \phi(C)^2 = 0 \), again contrary to the hypotheses of the proposition. Therefore the claim holds.

We set \( \Omega = R/(Y^2 + SXY + TX^2) \). Since \( \phi(Y^2 + SXY + TX^2) = 0 \), the map \( \phi \) factors through \( \Omega \) to yield a surjective homomorphism from \( \Omega \) onto \( \Lambda \); moreover, this homomorphism induces an isomorphism of the residue fields of \( \Omega \) and \( \Lambda \). To complete the proof of the proposition, we need only show that \( \Omega \) is an unsplit Dedekind-like ring.

First we recall that \( R \) is a unique factorization domain, and \( Y^2 + SXY + TX^2 \) is irreducible, so that \( (Y^2 + SXY + TX^2) \) is prime, and therefore \( \Omega \) is a (complete local noetherian one-dimensional) domain. Since we need only concern ourselves with the ring \( \Omega \) for the remainder of this proposition, let us denote by \( x, y, s, \) and \( t \) the images in \( \Omega \) of the elements \( X, Y, S, \) and
\[ T, \text{ respectively, of } R. \text{ Let us also denote by } m \text{ the maximal ideal of } \Omega, \text{ and note that } m = (x, y). \text{ By the definition of } \Omega, \text{ we have } y^2 + sxy + tx^2 = 0; \text{ that is, (3.6.1) still holds. Therefore } (y/x)^2 + s(y/x) + t = 0 \text{ in the quotient field of } \Omega, \text{ and hence } y/x \text{ is integral over } \Omega \text{ in its quotient field. Let us set } \\
\Gamma = \Omega[y/x] = \Omega + \Omega \cdot y/x \text{ (using the fact that } (y/x)^2 = -s(y/x) - t). \]

Now \( \Gamma \) is a finite, integral extension of \( \Omega \) in its quotient field, so that \( \Gamma \) is a semilocal integral domain. We claim that \( m \) is an ideal of \( \Gamma \). It suffices to prove that \( m \) is closed under multiplication by \( y/x \). We have \( (y/x)^2 = -s(y/x) - t \). Let us set \( \Gamma = \Omega\lbrack y/x \rbrack = \Omega + \Omega \cdot y/x \) (using the fact that \( (y/x)^2 = -s(y/x) - t \)). \n
Next we claim that \( m = \text{rad}(\Gamma) \). For the inclusion \( (\subseteq) \) it suffices to prove that, for every element \( m \) of the ideal \( m \) of \( \Gamma \), \( 1 - m \) is invertible in \( \Gamma \); and this follows from the fact that \( m = \text{rad}(\Omega) \). The inclusion \( (\supseteq) \) holds since \( F = \Gamma/m \) is a field.

Let \( F = \Gamma/m \). We claim that \( F \) is a 2-dimensional field extension of \( k \).

Since \( m \) is an ideal of \( \Gamma \), \( F = \Gamma/m \) is a ring containing \( k = \Lambda/m \). Now, \( F \) is generated as a \( k \)-algebra by the image \( \alpha \) of \( y/x \) in \( F \). As proved below (3.6.1), the polynomial \( Z^2 + sZ + t \) is irreducible over \( k \). Since the \( \alpha \) is a zero of this polynomial, by (3.6.1), we see that \( F \) is a 2-dimensional field extension of \( k \), as claimed.

Next we claim that \( m = \text{rad}(\Gamma) \). For the inclusion \( (\subseteq) \) it suffices to prove that, for every element \( m \) of the ideal \( m \) of \( \Gamma \), \( 1 - m \) is invertible in \( \Gamma \); and this follows from the fact that \( m = \text{rad}(\Omega) \). The inclusion \( (\supseteq) \) holds since \( F = \Gamma/m \) is a field.

\( \Gamma \) is a local ring since \( \Gamma/\text{rad}(\Gamma) \) is a field. We have \( m = \Gamma(\Lambda x + \Lambda y) = \Gamma x \) because \( y = (y/x)x \in \Gamma x \). Since the maximal ideal \( m \) of the local domain \( \Gamma \) is principal, we see that \( \Gamma \) is a DVR. Thus \( \Gamma \) is integrally closed and therefore must be the normalization of \( \Omega \). Finally, we note that \( \mu_\Omega(\Gamma) = 2 \), so that, by Definition 2.5, \( \Omega \) is unsplit Dedekind-like.

We remark that, since \( \Gamma \) is finitely generated as an \( \Omega \)-module, \( \Gamma \) is complete in the \( m \)-adic topology as an \( \Omega \)-module [N, Theorem 17.8]. But since \( m \) is also an ideal of \( \Gamma \), this is the same as the \( m \)-adic topology on \( \Gamma \) as a ring. Therefore, \( (\Gamma, m, F) \) is a complete DVR.

Remark 3.2 and Propositions 3.4, 3.5, and 3.6 together show that, for a given complete, local ring \( (\Lambda, m, k) \), at least one of the conditions (i) or (ii) of Theorem 3.1 must hold. It remains to show that these two conditions are mutually exclusive, which we show in the final lemma and proposition of this section.

**Lemma 3.7.** Let \( (\Lambda, m, k) \) be a complete Dedekind-like ring, but not a DVR.

(i) If \( w \in m - m^2 \) then \( w^2 \notin m^3 \).

(ii) The (composition) length of the ring \( \Lambda/m^3 \) is 5.

**Proof.** Let \( \Gamma \) be the normalization of \( \Lambda \).

(i) The important idea here is that \( m = \text{rad}(\Gamma) \) and therefore both the hypothesis \( w \in m - m^2 \) and the conclusion \( w^2 \notin m^3 \) say the same thing
whether we consider \(m\) to be an ideal of \(\Lambda\) or of \(\Gamma\). So we work in the ring \(\Gamma\) instead of \(\Lambda\).

Case 1: \(\Lambda\) unsplit Dedekind-like. Then \(\Gamma\) is a DVR and hence the desired conclusion is obvious.

Case 2: \(\Lambda\) strictly split Dedekind-like. Here \(\Gamma = V_1 \times V_2\), the direct product of two DVRs and \(m = \text{rad}(\Gamma) = \text{rad}(V_1) \times \text{rad}(V_2)\); and so the desired conclusion follows easily from the fact that each \(V_i\) is a DVR.

(ii) The bottom row of conductor square (2.13.1) shows that \(\Gamma/\Lambda\) is a simple \(\Lambda\)-module. Therefore so is \((\Gamma/m^3)/(\Lambda/m^3)\). So it suffices to show that \(\Gamma/m^3\) has length 6 as a \(\Lambda\)-module.

Since \(\Gamma\) is either a DVR or the direct product of two DVRs, and \(m = \text{rad}(\Gamma)\), it suffices to show that \(\Gamma/m\) has length 2 as a \(\Lambda\)-module; that is, as a \(k\)-vector space. See (2.5.1). \(\Box\)

To show that conditions (i) and (ii) of Theorem 3.1 are mutually exclusive, clearly it suffices to show that Klein rings and complete, strictly split or unsplit Dedekind-like rings cannot be mapped onto either an artinian triad or a Drozd ring. This we show in the final proposition of this section, which completes the proof of the theorem.

**Proposition 3.8.** Let \((\Lambda, m, k)\) be either a Klein ring or a complete, strictly split or unsplit Dedekind-like ring. Then \(\Lambda\) does not have either an artinian triad or a Drozd ring as homomorphic image.

**Proof.** Since \(\mu_\Lambda(m) = 2\), clearly \(\Lambda\) cannot map onto an artinian triad. Similarly, if \(\Lambda\) is a Klein ring, then \(\mu_\Lambda(m^2) = 1\), and hence \(\Lambda\) cannot map onto a Drozd ring.

Therefore, we can suppose that \(\Lambda\) is a complete split or unsplit Dedekind-like ring, and we must show that \(\Lambda\) cannot map onto a Drozd ring. Suppose, by way of contradiction, that \(\phi : \Lambda \to A\) were a homomorphism from \(\Lambda\) onto the Drozd ring \(A\). Then \(\phi(m^3) = \phi(m)^3 = 0\) implies that \(\phi\) would induce a map from \(\Lambda/m^3\) onto \(A\). Since both \(\Lambda/m^3\) and \(A\) have composition length five (by Lemma 3.7 and (2.4.1)), we would obtain the isomorphism \(\Lambda/m^3 \cong A\). That is, \(\Lambda/m^3\) would be a Drozd ring.

But then \(\Lambda\) would contain an element \(w\) whose image \(\bar{w}\) in \(\Lambda/m^3\) satisfies \(\bar{w} \in \bar{m} - \bar{m}^2\) and \(\bar{w}^2 = 0\). Therefore \(w \in m - m^2\) and \(w^2 \in m^3\), contrary to Lemma 3.7. \(\Box\)

### 4. Wildness of Drozd rings.

The purpose of this section is to prove its final result, Theorem 4.9, establishing finite-length wildness of Drozd rings.

*Caution.* When functions represent matrix multiplication, we write the functions as right-hand operators. This occurs many times in this section.
**Notation 4.1.** Recall (see Definition 2.4) that a Drozd ring is an artinian local ring \((\Lambda, m, k)\) such that \(\mu_\Lambda(m) = \mu_\Lambda(m^2) = 2\), \(m^3 = 0\), and \((\exists x \in m - m^2)\) \(x^2 = 0\). Since \(\mu_\Lambda(m) = 2\) we have \(m = (x, y)\) for some \(y \in m - m^2\). Fix such elements \(x, y\) for the rest of this section. Then \(m^2 = (xy, y^2)\), and the only nonzero monomials in \(x\) and \(y\) are \(x, y, xy, y^2\).

In Drozd’s original paper [D], \(\Lambda\) is a 5-dimensional \(k\)-algebra, and \(1, x, y, xy, x^2\) form a basis of \(\Lambda\). In our more general setting we partially repair the lack of a basis by establishing the following *standard form* for elements of \(m\).

**Lemma 4.2.** For all elements \(c \in m\), we can express \(c = u_1x + u_2y + u_3xy + u_4y^2\) with each \(u_i\) a unit or 0. (Note that we do not claim uniqueness of the coefficients \(u_i\).

**Proof.** There is an expression \(c = ax + by\). We can write \(a = u_1 + a'x + b'y\) where \(u_1\) is a unit or 0. (If \(u_1\) is a unit, we can take \(a' = 0 = b'\), but this is not important.) Similarly, \(b = u_2 + a''x + b''y\). Then \(c = u_1x + u_2y + (b' + a'')xy + b''y^2\). Making similar substitutions for \(b' + a''\) and \(b''\) completes the proof. \(\square\)

**Viewing** \(\Lambda/(y)\) and \(\Lambda/(xy)\) as \(\Lambda\)-modules, for \(\lambda \in \Lambda\) we denote by \(\lambda y\) (respectively \(\lambda xy\)) the coset of \(\lambda\) in \(\Lambda/(y)\) (respectively \(\Lambda/(xy)\)). Thus, \(\Lambda/(y) = \Lambda_1y\) has submodule \(\Lambda xy\), etc. The following lemma establishes some basic facts about \(\Lambda\) and its homomorphic images \(\Lambda_1y\) and \(\Lambda_1xy\).

**Lemma 4.3.**

(i) *All of the monomials* \(x, xy, x^2\), \(\ldots\) in diagram (4.3.1) *below are nonzero.*

(ii) \(\text{soc } \Lambda_1y = \Lambda xy\), \(\text{soc } \Lambda_1xy = \Lambda xy \oplus \Lambda y^2\), and \(\text{soc } \Lambda = \Lambda xy \oplus \Lambda y^2 = m^2\).

(iii) \(\Lambda x \cap \Lambda y = \Lambda xy\) and \(\Lambda x_{xy} \cap \Lambda y_{xy} = 0\).

(iv) \(\Lambda x \cong \Lambda y_{xy}\) as \(\Lambda\)-modules via the multiplication map \(cx \to cy_{xy}\).

(iii)

\[
\begin{array}{ccc}
\Lambda_1y & \Lambda_1xy & \Lambda
\\
\Lambda xy & \Lambda x_{xy} & \Lambda y_{xy}
\end{array}
\]

Proved. \(\Lambda.\) Since \(m^2 = (xy, y^2)\) and \(\mu_\Lambda(m^2) = 2\), the monomials displayed in \(\Lambda\) are nonzero. To prove the nontrivial inclusion in \(\Lambda x \cap \Lambda y = \Lambda xy\), assume that \(cx = dy\). The element \(c\) cannot be a unit, since this would imply \(\mu_\Lambda(x, y) = 1\). Therefore we can write \(c\) in the standard form of Lemma 4.2, getting \(cx = u_2xy\), as desired.
Next we verify the description of the socle of \( \Lambda \). Note that, for any \( \Lambda \)-module \( M \), we have \( \mathfrak{m}M = 0 \) if and only if \( M \) is (a \( k \)-module hence) a semisimple \( \Lambda \)-module. Therefore \( \mathfrak{m}^3 = 0 \) implies that \( \mathfrak{m}^2 \subseteq \text{soc} \Lambda \). Conversely, take \( c \in \text{soc} \Lambda \) and write \( c \) in the standard form of Lemma 4.2.

Then \( 0 = cx = u_2xy \) shows that \( u_2 \) is not a unit, and hence \( u_2 = 0 \). But then \( 0 = cy = u_1xy \), and so \( u_1 = 0 \), completing the proof that \( \mathfrak{m}^2 = \text{soc} \Lambda = \Lambda xy + \Lambda y^2 \). The sum \( \Lambda xy + \Lambda y^2 \) must be direct since \( \mu_\Lambda(\mathfrak{m}^2) = 2 \). This completes the proof of the assertions about \( \Lambda \).

\( \Lambda_1xy \). The second intersection in (iii) holds since \( \Lambda x \cap \Lambda y = \Lambda xy \). The monomial \( x_{xy} \) is nonzero since otherwise \( x \in \Lambda xy \) contradicting \( \mu_\Lambda(x, y) = 2 \). Similarly \( y_{xy} \neq 0 \); and \( y_{xy}^2 \neq 0 \) because \( \mu_\Lambda(xy, y^2) = 2 \).

We have \( \text{soc} \Lambda_1xy \supseteq \Lambda x_{xy} + \Lambda y_{xy}^2 \) because the right-hand side is annihilated by \( x \) and \( y \). For the opposite inclusion take \( c \in \Lambda \) such that \( c_{xy} \in \text{soc} \Lambda_1xy \). Since \( c_{xy} \mathfrak{m} = 0 \) we have \( \mathfrak{m} \subseteq \Lambda xy \). Now write \( c \) in the standard form of Lemma 4.2. Then \( cy = u_1xy + u_2y^2 \in \Lambda xy \). Since \( cy \in \Lambda xy \), directness of the sum \( \Lambda xy \oplus \Lambda y^2 \) shows that \( u_2 \) cannot be a unit, hence equals zero. Therefore \( c = u_1x + u_3xy + u_4y^2 \), showing that \( c_{xy} \in \Lambda x_{xy} + \Lambda y_{xy}^2 \). It remains to prove directness of this last sum; and this follows from \( \Lambda x_{xy} \cap \Lambda y_{xy} = 0 \), completing the proof of all assertions about \( \Lambda_1xy \).

\( \Lambda_1y \). All assertions here are obvious.

(iv). It suffices to prove that \( c y_{xy} = 0 \iff cx = 0 \). Since \( c \) is not a unit, we can write it in the standard form of Lemma 4.2. Then \( cy_{xy} = 0 \) is equivalent to \( u_1xy + u_2y^2 \in \Lambda xy \). Directness of the sum in our expression for \( \text{soc} \Lambda \) shows that \( u_2 \) cannot be a unit, hence equals 0. Therefore \( c y_{xy} = 0 \) implies that \( c = u_1x + u_3xy + u_4y^2 \), and therefore \( cx = 0 \). The converse is proved similarly.

\( \Box \)

**Definition 4.4** (\( S\left(1 \frac{1}{2}\right), S(2) \)). We define an object of the category \( S\left(1 \frac{1}{2}\right) \), or more completely, \( k\)-\( S\left(1 \frac{1}{2}\right) \) (“one and one-half \( k \)-similarity”) to be an ordered triple \( (m, n, \phi) \) where \( m \leq n \) are positive integers and \( \phi \) is an \( n \times n \) matrix over \( k \). We define a morphism of \( S\left(1 \frac{1}{2}\right) \) to be a pair of \( k \)-linear maps \( (\sigma, \tau) \) (equivalently, matrices of the appropriate sizes) such that the following diagram commutes.

\[
\begin{array}{ccc}
k^{(m)} & \overset{i}{\longrightarrow} & k^{(n)} \\
\sigma \downarrow & & \tau \downarrow \\
k^{(m')} & \overset{i'}{\longrightarrow} & k^{(n')}
\end{array}
\]

(4.4.1)

Here \( k^{(m)} \) denotes the direct sum of \( m \) copies of \( k \), written as row vectors, and \( i = i_{m,n} \) denotes the inclusion map \( \mathbf{b} \rightarrow (\mathbf{b}, \mathbf{0}) \), where \( \mathbf{b} \) denotes an arbitrary vector in \( k^{(m)} \) and \( \mathbf{0} \) denotes the zero vector in \( k^{(n-m)} \). The map \( i' \) is defined analogously.
Similarly, we define the category $S(2)$ (more completely, $k$-$S(2)$) to be “simultaneous similarity” of a pair of matrices over the field $k$. That is, the objects of $S(2)$ are ordered triples $(n, A, B)$ where $n$ is a positive integer and $A$ and $B$ are $n \times n$ matrices over $k$. We define a morphism of $S(2)$ to be a $k$-linear map (equivalently, matrix of the appropriate size) $\tau$ such that the following diagram commutes.

$$
\begin{array}{ccc}
 k^n & \xrightarrow{A} & k^n \\
 \downarrow{\tau} & & \downarrow{\tau} \\
 k^{n'} & \xrightarrow{A'} & k^{n'}
\end{array}
\quad \begin{array}{ccc}
 k^n & \xrightarrow{B} & k^n \\
 \downarrow{\tau} & & \downarrow{\tau} \\
 k^{n'} & \xrightarrow{B'} & k^{n'}
\end{array}
$$

(4.4.2) $S(2)$:

For the connection between these categories and wildness, first recall [Remarks 2.3(i)] that $\text{fdmod-} k \langle X, Y \rangle$ contains a copy of $\text{fdmod-} A$ for every finitely generated $k$-algebra $A$. The remaining (known) facts that we need are contained in the following lemma.

**Lemma 4.5.** Let $k$ be a field. Then:

(i) $S(2)$ is equivalent to the category $\text{fdmod-} k \langle X, Y \rangle$ of finite dimensional right $k \langle X, Y \rangle$-modules.

(ii) $S(1\frac{1}{2})$ contains a full subcategory that is equivalent to $S(2)$.

(iii) Let $k$ be a residue field of a commutative ring $\Lambda$. Suppose that there is full subcategory $W$ of the category of $\Lambda$-modules of finite length and an additive functor $\Phi: W \rightarrow S(1\frac{1}{2})$ that is a representation equivalence. Then $\Lambda$ is finite-length wild.

**Proof.** (i) It is easy to see that the category $S(2)$ is equivalent to the category $\text{fdmod-} k \langle X, Y \rangle$ of finite dimensional right $k \langle X, Y \rangle$-modules by sending the object $(n, A, B)$ to the $k \langle X, Y \rangle$-module consisting of the vector space $k^n$, on which $X$ and $Y$ act via right multiplication by the matrices $A$ and $B$, respectively.

(ii) This is due to Narazova [Nz, Lemma 1], except for a more functorial treatment here, and some changes and reversals of notation. For completeness, we include a proof.

Let $W'$ be the full subcategory of $S(1\frac{1}{2})$ consisting of objects of the form $(n, 2n, \phi)$ in which $\phi = \begin{bmatrix} 0 & I \\ A & B \end{bmatrix}$, where $A, B$ are $n \times n$ matrices over $k$ and $0, I$ are the $n \times n$ zero and identity matrices, respectively. Then let $F(n, 2n, \phi) = (n, A, B) \in S(2)$.

Let $(\sigma, \tau): (n, 2n, \phi) \rightarrow (n', 2n', \pi')$ be a morphism in $W'$. We claim that $\tau$ has the $2 \times 2$ block-diagonal form $\tau = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix}$. To see this, let $\tau = \begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix}$. 

The relation $i \tau = \sigma i'$ (as right operators) yields the equation

$$[I \ 0] \begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix} = [\sigma \ 0]$$

which shows that $T_1 = \sigma$ and $T_2 = 0$. Then the relation $\tau \phi' = \phi \tau$ yields the equation

$$\begin{bmatrix} T_1 & 0 \\ T_3 & T_4 \end{bmatrix} \begin{bmatrix} 0 & I \\ A' & B' \end{bmatrix} = \begin{bmatrix} 0 & I \\ A & B \end{bmatrix} \begin{bmatrix} T_1 & 0 \\ T_3 & T_4 \end{bmatrix}$$

which easily completes the proof of the claim. Moreover, the relation $\tau \phi' = \phi \tau$ now implies that $\sigma A' = A \sigma$ and $\sigma B' = B \sigma$, and hence these left and right multiplications by $\sigma$ form a morphism $F(\sigma, \tau) = \sigma$ in $\mathcal{S}(2)$.

The block diagonal form of $\tau$ together with the relations in the previous paragraph then show that $(\sigma, \tau)$ is an isomorphism if and only if $\sigma$ is an invertible matrix if and only if $F(\sigma, \tau) = \sigma$ is an isomorphism in $\mathcal{S}(2)$. It is now easy to check that $F$ yields an equivalence between $\mathcal{W}'$ and $\mathcal{S}(2)$.

(iii) In view of statements (i) and (ii), this is an immediate consequence of the definition [2.2] of finite-length wildness.

\[\square\]

**Definition 4.6** (Ringel’s $\Lambda$-module $M(m, n, \phi)$). Let $(\Lambda, m, k)$ be a Drozd ring; given an object $(m, n, \phi) \in \mathcal{S}(1\frac{1}{2})$, we define a module $M(m, n, \phi)$ as follows. First, let $Q$ be the $\Lambda$-module $Q = \Lambda^m \oplus \Lambda^n \oplus \Lambda^m$ (notation as in (4.3.1)), where $\Lambda^m$ denotes the direct sum of $m$ copies of the $\Lambda$-module $\Lambda$, written as row vectors. Next, define three maps $i$, mult, and $\phi$ between pairs of submodules of $Q$, as follows. [See diagram (4.6.1).]

The map $i: \Lambda^m x_y \to \Lambda^n y_{xy}$. Since the socle of every $\Lambda$-module is annihilated by $m$, we have $\Lambda^m x_y = k^m$; more precisely, this is a canonical $\Lambda$-module isomorphism via $b x_y \to b$ (where $b$ denotes the vector in $k^m$ formed by reducing each entry of $b$ modulo $m$). Similarly $\Lambda^n y_{xy}^2 = k^n$ via $c y_{xy} \to c$. We define $i = i_{m,n}$ by $(b x_y)i = (b, 0)y_{xy}^2$, the same map $i$ as in (4.4.1).

The map mult: $\Lambda^n x \to \Lambda^n y_{xy}$. This is by means of the multiplication map $d x \to d y_{xy}$. By Lemma 4.3(iv) this is indeed an $\Lambda$-module isomorphism.

The map $\phi: \Lambda^n x_{xy} \to \Lambda^n y^2$. We have $\Lambda^n x_{xy} = k^n$ via $c x_{xy} \to c$. Therefore we can right-multiply elements of $\Lambda^n x_{xy} = k^n$ by the matrix $\phi$ in our given object $(m, n, \phi)$ of $\mathcal{S}(1\frac{1}{2})$. Also, $\Lambda^n y^2 = k^n$ via $c y^2 \to c$. For $c \in \Lambda^n$ define $(c x_{xy})\phi = (c \phi)y^2$.

These three maps are displayed in the following diagram.
Finally, we define the $\Lambda$-module $M = M(m, n, \phi)$ to be $Q/R$ where $R$ denotes the submodule formed by the following set of relations, that is, elements of $Q$.

\[(4.6.2) \quad (bx_y, -(bi)y^2_{xy}, 0) + (0, -dy_{xy}, dx) + (0, cx_{xy}, -(\bar{c}\phi)y^2)\]

where $b, c, d$ range over all elements of $\Lambda(m), \Lambda(n), \Lambda(n)$ respectively. (Note that we can replace $b, c, d$ by $\bar{b}, \bar{c}, \bar{d}$ respectively, whenever convenient, but we cannot replace $\bar{c}$ by $c$ in $\bar{c}\phi$.)

Informally, we form $M$ from $Q$ by identifying elements of $Q$ with their image under the above “amalgamation” maps $i, \text{mult},$ and $\phi$. (It is possible that $\phi$ have a nontrivial kernel, in which case it is not a true amalgamation map.)

For every element $(m, n, \phi) \in S(1 \frac{1}{2})$ we have natural $\Lambda$-module surjections:

\[(4.6.3) \quad \rho: P = \Lambda(m) \oplus \Lambda(n) \oplus \Lambda(n) \xrightarrow{\rho_1} Q = \Lambda(m)1_y \oplus \Lambda(n)1_{xy} \oplus \Lambda(n) \xrightarrow{\rho_2} M(m, n, \phi)\]

where $(b, c, d)\rho_1 = (b_y, c_{xy}, d)$ and $\rho_2$ equals reduction modulo the relations $R$ displayed in (4.6.2). Note that $\ker \rho \subseteq mP$, and therefore:

\[(4.6.4) \quad \text{The map } \rho \text{ in (4.6.3) is a projective cover.}\]

**Historical Remark.** In Drozd’s original paper [D], $\Lambda$ is a $k$-algebra and the modules he uses in place of our $M$ are represented by very large matrices $(32n \times 32n$, in the notation of 4.4) over $k$. He assumes that $k$ is an algebraically closed field. But, in order to obtain wildness of $\Lambda$, the only property of algebraic closure that he uses is that $k$ has at least 5 elements. (Algebraic closure is, however, used elsewhere in his paper.) Ringel [R, (3.4)] uses much smaller matrices over $k$, by relating $M$ to $S(1 \frac{1}{2})$ rather than directly to $S(2)$, as Drozd did. The size of Ringel’s matrices is $(m+6n) \times (m+6n)$, in the notation of 4.4. Wildness of the category of $\Lambda$-modules that he obtains holds for any field $k$. 
Our module $M = M(m, n, \phi)$ is a modification of Ringel’s construction, and follows him by relating $\Lambda$-modules to $S(1^{2})$ rather than directly to $S(2)$. However, since our ring $\Lambda$ need not be a $k$-algebra, $M$ need not have a matrix representation. But we do represent homomorphisms in the category of modules $M(m, n, \phi)$ by matrices over the ring $\Lambda$. The critical fact is that any map of $\Lambda$-modules lifts to a map of their projective covers; and since projective modules over commutative local rings are free, this lifted map can be represented by a matrix over $\Lambda$. Our $\Lambda$-matrices are much smaller than Ringel’s $k$-matrices. But this is deceptive because an element of $\Lambda$ holds more information than a single element of $k$. There are two consequences of this: (i) The functor that we obtain, displaying wildness, is only a representation equivalence rather than the full exact embedding that Ringel gets. (ii) Our wildness proof is probably more intricate but shorter than Ringel’s (omitted) proof.

**Morphisms and Matrices.** Let $f: M = M(m, n, \phi) \rightarrow M' = M(m', n', \phi')$ be any $\Lambda$-module homomorphism. We wish to represent $f$ by a matrix $F$ over $\Lambda$. Since $P$ is a projective $\Lambda$-module, $f$ lifts to a $\Lambda$-homomorphism such that the following diagram commutes.

$$
\begin{array}{ccc}
P & \xrightarrow{F} & P' \\
\downarrow{\rho} & & \downarrow{\rho'} \\
M & \xrightarrow{f} & M'.
\end{array}
$$

Any such $F$ equals right multiplication by a unique $(m + 2n) \times (m' + 2n')$ matrix over $\Lambda$, and we again call this matrix $F$.

**Block form.** We always view $F$ in $3 \times 3$ block form, where the rows are partitioned into three blocks of lengths $m$, $n$, and $n$, respectively, and the columns are partitioned into three blocks of lengths $m'$, $n'$, and $n'$, respectively.

The crux of our wildness proof is the next lemma, which gives some information about the form of the matrix $F$.

**Notation 4.7.** We denote by $\text{ent } e$ the set of entries of any array $e$.

**Lemma 4.8.**

(i) The matrix $F$ has the block form

$$F = \begin{bmatrix} F_{11} & m & (x, y^2) \\ * & F_{22} & m \\ * & * & F_{33} \end{bmatrix}. \tag{4.8.1}$$

That is, $\text{ent } F_{12} \subseteq m$, $\text{ent } F_{23} \subseteq m$, and $\text{ent } F_{13} \subseteq (x, y^2)$. Moreover, when all entries are reduced modulo $m$, the diagonal blocks satisfy the following three properties (viewed as right operators).
(a) $F_{11}i_{m',n'} = i_{m,n}F_{22}$, where $i_{m',n'}$ and $i_{m,n}$ are as in (4.4.1).

(b) $F_{22} = F_{33}$.

(c) $\phi F_{22} = F_{22}\phi'$.

(We make no assertion about the blocks denoted by $\ast$.)

(ii) If $f$ is an isomorphism, then the diagonal blocks $F_{ii}$ are invertible matrices.

Proof. Let $b,c,d$ be elements of a local ring with residue field $k$ (e.g., $\Lambda$ or its homomorphic images).

(4.8.2) If $bd = cd$ and $d \neq 0$ then $b = c$ (natural images in $k$) because $b \equiv c$ modulo the annihilator of $d$, and this annihilator is contained in the maximal ideal of the ring. We use this fact many times in the long, intricate proof below.

(i) Since $F$ lifts $f$ we must have $(\ker \rho)F \subseteq \ker \rho'$. Recall that we factor $\rho = \rho_1 \rho_2$ as in (4.6.3). The information that we obtain about $F$ is information that can be obtained by dealing separately with five types of elements of $\ker \rho$. Type 1: Elements $(e,0,0) \in P$ such that $(e,0,0)\rho_1 = 0 \in Q$; type 2: Elements $(0,e,0) \in P$ such that $(0,e,0)\rho_1 = 0$ in $Q$; types 3–5: Elements of $P$ corresponding to the three types of amalgamations that were used to form $M$, one type for each term in (4.6.2). We label the five parts of the proof by the blocks of $F$ about which they give information.

$F_{12}, F_{13}$. Note that $(ey,0,0)\rho_1 = 0$ for every $e$ because $y_y = 0$. Therefore $(ey,0,0)F\rho' = 0$ and hence $(ey,0,0)F\rho_1' \subseteq \ker \rho_2'$. Writing $F$ in block form and using the definition of $\rho_1'$, it follows that

(4.8.3) $(0,eF_{12}y_y, eF_{13}y) \in \ker \rho_2'$

for every $e$. Therefore every element of this form must have an expression of the form (4.6.2) (more precisely, the version of (4.6.2) that applies to $M'$, hence contains $i'$ and $\phi'$ in place of $i$ and $\phi$).

Comparing coordinate 1 in (4.6.2) and (4.8.3) yields $\text{ent} b \subseteq m$. Therefore the first of the three terms in (4.6.2) equals zero, and can be ignored. Looking at coordinate 3 shows that

(4.8.4) $eF_{13}y = dx - (e\phi)y^2$.

Therefore $\text{ent} d \subseteq m$ (otherwise $x \in \Lambda y$ and $\mu_\Lambda(m) \leq 1$). On the other hand, looking at coordinate 2 shows that $eF_{12}y_y = -dy_y + cy_{xy}$. But $\Lambda x_{xy} \cap \Lambda y_y = 0$ by (4.3.1). Therefore $\text{ent} c \subseteq m$. A second look at coordinate 2 therefore shows that $\text{ent}(eF_{12}y_y) \subseteq m^2$ for every $e$. Hence $\text{ent} F_{12} \subseteq m$ as desired.

Since $\text{ent} c \subseteq m$, we have $\overline{c} = 0$. But $\text{ent} d \subseteq m$ also, so (4.8.4) now yields

$$\text{ent} eF_{13}y \subseteq \text{ent} dx \subseteq mx = \Lambda xy$$
for every \( \mathbf{e} \). Let \( c \) be any element of \( \text{ent} F_{13} \). Then \( cy \in \Lambda xy \), so that \( c \in m \).

Write \( c \) in the standard form of Lemma 4.2. Then

\[
cy = u_1 xy + u_2 y^2 \in \Lambda xy.
\]

Directness of the sum \( \Lambda xy + \Lambda y^2 \) then shows that \( u_2 \) is not a unit, and hence \( u_2 = 0 \). Therefore \( c \in (x, y^2) \) and hence \( \text{ent} F_{13} \subseteq (x, y^2) \) as desired.

\( F_{23} \). Note that \( (0, 0) \rho_1 = 0 \) for every \( \mathbf{e} \) because \( (xy)_{xy} = 0 \). Therefore, as before, \( (0, 0) F\rho'_1 \subseteq \ker \rho'_2 \) and hence

\[
(4.8.5) \quad (0, 0, eF_{23}xy) \in \ker \rho'_2
\]

for every \( \mathbf{e} \). As before, this triple must have the form \( (4.6.2) \). Looking at coordinate 1 in both triples shows that \( \text{ent} b \subseteq m \), hence term 1 of \( (4.6.2) \) equals zero and we can ignore it. Coordinate 2 then yields \(-dy_{xy} + cx_{xy} = 0\).

Since \( \Lambda x_{xy} \cap \Lambda y_{xy} = 0 \) we have \( dy_{xy} = 0 \). Therefore \( (4.8.6) \) yields \( cx_{xy} = 0 \) and \( dy_{xy} = 0 \).

In particular, \( \text{ent} c \subseteq m \) and \( \text{ent} d \subseteq m \). Write an arbitrary entry \( d \) of \( d \) in standard form \( d = u_1 x + u_2 y + u_3 xy + u_4 y^2 \). Then \( (4.8.6) \) yields \( 0 = dy_{xy} = u_2 y^2 \). Therefore \( u_2 \) is not a unit, and hence \( u_2 = 0 \). We conclude that \( d \in (x, y^2) \). That is, \( \text{ent} d \subseteq (x, y^2) \), from which it follows that \( dx = 0 \). Since \( \text{ent} c \subseteq m \) we have \( \tau = 0 \) and hence coordinate 3 of \( (4.6.2) \) equals 0. Therefore \( (4.8.5) \) yields \( eF_{23}xy = 0 \) for every \( \mathbf{e} \). This, in turn, implies that \( \text{ent} F_{23} \subseteq m \), as desired.

\( F_{11} i_{m', n'} = i_{m, n} F_{22} \). The first amalgamation term in \( (4.6.2) \) is carried by \( \rho \) to zero: \( (ex, -(eim_n)n^2, 0) \rho = 0 \) for every \( \mathbf{e} \). Therefore

\[
(ex, -(eim_n)n^2, 0) F\rho'_1 \subseteq \ker \rho'_2.
\]

Writing \( F \) in block form and using the definition of \( \rho'_2 \) then shows

\[
(eF_{11}xy - (eim_n)F_{21}y^2, eF_{12}xy - (eim_n)F_{22}y^2_{xy}, eF_{13}x - (eim_n)F_{23}y^2) \in \ker \rho'_2.
\]

Since \( y^2 = 0 \), \( F_{12}xy \in mx_{xy} = 0 \), \( \text{ent} F_{13} \subseteq (x, y^2) \), and \( \text{ent} F_{23} \subseteq m \) this simplifies to

\[
(4.8.7) \quad (eF_{11}xy, -(eim_n)F_{22}y^2_{xy}, 0) \in \ker \rho'_2.
\]

We have \( \Lambda x \cap \Lambda y^2 = \Lambda x \cap \Lambda y \cap \Lambda y^2 = \Lambda xy \cap \Lambda y^2 = 0 \) by (4.3.1). Comparing coordinate 3 of (4.8.7) with coordinate 3 of \( (4.6.2) \) yields \( dx - (\tau \phi)y^2 = 0 \); therefore \( dx = 0 \), and hence \( \text{ent} d \subseteq m \). Then writing a typical entry of \( d \) in standard form of Lemma 4.2, together with the relation \( dx = 0 \), shows that \( \text{ent} d \subseteq (x, y^2) \). Therefore term 2 in \( (4.6.2) \) equals zero.

Coordinate 2 of \( (4.8.7) \) and \( (4.6.2) \) now yields

\[
-(eim_n)F_{22}y^2_{xy} = -(b_{m', n'})y^2_{xy} + cx_{xy}.
\]
Then \( \Lambda x_{xy} \cap \Lambda y_{xy}^2 = 0 \) yields
\[
(4.8.8) \quad (e_{i,m,n}) F_{22} y_{xy}^2 = (b_{i,m',n'}) y_{xy}^2.
\]

Coordinate 1 yields \( e F_{11} x_y = b x_y \). Then (4.8.2) yields \( \overline{F}_{11} = \overline{b} \) and hence \( \overline{F}_{11} i m', n' = \overline{b} i m', n' \). But (4.8.8) yields \( (e_{i,m,n}) \overline{F}_{22} = \overline{b} i m', n' \) (also by (4.8.2)). We conclude that \( \overline{F}_{11} i m', n' = (e_{i,m,n}) \overline{F}_{22} \) for every \( e \), and therefore \( \overline{F}_{11} i m', n' = i_{m,n} F_{22} \), proving (a).

\( \overline{F}_{22} = \overline{F}_{33} \). We use the second amalgamation term in (4.6.2): \( (0, -e_y, e x) \rho = 0 \) and therefore \( (0, -e_y, e x) F \rho'_1 \subseteq \ker \rho'_2 \) for every \( e \). Writing \( F \) in block form and applying the definition of \( \rho' \) yields
\[
(4.8.9) \quad (\ldots, -e F_{22} x_y + e F_{32} x_y, -e F_{23} y + e F_{33} x) \in \ker \rho'_2.
\]

Comparing coordinate 2 in this and (4.6.2) and multiplying by \( y \) yields \( e F_{22} y_{xy}^2 = d y_{xy} \). Then (4.8.2) yields \( \overline{F}_{22} = \overline{d} \).

Comparing coordinate 3 and multiplying by \( y \) yields \( -e F_{23} y + e F_{33} x y = \overline{d} x y \). Since \( \Lambda x \cap \Lambda y^2 = 0 \), this yields \( \overline{F}_{33} = \overline{d} \) (by (4.8.2)). Therefore \( \overline{F}_{22} = \overline{e} \overline{F}_{33} \) for every \( e \), and hence \( \overline{F}_{22} = \overline{F}_{33} \), proving (b).

\( \phi \overline{F}_{22} = \overline{F}_{22} \phi' \). Use the third amalgamation term in (4.6.2): \( (0, e x, - (\overline{e} \phi, y^2) \rho = 0 \) and hence \( (0, e x, - (\overline{e} \phi, y^2) \rho = 0 \) for every \( e \). As before this yields
\[
(4.8.10) \quad (\ldots, e F_{22} x y - (\overline{e} \phi) 32 y_{xy}^2, e F_{23} x - (\overline{e} \phi) 33 y_{xy}^2) \subseteq \ker \rho'_2
\]

and this therefore has the form (4.6.2) for all \( e \). Comparing coordinate 2 in these expressions yields
\[
(4.8.11) \quad e F_{22} x y - (\overline{e} \phi) 32 y_{xy}^2 = -(b') y_{xy}^2 - d y_{xy} + c x y.
\]

Since \( \Lambda x \cap \Lambda y^2 = 0 \), comparing the \( x_{xy} \)-terms on both sides yields \( e F_{22} x y = c x y \), and therefore \( \overline{F}_{22} = \overline{c} \) (by (4.8.2)). We conclude that
\[
(4.8.12) \quad \overline{e} \overline{F}_{22} \phi' = \overline{c} \phi'
\]

for all \( e \).

Comparing coordinate 3 in (4.8.10) and (4.6.2) yields
\[
(4.8.13) \quad e F_{23} x - (\overline{e} \phi) 33 y^2 = d x - (\overline{e} \phi') y^2.
\]

As already observed, \( \Lambda x \cap \Lambda y^2 = 0 \). Therefore \( (\overline{e} \phi) 33 y^2 = (\overline{e} \phi') y^2 \), so that by (4.8.2) we get \( (\overline{e} \phi) \overline{F}_{33} = \overline{e} \phi' \). Comparing this with (4.8.12) yields \( (\overline{e} \phi) \overline{F}_{33} = \overline{e} \overline{F}_{22} \phi' \) for every \( e \), and hence \( \phi \overline{F}_{33} = \overline{F}_{22} \phi' \). Since \( \overline{F}_{33} = \overline{F}_{22} \), statement (c) follows, and the proof of part (i) of the lemma is now complete.

(ii) We assume that \( f \) is an isomorphism. Since \( \rho \) and \( \rho' \) are projective covers (see (4.6.4)), the map \( F \) is again an isomorphism, and therefore the matrix \( F \) is invertible. Hence \( \overline{F} \) is invertible. But \( \overline{F} \) is a block-triangular
matrix over the field $k$, by part (i) of this proof, and hence each diagonal block $F_{ii}$ is invertible. Therefore $F_{ii}$ is also invertible. □

**Theorem 4.9** (Wildness of Drozd Rings). *Every Drozd ring is finite-length wild.*

**Proof.** Let the Drozd ring be $(\Lambda, m, k)$. Since the $\Lambda$-modules $M(m, n, \phi)$ have finite length, by Lemma 4.5(iii) it suffices to show:

(4.9.1) Let $W$ the full subcategory of mod $\Lambda$ consisting of all modules of the form $M(m, n, \phi)$. Then there is an additive functor $E: W \to S(1^1_2)$ that is a representation equivalence.

For each $M = M(m, n, \phi) \in W$ fix a pair of surjective homomorphisms $\rho_1: P = \Lambda^m(\oplus) \Lambda^n(\oplus) \Gamma(n) \to Q$ and $\rho_2: Q \to M$, and let $\rho$ be their composition as in (4.6.3). In particular, $\text{ker}(\rho_2)$ is given by (4.6.2), and the matrix $\phi$ acts as shown in (4.6.1). These choices remain in effect throughout this proof, and our definition of $E$ depends on them. In particular, set $E(M) = (m, n, \phi) \in S(1^1_2)$.

Next let $f: M \to M'$ be a $\Lambda$-homomorphism, where $M = M(m, n, \phi)$ and $M' = M(m', n', \phi')$. We wish to define the morphism $E(f) = (\sigma, \tau)$; that is, we wish to define $\sigma$ and $\tau$ in diagram (4.4.1).

Recall that $\rho: P \to M$ and $\rho': P' \to M'$ are projective covers. Therefore $f$ lifts to a homomorphism $F: P \to P'$ such that diagram (4.6.5) commutes. The matrix of $F$ (again called $F$) has a $3 \times 3$ block upper triangular form, as displayed in (4.8.1). Then we define

(4.9.2) $E(f) = (F_{11}, F_{22})$.

We proceed to prove that $E$ has the required properties.

First we claim that $E(f)$ is well-defined; that is, choosing a different lifting $F$ of $f$ does not change $E(f)$. It suffices to show that if $f = 0$ then $F_{11}$ and $F_{22}$ equal 0. But $f = 0$ implies that $F(P) \subseteq \text{ker} \rho' \subseteq mP'$, and hence all entries of the matrix $F$ are elements of $m$. Hence the entire matrix $F$ equals 0, proving the claim.

Next, we claim that $E(f)$ is a morphism in $S(1^1_2)$, that is, that diagram (4.4.1) commutes. For that, we must show that, for every $b \in \Lambda^m$ and $c \in \Lambda^n$, both $bF_{22} = bF_{11}'$ and $cF_{22} = cF_{22}'$. But these two equations follow immediately from parts (i)(a) and (i)(c) of Lemma 4.8.

Since $E$ is well-defined on morphisms, it is now easy to complete the proof that $E$ is an additive functor. For example the identity map on a module can be lifted to the identity map on its projective cover and then again becomes an identity map modulo $m$, showing that $E(1) = 1$. Similarly $E$ preserves compositions and sums of morphisms.

We now check the three defining properties of a representation equivalence.
Dense, that is, $E$ maps onto all isomorphism classes.

This is clear from the definition of $W$ as the full subcategory of $\text{mod} \Lambda$ consisting of all modules of the form $M(m, n, \phi)$.

Full, that is, $E$ is a surjection on hom groups.

Let $(\sigma, \tau)$ be a morphism in $S(1_2)$, as displayed in (4.4.1). Each of $\sigma, \tau$ equals right multiplication by a $k$-matrix that we again call $\sigma, \tau$ respectively. Let $F$ be a block diagonal matrix over $\Lambda$ such that $F_{11} = \sigma, F_{22} = \tau, F_{33} = F_{22}$, and all other entries of $F$ are zero. Right multiplication by $F$ is a homomorphism of free $\Lambda$-modules $P \to P'$. To show that $F$ induces an $\Lambda$-homomorphism $M \to M'$ it suffices to check that $(\ker \rho) F \subseteq \ker \rho'$. The block diagonal form of $F$ shows that $F$ induces an $\Lambda$-module homomorphism $F' : Q = P \rho_1 \to Q' = P' \rho'_1$. Thus it now suffices to check that the induced map $F'$ induces a homomorphism $M \to M'$. For this, it suffices to check that $F'$ takes each of the three terms of (4.6.2) to terms of the same form with respect to $M'$. First we have

$$(bx_y, -(bi)y_{xy}, 0)(F_{11} \oplus F_{22} \oplus F_{33}) = (bF_{11}x_y, -(bi)F_{22}y_{xy}, 0).$$

Since we are working in socles of $\Lambda$-modules, we can replace $F_{11}$ and $F_{22}$ by their images mod $m$. Moreover, since $(F_{11}, F_{22}) = (\sigma, \tau)$ is a morphism in $S(1_2)$, we have $iF_{22} = F_{11}i'$, and this shows that the form of $(bx_y, -(bi)y_{xy}, 0)$ is preserved.

The form of the second term in (4.6.2) is preserved because $F_{22} = F_{33}$. For the third term we have

$$(0, cx_{xy}, -(c\phi)y^2)(F_{11} \oplus F_{22} \oplus F_{33}) = (0, cF_{22}x_{xy}, -(c\phi)F_{33}y^2).$$

Since $x_{xy}$ and $y^2$ are in the socles of their respective modules, we can replace each of $F_{22}$ and $F_{33}$, on the right-hand side, by their respective images (both $\tau$) mod $m$. Since $\phi \tau = \tau \phi'$ it is clear that the form of $(0, cx_{xy}, -(c\phi)y^2)$ has been preserved.

Faithful, that is, $E(M) \cong E(M')$ if and only if $M \cong M'$.

Since $E$ is a functor, clearly $E(M) \cong E(M')$ if $M \cong M'$.

For the converse, suppose that $(\sigma, \tau)$ is an isomorphism from $E(M)$ to $E(M')$ in the category $S(1_2)$. Let $F$ be a block diagonal matrix over $\Lambda$ such that $F_{11} = \sigma, F_{22} = \tau, F_{33} = F_{22}$, and all other entries of $F$ are zero; as in the proof of “full” above, right multiplication by $F$ induces a $\Lambda$-homomorphism $f : M \to M'$ such that $E(f) = (\sigma, \tau)$. But $(\sigma, \tau)$ is an isomorphism in the category $S(1_2)$, and so $\sigma$ and $\tau$ are invertible matrices over the field $k$. Then the matrix $F$ is invertible, too. The same argument shows that the (block diagonal) matrix $F^{-1}$ induces a $\Lambda$-homomorphism $g : M \to M'$ such that $E(g) = (\sigma^{-1}, \tau^{-1}) = (\sigma, \tau)^{-1}$. Moreover, since $FF^{-1}$ is the identity map on $P$ (the projective cover of $M$), it follows that the induced map $fg$ is the identity map on $M$. Similarly, $gf$ is the identity map on $M'$, and hence $f$ is an isomorphism between $M$ and $M'$. □
5. Klein and Dedekind-like rings.

In this section, we show that a Klein ring is a homomorphic image of a Dedekind-like ring if and only if its residue field is imperfect (Theorem 5.2). Thus, there are Klein rings which are homomorphic images of Dedekind-like rings, and others which are not. For example, the group algebra $kG$, of the Klein 4-group $G$ over a field $k$ of characteristic 2, is a Klein ring (with residue field $k$); and $kG$ is a homomorphic image of a Dedekind-like ring if and only if $k$ is imperfect. Thus, as remarked in §2, we cannot simplify Theorem 2.10 by omitting Klein rings from its statement.

We conclude this section with an example of a Klein ring of characteristic 4.

We begin by determining which Dedekind-like rings can be mapped onto Klein rings.

**Theorem 5.1.** Let $(\Lambda, m, k)$ be a Dedekind-like ring.

(i) If $\Lambda$ is split or a DVR, then $\Lambda$ cannot map onto a Klein ring.

(ii) If $\Lambda$ is unsplit with normalization $(\Gamma, m, F)$, then $\Lambda$ maps onto a Klein ring if and only if $F$ is an inseparable extension of $k$.

**Proof.** (i) Clearly a DVR, whose maximal ideal is principal, cannot map onto a Klein ring, whose maximal ideal requires two generators.

Thus let $(\Lambda, m, k)$ be a split Dedekind-like ring and $\Omega$ a Klein ring, and suppose, by way of contradiction, that there is a surjective ring homomorphism $\phi : \Lambda \to \Omega$. Then there would be a surjective homomorphism of the completion $\hat{\Lambda}$ onto $\Omega$, since artinian rings are already complete. Moreover, $\hat{\Lambda}$ is again split Dedekind-like, by Lemma 2.21. Therefore we may assume that $\Lambda$ is complete. As noted after (2.5.2), it follows that $\Lambda$ is strictly split.

Since $\Lambda$ is strictly split, its maximal ideal can be generated by two elements whose product is zero [Lemma 2.20]. Therefore the same is true of the maximal ideal $n$ of $\Omega$: say $n = (x, y)$. Since $xy = 0$ and the square of every element of the maximal ideal of any Klein ring is zero, we have $n^2 = 0$. This is the desired contradiction, since the definition of “Klein ring” requires that $\mu_\Omega(n^2) = 1$.

Before considering statement (ii) we prove two lemmas:

(5.1.1) Let $\Lambda$ be a Dedekind-like ring with normalization $\Gamma$, and $\varepsilon$ any element of $\Gamma - \Lambda$. Then $\Gamma = \Lambda + \Lambda \varepsilon$.

Since $m$ is an ideal of both rings $\Lambda$ and $\Gamma$, it suffices to prove this modulo $m$, whence it follows from the fact that $\Gamma/m$ is a 2-dimensional vector space over the field $\Lambda/m = k$.

(5.1.2) Let $(\Lambda, m, k)$ be an unsplit Dedekind-like ring with normalization $(\Gamma, m, F)$, and $\varepsilon \in \Gamma - \Lambda$. If $F$ has characteristic 2 and is separable over $k$, we have $\Gamma = \Lambda + \varepsilon^2 \Lambda$. 
To prove this it suffices, in view of (5.1.1), to prove that \( \varepsilon^2 \notin \Lambda \). Suppose, therefore that \( \varepsilon^2 \in \Lambda \). Then we have \( F = k[\overline{\varepsilon}] \) where \( \overline{\varepsilon} \) is the image of \( \varepsilon \) in \( F \), and \( \overline{\varepsilon}^2 \in k \). Since \( F \) has characteristic 2, we get the contradiction that \( F \) is inseparable over \( k \).

(ii) By Lemma 2.9, we can assume that \( k \) and \( F \) have characteristic 2.

Choose any \( \varepsilon \in \Gamma - \Lambda \), recall that \( \Gamma \) is DVR, and let \( x \) be any \( \Gamma \)-generator of \( m \) (i.e., \( m = \Gamma x \)).

First suppose that \( F \) is separable over \( k \). Multiplying the equation at the end of (5.1.2) by \( x^2 \) yields \( m^2 = \Lambda x^2 + \Lambda(\varepsilon x)^2 \). If there were a homomorphism of \( \Lambda \) onto a Klein ring \( (\Omega, n, k) \), the fact that the square of every element of \( n \) is zero would then yield the contradiction \( n^2 = 0 \).

Suppose instead that \( F \) is inseparable over \( k \). We claim that \( \Omega = \Lambda / (m^3 + \Lambda x^2) \) is a Klein ring.

Let \( n \) be the maximal ideal of \( \Omega \). Then we have \( \mu_\Omega(n^2) = 2 \) since \( \mu_\Lambda(m) = 2 \) [Lemma 2.7], and \( m^3 + \Lambda x^2 \subseteq m^2 = (\text{rad } \Lambda)m \).

Next we show that \( \mu_\Omega(n^2) = 1 \). Since \( \Gamma \) is a DVR with residue field \( F \), we have \( m^2/m^3 \cong F \) as \( \Gamma \)-modules, and hence as \( k \)-vector spaces, so that \( m^2/m^3 \) has dimension 2 as a vector space over \( k \). Again, since \( \Gamma \) is a DVR, we have \( x^2 \in m^2 - m^3 \), and therefore the dimension of the \( k \)-vector subspace \( (m^3 + \Lambda x^2)/m^3 \) of \( m^2/m^3 \) equals 1. But \( n^2 = m^2/(m^3 + \Lambda x^2) \), so it also must have dimension 1 as a vector space over \( k \), and therefore \( \mu_\Omega(n^2) = 1 \), as desired.

Obviously \( n^3 = 0 \). We claim that every element of \( n \) has square zero. First note that \( m = \Gamma x = \Lambda \varepsilon x + \Lambda \varepsilon x \) and \( m^2 = \Gamma x^2 = \Lambda x^2 + \Lambda \varepsilon x^2 \), by (5.1.1). Then take \( a, b \in \Lambda \). Since \( 2 \in m \) (because \( F \) has characteristic 2) and \( \varepsilon^2 \in \Lambda \) (because \( F = k[\overline{\varepsilon}] \) is inseparable over \( F \)) we have

\[
(ax + b \varepsilon x)^2 = a^2 x^2 + 2(abx)(\varepsilon x) + b^2 \varepsilon^2 x^2 \in \Lambda x^2 + m^3 + \Lambda x^2
\]

which proves the claim.

Taking the opposite point of view, we determine which Klein rings are homomorphic images of Dedekind-like rings.

**Theorem 5.2.** If \( (\Omega, m, k) \) is a Klein ring, then \( \Omega \) is a homomorphic of a (necessarily unsplit) Dedekind-like ring if and only if its residue field \( k \) is imperfect.

**Proof.** If the Klein ring \( (\Omega, m, k) \) is a homomorphic image of the Dedekind-like ring \( \Lambda \), then \( k \) is the residue field of \( \Lambda \). By Theorem 5.1, \( \Lambda \) must be unsplit, and \( k \) must be imperfect.

Conversely, suppose that \( k \) is the imperfect residue field (necessarily of characteristic 2, by Lemma 2.9) of the Klein ring \( (\Omega, m, k) \). We use the same idea as in the proofs of Propositions 3.4 and 3.6 to construct an unsplit Dedekind-like ring that maps onto \( \Omega \).
Fix generators \(x\) and \(y\) of \(\mathfrak{m}\). By assumption \(x^2 = y^2 = 0\) and \(\mathfrak{m}^2 = \Omega xy \neq 0\). Using the Structure Theorem of Complete Local Rings [N, Theorem 31.1], there is a surjective homomorphism \(\phi: V[[X,Y]] \rightarrow \Omega\) such that \(\phi(X) = x\) and \(\phi(Y) = y\), where \(V[[X,Y]]\) is a formal power series ring, and either \(V = k\), or \(V\) is a complete DVR of characteristic 0 with residue field \(k\) of characteristic 2 and the maximal ideal of \(V\) is generated by 2.

As in the proofs of Propositions 3.4 and 3.6, if \(V\) is a field, we set \(R = V[[X,Y]]\), while if \(V\) is not a field, we set \(R = V[[X,Y]]/(2 - (AX + BY))\), where \(A, B \in V[[X,Y]]\) are elements such that \(\phi(2 - (AX + BY)) = 0\), changing notation so that \(\phi\) is still the map from \(R\) onto \(\Omega\). Again, \(\phi: R \rightarrow \Omega\) is a surjective homomorphism such that \(\phi(X) = x\) and \(\phi(Y) = y\), where \(R\) is a complete two-dimensional regular local ring with maximal ideal \(\mathfrak{m}_R = (X,Y)\), and \(\phi\) induces an isomorphism between the residue fields of \(R\) and \(\Omega\).

Because \(k\) is imperfect of characteristic 2, not every element of \(k\) is a square in \(k\), so we can choose \(C \in R\) whose coset modulo \(\mathfrak{m}_R\) is not a square in \(k\). We claim that \(CX^2 + Y^2\) is an irreducible element in the unique factorization domain \(R\). Suppose, by way of contradiction, that \(CX^2 + Y^2\) is the product of two elements of \(\mathfrak{m}_R\):

\[
CX^2 + Y^2 = (SX + UY)(TX + VY) = STX^2 + (SV + UT)XY + UVY^2
\]

where \(S, T, U, V \in R\). We know that \(\mu_R(\mathfrak{m}_R^2) = 3\) (by [AM, Theorem 11.22]), and \(\mathfrak{m}_R^2 = (X^2, XY, Y^2)\). Therefore the images of \(X^2, XY,\) and \(Y^2\) in \(\mathfrak{m}_R^2/\mathfrak{m}_R^3\) are \(k\)-linearly independent. Comparing the coefficients of \(Y^2\) on both sides of (5.2.1) therefore shows that \(UV \equiv 1\) (all congruences in this part of the proof are modulo \(\mathfrak{m}_R\)). Therefore \(U\) and \(V\) are units in \(R\). After multiplying the first and second factors in the middle part of (5.2.1) by \(U^{-1}\) and \(U\) respectively, and changing notation, we now have \(U = 1\) and \(V \equiv 1\). Comparing coefficients of \(XY\) then shows that \(S + T \equiv 0\), which is equivalent to \(S \equiv -T\). But since \(k\) has characteristic 2, this is equivalent to \(S \equiv T\). Comparing coefficients of \(X^2\) therefore shows that \(C \equiv S^2\), contrary to our choice of \(C\) as an element of \(R\) whose coset modulo \(\mathfrak{m}_R\) is not a square. Thus, the claim is proved.

Set \(\Lambda = R/(CX^2 + Y^2)\), an integral domain because \(CX^2 + Y^2\) is an irreducible element of the unique factorization domain \(R\). We claim that \(\Lambda\) is unsplit Dedekind-like.

Let \(c, x,\) and \(y (\in \Lambda)\) denote the cosets in \(\Lambda\) of \(C, X,\) and \(Y,\) respectively, so that \(cx^2 + y^2 = 0\) in \(\Lambda\). (Since we have no further need of \(\Omega\), this duplication of earlier notation will cause no harm.) Let \(\Gamma = \Lambda[y/x]\), a subring of the total quotient ring of \(\Lambda\). Since \((y/x)^2 = c \in \Lambda, \Gamma\) is an integral extension of \(\Lambda\); in fact, \(\mu_\Lambda(\Gamma) \leq 2\).
We claim that \( m = \Lambda x + \Lambda y \), which we write as \( m \) from now on, is an ideal of \( \Gamma \). Since \( m = \Lambda x + \Lambda \bar{y} \), it suffices to check that \( x(y/x) \) and \( y(y/x) \) are elements of \( m \). The former is obvious, and the latter equals \( (y^2/x^2)x = cx \in m \).

Let \( F = \Gamma/m \), an algebra over its subfield \( k = \Lambda/m \). We claim that \( F \) is a 2-dimensional field extension of \( k \). Let \( \varepsilon = y/x \), and denote the natural images of \( \varepsilon \) and \( c \) in \( F \) by \( \overline{\varepsilon} \) and \( \overline{c} \) respectively. Then \( \overline{c} \in \Lambda/m = k \) and \( \varepsilon^2 - \overline{c} = 0 \). By our choice of \( C \), the element \( \overline{\varepsilon} \) is a nonsquare in \( k \), and therefore the polynomial \( z^2 - \overline{c} \) is irreducible over \( k \). This shows that \( F = k[\overline{\varepsilon}] \) is a 2-dimensional extension field of \( k \) as claimed.

Next we claim that \( m = \text{rad}(\Gamma) \). Since \( m \) is an ideal of \( \Gamma \) the inclusion \( (\subseteq) \) holds if we show that \( 1 - m \) is invertible for every \( m \in m \), and this holds since \( m = \text{rad}(\Lambda) \). The opposite inclusion holds because \( F \) is a field.

Finally we claim that \( \Gamma \) is a DVR (and is therefore the normalization of \( \Lambda \)). \( \Gamma \) is a local ring because \( \Gamma/\text{rad}(\Gamma) \) is a field. Therefore it suffices to show that \( m \) is a principal ideal of \( \Gamma \). But \( m = \Lambda x + \Lambda y \), so it suffices to observe that \( y = (y/x)x \in \Gamma x \). □

Remark 5.3 (Maximal tame?). Recall that, in the introduction, we called a ring “maximal tame” (following Ringel) if it is tame but not a proper homomorphic image of any other tame ring. In the commutative part of the situation considered by Ringel [R] — complete local rings that are algebras over their algebraically closed residue field — Klein rings turned out to be maximal tame. We compare this to the corresponding result in our more general situation.

Let \( (\Omega, m, k) \) be a Klein ring, not necessarily an algebra over a field. Then \( \Omega \) remains tame, by [KL2]. Moreover, \( \Omega \) remains maximal tame if \( k \) is perfect, because by our Theorem 5.2 \( \Omega \) is not a homomorphic image of a Dedekind-like ring.

On the other hand, if \( k \) is imperfect of characteristic 2, then the result might change. In this case \( \Omega \) is a homomorphic image of an unsplit Dedekind-like ring, say \( (\Lambda, m, k) \), with normalization \( (\Gamma, m, F) \) and \( F \) inseparable over \( k \). Here we are in the exceptional situation (1.0.1), in which we do not know whether \( \Lambda \) is tame, wild, or neither. Thus, if \( k \) is imperfect of characteristic 2, we do not know whether \( \Omega \) is maximal tame, although it is both tame and a homomorphic image of a Dedekind-like ring.

Example 5.4 (Klein ring of characteristic 4). Let \( \Lambda_p = \mathbb{Z}[X]/(X^2, p^2) \), where \( p \) is a prime number. In the case \( p = 2 \), we claim that \( \Lambda_2 \) is a Klein ring of characteristic 4.

For every \( p \), \( \Lambda_p \) is a local artinian ring with maximal ideal \( m_p = (x, \overline{p}) \), where \( x \) and \( \overline{p} \) denote the images of \( X \) and \( p \) respectively, in \( \Lambda_p \). Clearly, \( m_p^3 = 0 \), \( \mu_A(m_p) = 2 \), and \( \mu_A(m_p^2) = 1 \) (generated by \( \overline{px} \)).
We check that the square of every element of \(m_2\) is zero. For any \(w = ax + 2b \in m_2\) we have \(w^2 = (ax + 2b)^2 = a^2x^2 + 4abx + 4b^2 = 0\) in \(\Lambda_2 = \mathbb{Z}[x]/(x^2, 4)\).

Note that \(\Lambda_p\) cannot be a Klein ring when \(p \neq 2\), because its characteristic is neither 2 nor 4 [Lemma 2.9]. In fact, \(\Lambda_p\) is a homomorphic image of a Dedekind-like ring when \(p \neq 2\). To prove this, note that \(\Lambda_p\) does not map onto an artinian triad since \(\mu(m_p) = 2\), and does not map onto a Drozd ring since \(\mu(m_p^2) = 1\). Therefore, by our Ring-theoretic Dichotomy Theorem 3.1, \(\Lambda_p\) is either a Klein ring (which we already know it is not) or a homomorphic image of a Dedekind ring (as claimed).

### 6. Drozd rings and ramification.

This section begins with two simple examples: A Drozd ring that is not an algebra over a field, and a ring that we call “superwild,” something that cannot happen for algebras over fields. The main theorem of this section [Theorem 6.4] shows that every Drozd ring is a homomorphic image of a complete ramified order. Conversely, every complete ramified order is finite-length wild [Proposition 6.3]. We apply this to give the promised constructive definition of Drozd rings as a certain type of subring of local artinian principal ideal rings. Finally, we display in this form the two Drozd rings previously discussed.

**Example 6.1.** The simplest example of a Drozd ring that is not an algebra over a field, is the ring \(A_p = \mathbb{Z}[X]/(X^2, p^3, p^2X)\), where \(p\) denotes any prime number.

**Example 6.2 (Superwild).** The innocuous-looking (non-local) ring \(\Lambda = \mathbb{Z}[X]/(X^2)\) of Krull dimension 1 is superwild in the following sense: For every prime number \(p\), \(\Lambda\) is finite-length wild with respect to some residue field of characteristic \(p\).

This holds because, for each prime number \(p\), the ring \(A_p\) in Example 6.1 is a Drozd ring with residue field \(\mathbb{Z}/(p)\), and \(A_p\) is a homomorphic image of \(\Lambda\).

Three basic concepts in algebraic number theory, relating an order to an overorder, are those of splitting, residue-field growth, and ramification of a maximal ideal. The first two of these concepts occur in the present paper in the definitions of split and unsplit Dedekind-like rings, but ramification has yet to make an appearance. In fact, for orders ramification leads to wildness.

**Proposition 6.3 (Ramification).** Let \((\Lambda, m, k)\) be any complete local reduced ring that ramifies in its normalization \(\Gamma\), in the sense that \(m\) is contained in the square of some maximal ideal of \(\Gamma\). Then \(\Lambda\) is finite-length wild.
Proof. Suppose that \( \Lambda \) is not finite-length wild. First note that \( \Lambda \) is not artinian, since then it would be a field. Therefore, by our Main Wildness Theorem 2.10, \( \Lambda \) is a homomorphic image of a Dedekind-like ring, say \( \Omega \). Moreover, \( \Lambda \not\cong \Omega \) because the maximal ideal of a Dedekind-like ring never ramifies in its normalization. Thus the homomorphism \( \Omega \to \Lambda \) is proper.

The Dedekind-like ring \( \Omega \) cannot be unsplit, since then it is an integral domain; and all proper homomorphic images of noetherian domains of dimension 1 are artinian. Thus the only remaining possibility is that \( \Omega \) is strictly split.

The normalization of \( \Omega \) is therefore the direct sum of two DVRs, say \( V_1 \oplus V_2 \). The imbedding \( \Omega \subseteq V_1 \oplus V_2 \) is described in detail in Lemma 2.15. In particular, it is easy to see that \( \Omega \) has only two dimension 1 prime ideals — the two projection maps \( \Omega \to V_i \) — and therefore the only possible non-artinian, reduced homomorphic images of \( \Omega \) are the coordinate rings \( V_i \). But since \( V_i \) is its own normalization, the maximal ideal of \( V_i \) does not ramify in its normalization. This contradiction completes the proof. \( \square \)

It is an interesting fact that all Drozd rings arise in this way; that is, every Drozd ring is a homomorphic image of a ramified order.

**Theorem 6.4** (Ramified onto Drozd). Let \((\Lambda, n, k)\) be a Drozd ring. Then \( \Lambda \) is a homomorphic image of a completely ramified integral domain \( \Omega \). In more detail, there is a commutative diagram of ring homomorphisms (6.4.1) in which:

(i) \((\Omega, m, k)\) is a complete local domain whose normalization is a DVR \((\Gamma, (z), k)\) such that \( m = \Gamma z^2 \);

(ii) The vertical maps are surjections; and the left-hand vertical maps are the restrictions of \( \rho' \) and \( \rho'' \).

\[
\begin{array}{ccc}
\Omega & \subseteq & \Gamma \\
\downarrow & & \downarrow \\
\Lambda & \subseteq & \Gamma/(z^6) \\
\downarrow & & \downarrow \\
k & \subseteq & k + k\overline{z} \quad (\overline{z}^2 = 0, \overline{z} \neq 0)
\end{array}
\]

(6.4.1)

*Proof.* The approach is similar to that in the proofs of Propositions 3.4 and 3.6. We can write the maximal ideal of the Drozd ring \( \Lambda \) as \( n = (x, y) \), where \( x^2 = y^3 = 0 \) and \( xy \neq 0 \). As in the proofs of Propositions 3.4 and 3.6, by the Structure Theorem of Complete Local Rings [N, Theorem 31.1], there is a surjective homomorphism \( \phi : V[[X, Y]] \to \Lambda \) such that \( \phi(X) = x \) and \( \phi(Y) = y \), where \( V[[X, Y]] \) is a power series ring and either \( V = k \) or \( V \) is a complete DVR of characteristic 0 with residue field \( k \). Moreover, in
this latter case, $k$ has characteristic $p \neq 0$ and the maximal ideal of $V$ is generated by $p$.

If $V$ is a field, we set $R = V[[X,Y]]$, a complete, two-dimensional, regular local ring, with maximal ideal $(X,Y)$. If $V$ is a DVR, then again exactly as in the proofs of Propositions 3.4 and 3.6, we note that there are elements $A, B \in V[[X,Y]]$ such that $\phi(p - (AX + BY)) = 0$, and we set $R = V[[X,Y]]/(p - (AX + BY))$, a regular local ring by [AM, Theorem 11.22]. In this latter case, we change notation and let $\phi$ denote the induced map from $R$ onto $\Lambda$, and we let $X$ and $Y$ denote the cosets of $X$ and $Y$, respectively, in $R$. Thus, in either case we have a surjective homomorphism $\phi : R \to \Lambda$ such that $\phi(X) = x$ and $\phi(Y) = y$, where $R$ is a complete two-dimensional regular local ring and $(X,Y)$ is the maximal ideal of $R$.

Now $\phi(X^2 - Y^3) = x^2 - y^3 = 0$, and so $\phi$ induces a surjective homomorphism from $\Omega = R/(X^2 - Y^3)$ onto $\Lambda$. We claim that $X^2 - Y^3$ is irreducible in $R$, so that $\Omega$ is an integral domain.

Suppose, by way of contradiction, that $X^2 - Y^3$ were reducible, say

\[(6.4.2) \quad X^2 - Y^3 = (aX + bY)(cX + dY) = acX^2 + (ad + bc)XY + bdY^2\]

for some elements $aX + bY$ and $cX + dY$ of the maximal ideal $(X,Y)$, with $a, b, c, d \in R$. Note that $(X,Y)^2 = (X^2, XY, Y^2)$ and, since $R$ is a regular local ring of dimension 2, the elements $X^2, XY, Y^2$ form a basis of $(X,Y)^2/(X,Y)^3$. Thus reading (6.4.2) modulo $(X,Y)^3$, we see that $ac$ is a unit in $R$, and hence $a$ and $c$ are units. Moreover, by this same reasoning, $1 - ac \in (X,Y)$.

Similarly, looking at the coefficient of $Y^2$ and reading (6.4.2) modulo $(X,Y)^3$ shows that $bd \in (X,Y)$, and hence at least one of $b$ and $d$ is in $(X,Y)$. But then both $b$ and $d$ must be in $(X,Y)$, because otherwise, (using the fact that $a$ and $c$ are both units) the coefficient of $XY$ would be a unit on the right-hand side of (6.4.2) but zero on the left-hand side.

We have $(X,Y)^3 = (X^3, X^2Y, XY^2, Y^3)$ and the elements $X^3, X^2Y, XY^2,$ and $Y^3$ form a basis of $(X,Y)^3/(X,Y)^4$ (again because $R$ is a regular local ring of dimension 2). Rearranging (6.4.2), we get that

\[(6.4.3) \quad (1 - ac)X^2 - Y^3 - (ad + bc)XY \equiv 0 \pmod{(X,Y)^4}\]

since $b$ and $d$ are in $(X,Y)$, so that $bdY^2 \in (X,Y)^4$. Since $1 - ac, b,$ and $d$ are all in $(X,Y)$ we can substitute expressions of the form $1 - ac = eX + fY$ and $ad + bc = gX + hY$ into (6.4.3). Since the coefficient of $Y^3$ is a unit, this gives a nontrivial linear relation among the basis elements of $(X,Y)^3/(X,Y)^4$, and this contradiction proves our claim that $X^2 - Y^3$ is irreducible in $R$. Therefore $\Omega$ is an integral domain.

Let $s$ and $t$ denote the cosets of $X$ and $Y$, respectively, in $\Omega = R/(X^2 - Y^3)$, and recall that $m$ denotes the maximal ideal of $\Omega$, so that $m = (s,t)$. Since $s^2 = t^3$, we have $(s/t)^2 = t$ in the quotient field of $\Omega$. It follows that
\[ z = s/t \text{ is integral over } \Omega, \text{ and hence } \Gamma = \Omega[z] \text{ is an integral extension of } \Omega. \]

Moreover, since \( z \cdot t = s \in m \) and \( z \cdot s = s^2/t = t^2/t = t^2 \in m \), we have that \( m \) is an ideal of \( \Gamma \) also. Since \( m \) is the maximal ideal of \( \Omega \), it follows that \( m \) is the conductor from \( \Gamma \) to \( \Omega \). Also, \( m = \text{rad}(\Omega) \), and \( \Gamma \) is finitely generated as an \( \Omega \)-module, so \( m \subseteq \text{rad}(\Gamma) \). We claim that

\[
\mu_\Omega(m) = 2, \quad z^2 \in m, \quad z \not\in m.
\]

(6.4.4)

We already have that \( m = (s, t) \) as an ideal of \( \Omega \). The fact that two generators are required follows from the fact that \( m \) maps onto the maximal ideal of the Drozd ring \( \Lambda \). Next, note that \( z^2 = (s/t)^2 = t \in m \). Finally, if \( z \) were an element of \( m \), then \( s/t = z \) would imply that \( s = tz \in m^2 \), which would make the ideal \( m = (s, t) \) of \( \Omega \) principal, contrary to what was just shown.

Since \( \Gamma = \Omega[z] \), (6.4.4) yields

\[
\Gamma/m = k + k\overline{z} \quad \text{with} \quad \overline{z}^2 = 0 \quad \text{and} \quad \overline{z} \neq 0.
\]

(6.4.5)

This yields the first equality in

\[
\text{rad}(\Gamma) = m + \Gamma z = \Gamma z.
\]

(6.4.6)

The second equality holds since \( s = z \cdot t \in \Gamma z \) and \( t = z^2 \in \Gamma z \).

The fact that \( \text{rad}(\Gamma/m) = (\overline{z}) \) is the maximal ideal of \( \Gamma/m \) now shows that \( \text{rad}(\Gamma) = \Gamma z \) is the maximal ideal of \( \Gamma \); that is, \( \Gamma \) is a local domain with principal maximal ideal \( \Gamma z \). Therefore \( \Gamma \) is a DVR and hence is the normalization of \( \Omega \).

Next we note that, since \( \Gamma \) is a DVR with maximal ideal \( \Gamma z \), (6.4.4) implies that \( m = \Gamma z^2 \).

We have now completed the proof of the top line in diagram (6.4.1). We also have established the existence of a surjective ring homomorphism \( \rho' : \Omega \to \Lambda \). The map \( \rho' \) takes \( m = m_\Omega \) onto \( n = m_\Lambda \); and \( n^3 = 0 \) since \( \Lambda \) is a Drozd ring. We conclude that \( m^3 \subseteq \ker \rho' \); we claim that equality holds.

It suffices to prove that the composition length of \( \Omega/m^3 \) equals the length of \( \Lambda \), namely 5. The length of \( \Omega/m^3 \) is at least 5, since \( \Omega \) maps onto \( \Lambda \). Therefore it suffices to prove that the length of the \( \Omega \)-module \( \Omega/m^3 \) is less than 6.

Since \( \Omega \) is properly contained in \( \Gamma \), it suffices to show that the length of the \( \Omega \)-module \( \Omega/m^3 \) equals 6. Now, the simple \( \Gamma \)-module \( k \) remains simple as an \( \Omega \)-module, by (6.4.5) and the fact that \( \Omega/m^3 = k \). Therefore, it suffices to show that the length of the \( \Gamma \)-module \( \Gamma/m^3 \) equals 6. This follows from the fact that \( \Gamma \) is a DVR whose maximal ideal \( \Gamma z \) satisfies \( \Gamma z^2 = m \).

We have now established the existence of the top commutative square in (6.4.1). The existence of the bottom commutative square now follows immediately from (6.4.5), completing the proof of the theorem. \( \square \)

By an AVR (artinian valuation ring) of length \( n \) we mean a local principal ideal ring \( (V, (z), k) \), with lattice of ideals \( V \supset Vz \supset Vz^2 \supset \ldots \supset Vz^n = 0 \).
for some \( n \). We say that \( V \) contains its residue field if \( V \) contains a subfield that maps isomorphically onto \( k \) via the natural homomorphism \( V \to k \).

We are now ready to give our constructive definition of “Drozd ring” as a certain type of subring of an AVR of length 6. This AVR seems to be a kind of “artinian normalization” of the Drozd ring. In more detail, the Drozd ring \( \Lambda \) is the pullback of square (6.5.1), and if we replace the AVR by a DVR that contains its residue field (and delete the condition “\( x^2 = 0 \)”), we get the ordinary conductor square for a simple cusp whose normalization is the DVR.

**Theorem 6.5 (Constructive Definition of Drozd Rings).** Let \((V, (z), k)\) be an AVR of length 6 such that the ring \( V/(z^2) \) contains its residue field (and hence \( V/(z^2) = k + k\bar{z} \) with \( \bar{z}^2 = 0 \) and \( \bar{z} \neq 0 \)), and let \( \Lambda \) be the pullback of diagram (6.5.1). Then:

(i) \( \Lambda \) is a Drozd ring, and every Drozd ring arises in this way.

(ii) \( V = \Lambda + \Lambda z \) and \( m_\Lambda = \Lambda x + \Lambda y \), where \( x = z^3 \), \( y = z^2 \), \( x \in m_V - m_V^2 \), and \( x^2 = 0 \).

\[
\begin{array}{ccc}
\Lambda & \longrightarrow & V \\
\downarrow & & \downarrow \rho \\
k & \longrightarrow & k + k\bar{z}
\end{array}
\]

\[(6.5.1)\]

Proof. The fact that every Drozd ring arises in this way is expressed in the bottom square of diagram (6.4.1).

Conversely, given that \( \Lambda \) is the pullback of diagram (6.5.1), we want to prove that \( \Lambda \) is a Drozd ring satisfying conditions (ii).

Since the image of the left-hand vertical map in (6.5.1) is the field \( k \) and the kernel of this map is nilpotent, we see that \( \Lambda \) is a local ring with maximal ideal \( m_\Lambda = Vz^2 \) and residue field \( k \). Moreover, \( V = \Lambda + \Lambda z \) holds because it obviously holds modulo \( Vz^2 \). Multiplying by \( z^2 \) then shows

\[
m_\Lambda = \Lambda z^2 + \Lambda z^3.
\]

\[(6.5.2)\]

To see that \( m_\Lambda \) requires two generators (as a \( \Lambda \)-module), note that \( m_\Lambda/m_\Lambda^2 = (Vz^2)/(Vz^4) \). Since \( V \) is an AVR, this is \( V \)-isomorphic to \( V/(Vz^2) \), a \( k \)-vector space of dimension 2 by (6.5.1). Similarly, to see that \( m_\Lambda^2 \) requires two generators as a \( \Lambda \)-module, we note that \( m_\Lambda^2/m_\Lambda = Vz^{-1} \) has \( k \)-dimension 2.

Thus, to complete the proof that \( \Lambda \) is a Drozd ring, it suffices to note that \( x = z^3 \) is an element of \( m - m^2 = Vz^2 - Vz^4 \) (obvious, since \( V \) is an AVR), \( x^2 = 0 \), and \( m_\Lambda^3 = Vz^6 = 0 \). \( \square \)

**Examples 6.6.** We display, as subrings of their AVR “normalization” \( V \), the two Drozd rings explicitly discussed in this paper:
(i) \( \Lambda = k + kxky + kxy + ky^2 \), the Drozd \( k \)-algebra originally studied by Drozd. Here \( k \) is any field, and all monomials in \( x \) and \( y \) other than the displayed ones equal 0. Choose \( V = k[z] \) where \( z \) satisfies the defining relation \( z^6 = 0 \). Then \( \Lambda \) is the \( k \)-subalgebra with basis 1, \( z^2 = y \), \( z^3 = x \), \( z^4 = y^2 \), \( z^5 = xy \).

(ii) \( \Lambda = \mathbb{Z}[X]/(X^2, p^3, p^2X) \), the Drozd ring in Example 6.1. Choose \( V = \mathbb{Z}_{p^3}[W]/(W^6, W^2 - p) \) where \( \mathbb{Z}_{p^3} \) denotes the integers modulo \( p^3 \) and \( W \) is an indeterminate. Letting \( w \) be the natural image of \( W \) in \( V \), we see that \( V \) is an AVR of length 6 with composition series

\[ V \supset Vw \supset Vw^2 = Vp \supset Vw^3 = Vpw \supset Vw^4 = Vp^2 \supset Vw^5 = Vp^2w \supset 0 \]

and \( V/Vw^2 \) contains its residue field \( \mathbb{Z}_p \).

The subring of \( V \) additively spanned by the monomials 1, \( w^2 = p \), \( w^3 = pw \), \( w^4 = p^2 \), \( w^5 = p^2w \) is isomorphic to \( \Lambda \), under the correspondence \( p \rightarrow p \) and \( pw \rightarrow x \).

7. Infinite lattice type.

If a local ring \( \Lambda \) is reduced (no nonzero nilpotent elements), then submodules of free \( \Lambda \)-modules of finite rank are often called lattices. (This is equivalent to the definition of “lattice” used in integral representation theory.) \( \Lambda \) is said to have infinite lattice type, or less precisely, “infinite representation type,” if it has infinitely many indecomposable lattices. There exist many such \( \Lambda \) with infinite but tame lattice type. However, in the present context, such commutative noetherian rings are finite-length wild.

**Theorem 7.1.** Let \((\Lambda, m, k)\) be a complete local reduced ring. If \( \Lambda \) has infinite lattice type, then \( \Lambda \) is finite-length wild.

**Proof.** First recall that a reduced, non-artinian, proper homomorphic image of a strictly split or unsplit Dedekind-like ring must be a DVR.

Next note that \( \Lambda \) is not artinian, since then it would be a direct product of fields, and hence have finite lattice type.

It is easiest to prove the contrapositive of our theorem, so suppose that \( \Lambda \) is not finite-length wild. Then by our Main Wildness Theorem 2.10 and the fact that \( \Lambda \) is not artinian, \( \Lambda \) must be a homomorphic image of a complete split (and hence strictly split) or unsplit Dedekind-like ring \( \Omega \). Since \( \Lambda \) is reduced, it must therefore be either one of these types of Dedekind-like rings or else (see the first paragraph of this proof) a DVR.

Since DVRs are well-known to have finite lattice type, we may suppose that \( \Lambda \) is split or unsplit Dedekind-like, with normalization \( \Gamma \). Then \( \mu_\Lambda(\Gamma/\Lambda) = 1 \) and \( \mu_\Lambda(\rad(\Gamma/\Lambda)) = 0 \) (since \( m \cdot \Gamma/\Lambda = 0 \), so that \( \Gamma/\Lambda \) is a \( k \)-module, hence a simple \( \Lambda \)-module). Therefore \( \Lambda \) satisfies the Drozd-Roiter conditions and hence, by a result of Cimen and Wiegand (extending
earlier results of Drozd, Roiter, Jacobinski, Green, and Reiner on lattices over orders) must have finite lattice type [CWW]. □

References


Received September 9, 1999 and revised November 29, 2000. Levy’s research was partially supported by an NSA grant.

**Mathematics Department**  
**Florida Atlantic University**  
**Boca Raton, Florida 33431-0991**  
*E-mail address*: klingler@fau.edu

**Mathematics Department**  
**University of Wisconsin**  
**Madison, Wisconsin 53706-1388**  
*E-mail address*: levy@math.wisc.edu
REPRESENTATION TYPE OF COMMUTATIVE NOETHERIAN RINGS II: LOCAL TAMENESS

Lee Klingler and Lawrence S. Levy

We describe all isomorphism classes of finitely generated \( \Lambda \)-modules, where \( \Lambda \) is any complete local (commutative noetherian) ring whose category of modules of finite length does not have wild representation type. (There is a possible exception to our results, involving characteristic 2.)

1. Introduction.

This is the second of a 4-paper series \([KL1]-[KL4]\) whose purpose is to determine all isomorphism classes of finitely generated modules over all commutative noetherian rings that do not have wild representation type. The first two papers of the series consider the complete local case of this problem. If all isomorphism classes of finitely generated modules over a ring \( \Lambda \) can be described, we call \( \Lambda \) finitely-generated tame.

Let \( \Lambda \) be a complete local (commutative, noetherian) ring. In \([KL1]\) we showed that the category of \( \Lambda \)-modules of finite length has wild representation type unless \( \Lambda \) belongs to one of the following quite small classes of rings: (i) a homomorphic image of one of two types of rings of Krull dimension 1 that we call “strictly split Dedekind-like” and “unsplit Dedekind-like,” and (ii) a type of artinian ring that we call a “Klein ring.” See \([KL1], \S\S1,2]\) for the definition of “wild representation type” and a very brief introduction to the subject from the point of view of commutative noetherian rings.

The purpose of the present paper is to describe all isomorphism classes of finitely generated modules over the rings of types (i) and (ii) above — with a possible exception involving characteristic 2 [see (1.1.3) for the extra hypothesis needed to avoid this exception]. The bulk of the present paper is devoted to establishing finitely-generated tameness of strictly split and unsplit Dedekind-like rings themselves. Extending this to homomorphic images of these rings, and to Klein rings, is then not difficult.

We begin by defining the rings with which the bulk of this paper deals.

**Notation 1.1.** Throughout this paper, ring means “commutative ring,” unless otherwise stated. By a local ring \((\Lambda, \mathfrak{m}, k)\) we mean a (commutative) noetherian local ring \( \Lambda \) with maximal ideal \( \mathfrak{m} \) and residue field \( k = \Lambda/\mathfrak{m} \).
We use the abbreviation DVR for discrete valuation ring; we do not consider fields to be DVRs.

Consider a commutative diagram of rings:

\[
\begin{array}{ccc}
\Lambda & \subset & \Gamma \\
\downarrow \rho & & \downarrow \rho \\
k & \subset & \Gamma
\end{array}
\]

(1.1.1) \((\ker(\rho) = m = \text{rad } \Gamma)\)

where the right-hand map \(\rho\) is a surjective ring homomorphism, the left-hand map \(\rho\) is the restriction of this map to \(\Lambda\), “rad” denotes “Jacobson radical,” the conditions listed at the right of the diagram hold, and \(\Lambda\) is the pullback of this diagram; that is,

\[
\Lambda = \{ x \in \Gamma \mid \rho(x) \in k \}.
\]

In this situation, \(m = \ker(\rho) \subset \Lambda\), and \(\Lambda\) is noetherian. (See Lemma 4.2 for a generalization of this last fact.) Moreover, \(m = \text{rad } (\Lambda)\), and therefore \((\Lambda, m, k)\) is a local ring. The following two specializations of this situation are the principal rings considered in this paper.

We call \((\Lambda, m, k)\) an unsplit Dedekind-like ring if \((\Gamma, m, \Gamma)\) is a DVR, \(\Gamma = F\) is a field such that \(\dim_k F = 2\), and:

\[
\text{The 2-dimensional field extension } k \subset F \text{ is separable.}
\]

The only reason for including condition (1.1.3) in this definition is that, when it fails, so does our tameness proof, and we do not know whether \(\Lambda\) is tame, wild, or neither. However, this exception cannot occur for rings of number-theoretic origin, because finite-dimensional extensions of finite fields are always separable.

For the rest of this paper — when \(\Lambda\) is unsplit Dedekind-like — we fix an element \(\pi \in \Gamma\) such that \(\Gamma \pi = m\), and call \(\pi\) the standard \(\Gamma\)-generator of \(m\).

Returning to diagram (1.1.1), we call \((\Lambda, m, k)\) a strictly split Dedekind-like ring if \(\Gamma = \Gamma_1 \oplus \Gamma_2\), where \((\Gamma_1, m_1, k)\) and \((\Gamma_2, m_2, k)\) are DVRs with the common residue field \(k\), and \(m = m_1 \oplus m_2\). In this situation, \(\Gamma = k \oplus k\). We usually identify \(k\) with the set of elements \((x, x) \in \Gamma = k \oplus k\), and this is the way that we interpret the inclusion \(k \subset \Gamma\) shown in (1.1.1).

Note that we again have \(\dim_k (\Gamma) = 2\). For the rest of this paper — when \(\Lambda\) is strictly split Dedekind-like — we fix elements \(\pi_i \in \Gamma_i\) \((i = 1, 2)\) such that \(\Gamma_i \pi_i = m_i\), and we call \(\pi_i\) the standard \(\Gamma_i\)-generator of \(m_i\).

In both the unsplit and strictly split case \(\Gamma\) is clearly the normalization of \(\Lambda\).

**Remarks 1.2.** (i) Two other types of local Dedekind-like rings are defined in [KL1, Definition 2.5], and will play a role later in this series of papers, when we drop the “complete local” hypothesis: (a) “Nonstrictly split
Dedekind-like rings.” These cannot occur in the complete case [KL1, Definition 2.5] and therefore play no role in the present paper. (b) DVRs are considered to be Dedekind-like in this series of papers, so that “Dedekind-like” generalizes “Dedekind.” But this paper has nothing new to say about modules over DVRs, so in this paper DVRs are used only as the starting point for building unsplit and strictly split Dedekind-like rings.

(ii) In the present paper, we do not assume that our unsplit and strictly split Dedekind-like rings \((\Lambda, \mathfrak{m}, k)\) are \(\mathfrak{m}\)-adically complete. The only property we need, that completeness would provide — the Krull-Schmidt Theorem for finitely generated modules — already holds for these two types of Dedekind-like rings [Lemma 1.3]. Not having to worry about the complications of completions will simplify a few of our proofs.

(iii) For interesting examples of naturally-occurring Dedekind-like rings, see §12.

(iv) An abstract definition of “Dedekind-like local ring” is given in [KL1, Definition 2.5]. This definition is equivalent to the present one by [KL1, Lemmas 2.14–2.16]. But the present, more constructive definition is better suited to the purposes of the present paper.

**Lemma 1.3.** The Krull-Schmidt Theorem holds for finitely generated modules over unsplit and strictly split Dedekind-like rings.

**Proof.** Note that \(\Lambda\) has no nonzero nilpotent elements, because \(\Lambda\) is contained in the direct sum of one or two integral domains. Therefore the total quotient ring of \(\Lambda\) is the localization \(\Lambda_Q\) that inverts all elements of \(\Lambda\) that lie outside of every minimal prime ideal. Also, \(\Lambda\) is noetherian, \(\Lambda \neq \Gamma\), and the local ring \(\Lambda\) is an indecomposable ring. Under these circumstances, a special case of [LO, Theorem 1.1 and Remark 2.2] is that \(\Lambda\) satisfies the Krull-Schmidt Theorem for finitely generated modules if and only if primitive idempotents of \(\Gamma\) remain primitive in \(\Gamma/\text{rad}(\Gamma)\). This last condition is clearly satisfied by unsplit and strictly split Dedekind-like rings. \(\square\)

Because of the length and complexity of this paper, we now state the contents of each section, and very briefly sketch the main ideas in the proofs leading to our main structure theorems.

§2 Indecomposable modules, unsplit case. This section is a detailed description of the structure of indecomposable finitely generated \(\Lambda\)-modules, when \(\Lambda\) is unsplit Dedekind-like. The proofs of the theory that yields this structure, together with the analogous structure in the strictly split case, occupy Sections 4–10. In particular, proofs of the specific assertions of §2 occur in §9.

§3 Indecomposable modules, strictly split case. This section is a detailed description of the structure of indecomposable finitely generated \(\Lambda\)-modules, when \(\Lambda\) is strictly split Dedekind-like. Proofs of the specific assertions of the §3 occur in §10.
§4 Separated covers and almost functorial property. In this section we introduce the basic abstract concept that allows us to deal with module structure over rings that are not necessarily algebras over fields. Let $\Lambda$ be an unsplit or strictly split Dedekind-like ring with normalization $\Gamma$. Since $\Gamma$ is either a DVR or the direct sum of two DVRs, the structure of finitely generated $\Gamma$-modules is well-known. The basic difficulty in dealing with $\Lambda$-modules — a difficulty that does not arise when studying only lattices — is that not all $\Lambda$-modules are contained in $\Gamma$-modules.

To deal with this difficulty, we define a “best approximation” to each finitely generated $\Lambda$-module $M$ by means of a $\Lambda$-submodule $S$ of some finitely generated $\Gamma$-module $X = \Gamma S$, together with a surjective $\Lambda$-homomorphism $S \to M$. We call this homomorphism $S \to M$ a “separated cover.”

It turns out to be a triviality that separated covers of finitely generated $\Lambda$-modules exist. The main result of this section is that separated covers of $\Lambda$-modules have an “almost functorial” property: Every map $f : M' \to M$ of $\Lambda$-modules can be lifted to a map $\theta : S' \to S$ between their separated covers. Moreover, if $f$ is one-to-one or onto, so is $\theta$. In addition, there is a natural extension of $\theta$ to a $\Gamma$-homomorphism $\theta^* : X' \to X$ of the $\Gamma$-modules $X'$ and $X$ generated by $S'$ and $S$ respectively. Thus we can think of $\theta^*$ as the “best approximation” to $f$ by a homomorphism of $\Gamma$-modules.

In this section, $\Lambda$ is a much more general ring than in the rest of the paper. In fact, $\Lambda$ is not necessarily local or commutative. We assume that $\Lambda$ is the pullback of a diagram analogous to (1.1.1), in which all rings are noetherian, $k$ is semisimple artinian, $\Gamma$ and $\Gamma$ are finitely generated modules over $\Lambda$ and $k$ respectively, and $\rho$ is a ring homomorphism. We do not assume that $\ker \rho = \text{rad} \Gamma$. (For more detail, see (4.1.1).)

We note that additional hypotheses about commutativity, completeness, or the radical would not make anything easier in this section. Moreover we will need some of the additional generality in the final paper of this series, when we consider the nonlocal situation.

Earlier versions of the results in this section were used to find all finitely generated modules over a class of rings that includes some commutative noetherian rings whose completions at nonsingular maximal ideals are strictly split Dedekind-like $[L1, L2, L3]$, and also includes the integral group ring $\mathbb{Z}G$, where $G$ is a nonabelian group of order $pq$ ($p, q$ distinct primes) $[K]$. These earlier versions, however, do not apply to unsplit Dedekind-like rings. But the basic idea of separated covers (called “separated representations” in this earlier work) applies to both nonfinitely generated and finitely generated modules, in this previous work. We do not know whether separated covers of nonfinitely generated modules exist, when $\Lambda$ is unsplit Dedekind-like. See Remarks 4.8.

§5 Isomorphism as matrix problem, unsplit case. In this section we return to unsplit Dedekind-like rings and use separated covers to reformulate the
problem of classifying isomorphism classes of finitely generated \( \Lambda \)-modules as a matrix problem involving \( k \) and \( F \), the residue fields of \( \Lambda \) and its normalization \( \Gamma \), respectively. In this matrix problem, each \( \Lambda \)-module is represented by a \( \Gamma \)-module and a pair of matrices \((A, B)\) with entries in \( F \); and two modules turn out to be isomorphic if and only if they have the same associated \( \Gamma \)-module and their corresponding matrix pairs can be transformed into each other by means of left multiplication by matrices in \( \text{GL}(m, k) \) and right multiplication by matrices in a certain subgroup of \( \text{GL}(n, F) \). A precise statement of these matrix operations is the last result of this section.

\( \S 6 \) Isomorphism as matrix problem, strictly split case. In this section we do the analogous reduction to a matrix problem in the strictly split case. Here — since \( \Gamma = k \oplus k \), a direct sum of two rings — we deal with matrix 4-tuples \((A_1, A_2, B_1, B_2)\) with entries in the field \( k \), instead of pairs as in \( \S 5 \), and the matrix operations are more complicated. But the matrix problem is easier to solve because all matrix operations involve the single field \( k \).

\( \S 7 \) Solution of matrix problem, strictly split case. In this section we recall the solution of this “strictly split” matrix problem, from [KL0]. The results of \( \S 6 \) and \( \S 7 \) were previously obtained in [L1, L3], so we omit most of the details, concentrating on making the terminology and notation consistent with that of the present paper. But we do note that these theorems can be more cleanly proved by quoting [KL0].

\( \S 8 \) Solution of matrix problem, unsplit case. In this section we solve the matrix problem in the unsplit case. When \( \Lambda \) is a \( k \)-algebra, we can tensor it by \( F \). This converts the \( k-F \) matrix problem, in the unsplit case, to the split matrix problem over \( F \) alone. It can then be solved by using the main result in the split case — with \( F \) in place of \( k \). The details of this change of scalars argument use Galois theory, and therefore require our hypothesis that \( F \) be separable \( k \).

Of course, one of the main purposes of this series of papers is to not require that \( \Lambda \) be a \( k \)-algebra (because we want to deal with rings of algebraic integers). To deal with this more general situation, we take the unsplit \( k-F \) matrix problem out of context, allowing us to reinterpret the problem as a problem about modules over \( k \)-algebras. We then solve the problem by reducing it to the split case, getting exactly the same answer as if \( \Lambda \) had been a \( k \)-algebra. The main results of this section give the answer to this \( k-F \) matrix problem.

\( \S 9 \) Proofs: Indecomposable modules, unsplit case. In this section we use the solution of the unsplit matrix problem, given in \( \S 8 \), to prove that the structure of indecomposable \( \Lambda \)-modules is as described in \( \S 2 \).

\( \S 10 \) Proofs: Indecomposable modules, split case. In this section we use the solution of the split matrix problem, given in \( \S 7 \), to prove that the structure of indecomposable \( \Lambda \)-modules is as described in \( \S 3 \). As in previous
sections about the split case, we omit most of the details since these are not new results, and we concentrate on making the terminology and notation consistent with that of the rest of this paper.

§11 Klein rings and homomorphic images of Dedekind-like rings. In this section we show that, if $\Lambda$ is a Klein ring, then $\Lambda$ is quasi-Frobenius, with a unique minimal ideal modulo which $\Lambda$ becomes a homomorphic image of a strictly split Dedekind-like ring. Thus every module over a Klein ring is the direct sum of a free $\Lambda$-module and a module over a strictly split Dedekind-like ring. It follows immediately that Klein rings are tame.

We also show how to apply the previous results (about modules over Dedekind-like rings) to modules over homomorphic images of Dedekind-like rings. Although these descriptions follow from the previous results in this paper, some work is required, because our description of $\Lambda$-modules involves both $\Lambda$ and $\Gamma$, and translating this description to $\Lambda/I$-modules is complicated by the fact that the ideal $I$ of $\Lambda$ might not be an ideal of $\Gamma$.

§12 Examples. We review some previously known examples of rings whose completions are Dedekind-like, and establish two new examples: The quadratic order $\mathbb{Z}[\sqrt{n}]$ for every square-free integer $n$, and all subrings of square-free index in $\mathbb{Z}((n))$, where $n$ is any positive integer. More precisely, we show that all completions of these rings are either DVRs or else unsplit or strictly split Dedekind-like.

§13 Terminological index. An index of definitions and named theorems.

Remark 1.4 (Typographical error in [L1]). The structure of $\Lambda$-modules of finite length, in the strictly split case, was worked out in [L1]. Unfortunately, a nonrepeatedness condition was omitted when block cycle indecomposable modules were discussed in the introduction to that paper, [L1, p. 68]. However, the theorem is correctly stated in the paper itself [L1, Theorem 8.2 and the preceding paragraph]. This omission is doubly unfortunate because it was also made in [NR], although subsequently corrected in [NRSB].

Since the rings $\Lambda$ considered in the present paper vary from section to section, each section begins by stating what class of rings $\Lambda$ can represent. When the setting is not left-right symmetric, “module” means “left module” unless otherwise stated.

We extend special thanks to Markus Schmidmeier for pointing out to us a serious oversight in the penultimate version of this paper.

2. Indecomposable modules, unsplit case.

In this section we describe (but do not prove) how to construct all indecomposable finitely generated $\Lambda$-modules from indecomposable (necessarily uniserial) finitely generated $\Gamma$-modules, in the unsplit case. We also describe, very briefly, how homomorphisms of $\Lambda$-modules arise from homomorphisms of $\Gamma$-modules. See §9 for proofs of the main results of this section.
Notation 2.1. In this section \((\Lambda, m, k)\) is unsplit Dedekind-like with normalization \((\Gamma, m, F)\). Recall that \(\rho\) denotes the natural surjection \(\Gamma \to \Gamma / m = F\), and let \(\bar{\rho}: \Gamma \to F\), the *conjugate* of \(\rho\), be the composition of \(\rho\) with the nonidentity \(k = \Lambda / m\)-automorphism of the 2-dimensional Galois extension \(F\) of \(k\). Similarly, for \(\alpha \in F\) the notation \(\bar{\alpha}\) (the *conjugate* of \(\alpha\)) denotes the image of \(\alpha\) under this automorphism, and for a matrix \(A\) over \(F\), \(A\) denotes the matrix whose entries are the conjugates of those of \(A\).

Let \(t\) be a positive integer or \(\infty\). For each such \(t\) let \(\rho: \Gamma / m^t \to F\) denote the map induced by the map called \(\rho\) in the preceding paragraph. We set \(m^\infty = 0\). Thus \(\Gamma / m^t\) is always a uniserial \(\Gamma\)-module, and has infinite length when \(t = \infty\).

Recall that we chose a standard \(\Gamma\)-generator \(\pi\) of \(m\) in Notation 1.1. When \(t\) is finite, note that, once \(\pi\) has been chosen, there is a natural \(F\)-linear isomorphism \(\sigma: F \cong m^t / m^t - 1 / m^t\) defined by \(\sigma(\gamma + m) = \gamma t^{-1} + m^t\). We usually regard this map as an identification, in which case we have \((\Gamma / m^t)\pi t^{-1} = m^t / m^t = F\), the standard copy of \(F\) in \(\Gamma / m^t\). This copy of \(F\) is the \(\Gamma\)-socle of \(\Gamma / m^t\). The following simple fact will be used many times.

(2.1.1) When \(t \neq \infty, \neq 1\), the standard copy of \(F\) in \(\Gamma / m^t\) satisfies \(\rho(F) = 0\).

Notation 2.2 (Diagrams). Our constructions of indecomposable finitely generated \(\Lambda\)-modules begin with a nonempty direct sum \(X\) of nonzero uniserial \(\Gamma\)-modules and a diagram \(\mathcal{D}\) associated with \(X\), after which we define a \(\Lambda\)-module \(M(\mathcal{D})\) associated with \(\mathcal{D}\). The module \(\Gamma\)-module \(X\) has the form:

\[
X = (\Gamma / m^{i_1})^{(m)} \oplus (\Gamma / m^{j_1})^{(m)} \oplus \ldots \oplus (\Gamma / m^{i_d})^{(m)} \oplus (\Gamma / m^{j_d})^{(m)}
\]

where the superscript \((m)\) denotes “direct sum of \(m\) copies of,” and where the brackets at the end indicate that, in some cases, the final term is not present. We call \(m\) the *block size* of \(\mathcal{D}\) and of \(X\). (For examples of the diagrams that we are referring to, see (2.4.1) and (2.6.2).)

We now give the set of rules for forming and interpreting diagrams. Let \(X\) be given, as in (2.2.1). Each block of summands \((\Gamma / m^t)^{(m)}\) is represented, in our diagrams, by a vertical bar with the *length label* \(t\) written over it, as shown in the diagram (2.2.2). Let \(\mathcal{D}_0\) denote this diagram.

\[
\begin{array}{ccc}
\mathcal{D}_0 : & i_1 & j_1 & \ldots & i_d & j_d \\
\end{array}
\]

We define \(\mathcal{D}\) to be any diagram that can be formed from \(\mathcal{D}_0\) by a finite number of applications of the following four operations, each of which attaches an “edge” to these vertical bars.

*Top-glue.* Choose some pair of vertical bars, neither of whose tops has an edge attached to it and connect the tops of these bars by an edge, as shown
in the two “top-glue” diagrams in (2.2.3). Then label the left or right end of this edge with an invertible $m \times m$ matrix $U$ over $F$, as shown. We view $U$ as being attached to the top of the corresponding vertical bar, as well as to the gluing edge.

\[
\begin{array}{ccc}
\text{top-glue} & \text{top-glue} & \text{top-reduce} \\
(2.2.3) & \begin{cases}
(U) \\
(U) \\
(U)
\end{cases}
\end{array}
\]

In the case of the first top-glue diagram in (2.2.3), replace the (external) direct sum $(\Gamma/m^i)^{(m)} \oplus (\Gamma/m^j)^{(m)}$ by the $\Lambda$-submodule given by the following pullback:

\[
\{ (x, y) \in (\Gamma/m^i)^{(m)} \oplus (\Gamma/m^j)^{(m)} \mid \rho(x) = \overline{\rho(y)} \cdot U \}.
\] (2.2.4)

(One applies $\rho$ to a tuple by applying $\rho$ to each entry.) The reason for considering $U$ to be attached to the $i$-labeled (rather than $j$-labeled) vertical bar is that, since $U$ is invertible, the set of ordered pairs $(\rho(x), \rho(y))$ that arise from (2.2.4) is the set of all pairs

\[
\{ (\alpha U, \alpha) \in F^m \oplus F^m \mid \alpha \in F^m \}.
\] (2.2.5)

This will be consistent with our definition of bottom-gluing, below, and with the matrix pairs that will appear when we prove the main theorems of this section.

In the case of the second top-glue diagram, replace the condition $\rho(x) = \overline{\rho(y)} \cdot U$ in (2.2.4) by $\overline{\rho(x)} \cdot U = \rho(y)$. The effect of this is to move $U$ to the “$j$” side of the equal sign in (2.2.5).

If $U = I$ we usually do not explicitly display it.

To explain our view of what has happened in this operation, first note that the pullback in (2.2.4) contains $(m/m^i)^{(m)} \oplus (m/m^j)^{(m)}$; that is, it contains all but the topmost part of $(\Gamma/m^i)^{(m)} \oplus (\Gamma/m^j)^{(m)}$. We think of the residue module

\[
\frac{(\Gamma/m^i)^{(m)} \oplus (\Gamma/m^j)^{(m)}}{(m/m^i)^{(m)} \oplus (m/m^j)^{(m)}} = F^{(m)} \oplus F^{(m)}
\] (2.2.6)

(canonical isomorphism via $\rho$) as the “top” of $(\Gamma/m^i)^{(m)} \oplus (\Gamma/m^j)^{(m)}$.

If $U = I_m$, then forming the pullback keeps only half of $F^{(m)} \oplus F^{(m)}$ in (2.2.6), namely the set of ordered pairs of the form $(\alpha, \overline{\alpha})$. A general invertible matrix $U$ “twists” the half of $F^{(m)} \oplus F^{(m)}$ that we are keeping.

**Top-reduce.** Choose some vertical bar whose top has no attached edge. Say its label is $i$. Attach an edge to this bar — displayed in diagram (2.2.3) by a short thick horizontal line segment — and label the edge with an invertible $m \times m$ matrix $U$ over $F$, as shown. Then replace the corresponding
summand \((\Gamma/\mathfrak{m}^t)^{(m)}\) of \(X\) by its \(\Lambda\)-submodule

\[
\{ x \in (\Gamma/\mathfrak{m}^i)^{(m)} \mid \rho(x) \in k^{(m)} U \}.
\]

As before, we usually do not display \(U\) if \(U = I\), and we consider \(U\) to be attached to the vertical bar as well as to the reduction edge.

To understand what has happened here, note that the module in \((2.2.7)\) contains the submodule \((\mathfrak{m}/\mathfrak{m}^t)^{(m)}\) of \((\Gamma/\mathfrak{m}^t)^{(m)}\) — all but the top, as before. When \(U = I_m\) the replacement also keeps the half \(k^{(m)}\) of \(F^{(m)} = (\Gamma/\mathfrak{m}^t)^{(m)}/(\mathfrak{m}/\mathfrak{m}^t)^{(m)}\), the top of \((\Gamma/\mathfrak{m}^t)^{(m)}\). A general invertible matrix \(U\) again “twists” half we are keeping.

Let \(S(\mathcal{D})\) be the \(\Lambda\)-submodule of \(X\) that results from whatever top-gluing and top-reductions that have been done. (The various matrix labels \(U\) do not need to be the same.) Our discussion of which parts of the original summands of \(X\) remain intact by top-gluing and top-reduction shows:

\[
(2.2.8) \ S(\mathcal{D}) \text{ contains the standard copy of } F^{(m)} \text{ in every summand } (\Gamma/\mathfrak{m}^t)^{(m)} \text{ of } X \text{ for which } t \neq \infty, \neq 1.
\]

Our remaining two operations on diagrams are shown symbolically below. As before we do not usually display \(U\) if \(U = I\), and we regard \(U\) as being attached to the appropriate vertical bar as well as to the gluing or reduction edge.

\[
\begin{array}{ccc}
\text{bottom-glue} & \text{bottom-glue} & \text{bottom-reduce} \\
\begin{array}{c}
\hat{j} \\
\begin{array}{c}
U \\
\end{array}
\end{array} & \begin{array}{c}
\hat{i} \\
\begin{array}{c}
(\hat{U}) \\
\end{array}
\end{array} & \begin{array}{c}
i \\
\begin{array}{c}
(\hat{U}) \\
\end{array}
\end{array}
\end{array}
\]

**Bottom-glue.** Choose some pair of vertical bars whose bottoms have no attached edges and whose length labels \(j, i\) are neither \(\infty\) nor 1. Connect the bottoms of these bars with an edge, in the form of an elongated equal sign (as shown), and label the left or right end of this edge with an invertible \(m \times m\) matrix \(U\) over \(F\), as shown in \((2.2.9)\). Since \(j \neq \infty\) and \(i \neq \infty\), the module \((\Gamma/\mathfrak{m}^j)^{(m)} \oplus (\Gamma/\mathfrak{m}^i)^{(m)}\) contains our standard copy of \(F^{(m)} \oplus F^{(m)}\). In the case of the first diagram in \((2.2.9)\), form the \(\Lambda\)-submodule \((2.2.5)\) of \(F^{(m)} \oplus F^{(m)}\), and call this the **bottom-gluing module** associated with the diagram. In the case of the second diagram in \((2.2.9)\), move \(U\) to the “\(i\)” side of the comma in \((2.2.5)\), and again call the resulting collection of ordered pairs the **bottom-gluing module** associated with the diagram.

**Bottom-reduce.** Choose some vertical bar whose bottom has no attached edge and whose label is neither \(\infty\) nor 1. Attach an edge to the bottom of this bar, displayed as a short thick horizontal line segment, and label this edge with an \(m \times m\) invertible matrix \(U\) over \(F\), as shown. Then form the
following \( \Lambda \)-submodule of \( F^{(m)} = (\mathfrak{m}^{i-1}/\mathfrak{m}^i)^{(m)} \)

\[ k^{(m)} \cdot U \]

(2.2.10)

and call it the bottom-reduction module associated with the diagram.

Let \( \mathcal{D} \) be the diagram that results from any such bottom and top operations, and let \( K(\mathcal{D}) \) be the (necessarily direct) sum of the bottom-gluing and the bottom-reduction modules thus formed. Then let

\[ M(\mathcal{D}) = S(\mathcal{D})/K(\mathcal{D}) \]

(2.2.11)

which we call the \( \Lambda \)-module associated with \( \mathcal{D} \). For (2.2.11) to make sense we need to prove that \( K(\mathcal{D}) \subseteq S(\mathcal{D}) \), but this follows from (2.2.8).

In order to see what passing modulo \( K(\mathcal{D}) \) does, first consider the case that every \( U \) equals \( I_m \). Then every bottom-gluing amalgamates one standard copy of \( F^{(m)} \) with another; but \( \alpha \) in one summand is identified with \(-\alpha\) in the other. Every bottom-reduction, on the other hand, reduces \( S(\mathcal{D}) \) modulo the copy of \( k^{(m)} \) in the standard copy of \( F^{(m)} \). When \( U \neq I_m \), basically the same thing happens (since \( U \) is invertible), but it is twisted by \( U \).

The \( \Lambda \)-modules \( M(\mathcal{D}) \) thus formed are not always indecomposable. Connectivity of diagram \( \mathcal{D} \) is clearly a necessary condition for indecomposability, but is not sufficient.

Reversals. We close these introductory definitions by explicitly noting the effect of moving a matrix label \( U \) from one end of a (bottom or top) gluing edge to the other. Top-gluing and bottom-gluing — with \( U \) attached to the left-hand end of the edge — involve submodules of \( F^{(m)} \oplus F^{(m)} \) consisting of all elements of the form \( (\alpha U, \overline{\alpha}) \), which can be rewritten in the form \( (\alpha U, (\overline{\alpha U}) U^{-1}) = (\beta, \overline{\beta} U^{-1}) \). We shall use this in the following form:

(2.2.12) Let \( U \) be the matrix label attached to one end of a gluing edge in \( \mathcal{D} \). Then moving \( U \) to the opposite end of that edge and then replacing \( U \) by \( \overline{U}^{-1} \) leaves \( M(\mathcal{D}) \) (not just its isomorphism class!) unchanged.

**Notation 2.3** (Sequence manipulation). As in Notation 8.6, let \( \mu \) be the mirror image permutation, the permutation that reverses the order of a finite sequence. Thus \( \mu\{i_1, i_2, \ldots, i_d\} = \{i_d, \ldots, i_2, i_1\} \). Let \( \nu \) be the unit forward rotation defined by \( \nu\{i_1, i_2, \ldots, i_d\} = \{i_2, i_3, \ldots, i_d, i_1\} \). A cycle is any cyclic permutation of the form \( \nu^t \), where \( t \) is an integer. If \( I \) and \( J \) are finite sequences, we let \( \{I, J\} \) denote the concatenation of \( I \) and \( J \), that is, the sequence consisting of the terms of \( I \) followed by those of \( J \). Thus \( \{I, \mu(J)\} \) denotes \( I \) followed by the mirror image of \( J \), and we have \( \mu(\{I, \mu(J)\}) = \{J, \mu(I)\} \).

The definitions, below, of our seven “standard diagrams” all make use of the pair of label sequences in decomposition (2.2.1):

(2.3.1) \[ I = \{i_1, i_2, \ldots, i_d\} \quad J = \{j_1, j_2, \ldots, j_{d-1}, [j_d]\} \]
Thus every $i_\nu$ and $j_\nu$ is a positive integer or $\infty$, and the brackets around $j_d$ indicate that $j_d$ sometimes does not occur.

We now define some connected diagrams whose associated $\Lambda$-modules turn out to be indecomposable. We reserve the term *standard diagram* for those diagrams defined in 2.4 and 2.6.

**Definitions 2.4** (Standard diagrams, block size 1). The three standard diagrams in this series occur only with block size 1. The matrix labels $U$ are all one-by-one identity matrices, and are therefore not displayed. (Thus, we must learn from context that the block size is 1.)

\[
\begin{align*}
\mathcal{D}_{Nrd} &: \quad \begin{array}{cccc}
  i_1 & j_1 & i_2 & j_2 & \cdots & i_d & j_d
\end{array} \\
\mathcal{D}_{Brd} &: \quad \begin{array}{cccc}
  i_1 & j_1 & i_2 & j_2 & \cdots & i_d & j_d
\end{array} \\
\mathcal{D}_{Trd} &: \quad \begin{array}{cccc}
  i_1 & j_1 & i_2 & j_2 & \cdots
\end{array}
\end{align*}
\]

We now complete these definitions.

*Nonreduced diagram, $\mathcal{D}_{Nrd}$. Here the following conditions must be satisfied.*

(2.4.2) (i) The block size is $m = 1$.

(ii) Only length labels $i_1$ and $j_d$ can equal $\infty$ or 1.

(iii) The pair of label sequences $I$ and $J$ must be *unsymmetrical* in the sense that $J \neq \mu(I)$.

Since the one-by-one identity matrix is its own conjugate inverse, the following is an immediate consequence of (2.2.12):

(2.4.3) Replacing the pair of label sequences $I$ and $J$ in diagram $\mathcal{D}_{Nrd}$ by $\mu(J)$ and $\mu(I)$, respectively — that is, drawing the left-right mirror image of the diagram — does not change the isomorphism class of $M(\mathcal{D}_{Nrd})$.

In the extreme case $d = 1$, no bottom-gluing occurs. When condition (2.4.2)(iii) fails, $M(\mathcal{D}_{Nrd})$ becomes the direct sum of two indecomposable modules [Proposition 9.5].

*Bottom-reduced and top-reduced diagram, $\mathcal{D}_{Brd}$ and $\mathcal{D}_{Trd}$. The restrictions here are:*

(2.4.4)

(i) The block size is $m = 1$.

(ii) Only length label $i_1$ can equal $\infty$ or 1.

Note that $j_d$ does not occur in the top-reduced diagram $\mathcal{D}_{Trd}$. In the extreme case $d = 1$, $\mathcal{D}_{Trd}$ consists of a single top-reduced vertical bar.

Before proceeding, we note that there are other possibilities for diagrams with block size 1. Some of these [diagrams (2.6.2)] yield indecomposable
modules in all block sizes, as we shall see, while others are not standard diagrams, even if they are connected. (See, for example, (2.5.1) and Remark 2.10.)

**Examples 2.5.** The first two diagrams in (2.5.1) are the standard diagrams whose associated Λ-modules are Λ and its residue field $k$, respectively.

\[
\begin{align*}
\Lambda : & \quad \begin{array}{c}
\infty \\
\uparrow
\end{array} \\
\kappa : & \quad \begin{array}{c}
1 \\
\uparrow
\end{array} \\
k \text{ (illegal)} : & \quad \begin{array}{c}
1}
\end{array}
\end{align*}
\]

The third diagram in (2.5.1) is illegal because the definition does not allow bottom reduction on a vertical bar with length-label 1. Disregarding this for a moment, note that the Λ-module associated with a single vertical bar with length-label 1 is $\Gamma/\mathfrak{m} = F$, which is isomorphic to $k \oplus k$ as a Λ-module. Thus either top-reducing this module or bottom-reducing this module will result in a Λ-module isomorphic to $k$. Allowing only one of these to be standard associates a unique standard diagram with the Λ-module $k$. Moreover, this choice is made in such a way that Theorem 2.11 is true for the Λ-module $k$.

We have not investigated whether or not the theorem is true with respect to the second diagram for $k$.

Caution: As just mentioned, the Λ-module corresponding to single vertical bar with length label 1 and no top or bottom reduction is $\Gamma/\mathfrak{m} = F$. But this is not the standard diagram for $\Gamma/\mathfrak{m}$, because it is not among the diagrams displayed in (2.4.1) and (2.6.2). The standard diagram for $\Gamma/\mathfrak{m}$ is displayed in (2.10.1). Again, the reason for this seeming peculiarity is that Theorem 2.11 holds for this choice of diagrams.

**Definitions 2.6** (Standard diagrams, arbitrary block size). The remaining four standard diagrams occur with arbitrary block size. Before stating details, we mention some important differences between these and the standard diagrams of block size 1. (i) None of the length labels is $\infty$ or 1; in particular, the associated Λ-modules $M(D)$ all have finite length. (ii) Both the top and bottom of every vertical bar is either glued or reduced. (iii) The additional conditions that guarantee indecomposability are more complicated. (iv) Each diagram explicitly displays one invertible $m \times m$ matrix label, $U$ or $U^{-1}$ over $F$, which we call the blocking matrix of the diagram. The reason for sometimes using $U$ and sometimes $U^{-1}$ is to achieve uniformity in the uniqueness formulas in Theorem 2.8.

According to our convention, the remaining matrix labels are identity matrices, and are not displayed explicitly. The placement of the blocking matrix in the diagram is somewhat arbitrary. (See Proposition 2.9 about moving $U$.)

We require the following definition.
(2.6.1) A finite sequence $W$ is repetition-free if there is no strictly shorter sequence $W'$ such that $W$ consists of repetitions of $W'$.

The following are the standard diagrams associated with arbitrary block size.

**Bottom-bottom-reduced diagrams, $\mathcal{D}_{BBrd}$.** We require the following conditions:

(2.6.3) (i) $UU^{-1}$ is indecomposable under similarity.
(ii) None of the length labels is $\infty$ or 1.
(iii) The concatenated sequence $\{I, \mu(J)\}$ is repetition-free [see (2.6.1)].

Although it is not obvious, for any positive integer $n$, there exist $n \times n$ matrices $U$ satisfying condition (i). (See Remark 2.12.) Note that, when $d = 1$, no bottom-gluing occurs.

**Bottom-top-reduced diagrams, $\mathcal{D}_{BTrd}$.** We require conditions (2.6.3). Note that $j_d$ does not occur in this diagram. When $d = 1$ the diagram becomes a single vertical bar, reduced both at the bottom and the top.

**Top-top-reduced diagrams, $\mathcal{D}_{TTrd}$.** We require conditions (2.6.3) here, too. When $d = 1$ no top-gluing occurs.

**Cycle diagrams, $\mathcal{D}_{Cy}$.** We require:

(2.6.4) (i) The blocking matrix $U$ is indecomposable under similarity.
(ii) Either the blocking matrix $U$ is not similar to $U^{-1}$ or, for all cycles $\nu^t$, $J \neq \nu^t \mu(I)$.
(iii) None of the length labels is $\infty$ or 1.
(iv) The sequence of pairs $\{(i_1, j_1), \ldots, (i_d, j_d)\}$ is repetition-free.

When condition (ii) is not satisfied, the resulting module becomes the direct sum of two indecomposable modules [Proposition 9.5]. Unlike the previous types, there is nothing exceptional about the case $d = 1$. 

---

**Diagram Notes:**
- $\mathcal{D}_{BBrd}$: 
  \[
  \begin{array}{cccc}
  i_1 & j_1 & i_2 & j_2 \\
  \end{array} \ldots \begin{array}{cc}
  i_d & j_d \\
  \end{array} \begin{array}{c}
  (U^{-1}) \\
  \end{array}
  \]

- $\mathcal{D}_{BTrd}$: 
  \[
  \begin{array}{cccc}
  i_1 & j_1 & i_2 & j_2 \\
  \end{array} \ldots \begin{array}{cc}
  i_d & j_d \\
  \end{array} \begin{array}{c}
  (U) \\
  \end{array}
  \]

- $\mathcal{D}_{TTrd}$: 
  \[
  \begin{array}{cccc}
  i_1 & j_1 & i_2 & j_2 \\
  \end{array} \ldots \begin{array}{cc}
  i_d & j_d \\
  \end{array} \begin{array}{c}
  (U) \\
  \end{array}
  \]

- $\mathcal{D}_{Cy}$: 
  \[
  \begin{array}{cc}
  (U) \\
  \end{array} \ldots \begin{array}{cc}
  (U) \\
  \end{array}
  \]
In all four types, unless the repetition-freeness condition holds, the module might be decomposable, and uniqueness of $D$ can fail [Remark 9.7].

For square matrices $U, V$ over $F$ we write $U \sim V$ for “$U$ is similar to $V$.”

**Theorem 2.7.** Every indecomposable finitely generated $\Lambda$-module is isomorphic to $M(D)$ where $D$ is one of the standard diagrams in (2.4.1) or (2.6.2), and all such modules are indecomposable. Moreover, every indecomposable finitely generated $\Lambda$-module of infinite length is isomorphic to $M(D)$ for one of the standard diagrams $D$ in (2.4.1).

**Theorem 2.8.** Let $D$ and $D'$ be standard diagrams, with pairs of label sequences $I$ and $J$, and $I'$ and $J'$ respectively. If these diagrams have blocking matrices, call them $U$ or $U^{-1}$, and $V$ or $V^{-1}$ respectively, and assume that the blocking matrices are located as shown in Diagram (2.6.2).

Then $M(D) \cong M(D')$ if and only if $D$ and $D'$ are of the same type (i)-(vii) below, and the conditions listed for their type hold.

(i) (Nonreduced) Either $I' = I$ and $J' = J$, or $I' = \mu(J)$ and $J' = \mu(I)$.

(ii) (Bottom reduced) $I' = I$ and $J' = J$.

(iii) (Top reduced) $I' = I$ and $J' = J$.

(iv) (Bottom-bottom reduced) Either $I' = I$ and $J' = J$, or $I' = \mu(J)$ and $J' = \mu(I)$; and $V^{-1} \sim U^{-1}$.

(v) (Bottom-top reduced) $I' = I$ and $J' = J$, and $V^{-1} \sim U^{-1}$.

(vi) (Top-top reduced) Either $I' = I$ and $J' = J$, or $I' = \mu(J)$ and $J' = \mu(I)$; and $V^{-1} \sim U^{-1}$.

(vii) (Cycle) Either $I' = \nu^t(I)$ and $J' = \nu^t(J)$ (for some cyclic permutation $\nu^t$) and $V \sim U$; or $I' = \nu^t \mu(J)$ and $J' = \nu^t \mu(I)$ (for some cyclic permutation $\nu^t$) and $V \sim U^{-1}$.

We note that, in every case where the pair of label sequences is not uniquely determined by the isomorphism class of $M(D)$, the alternative pair corresponds to the geometric left-right or rotational symmetry of the diagram. However (according to the statement of the theorem), when applying this symmetry to the bars, edges, and length-labels, one leaves the location of the blocking matrix unchanged, in the position shown in standard diagrams (2.6.2).

As mentioned earlier, the placement of the blocking matrix in standard diagrams (2.6.2) is somewhat arbitrary. The precise rules for moving the blocking matrix — and appropriately modifying it — are the subject of our next result.

**Proposition 2.9 (Moving $U$).** Let $D$ be any standard diagram in the “arbitrary block size” family (2.6.2), except that the blocking matrix $U$ is located at an arbitrary position in the diagram.
Then each of the following operations on \( U \) leaves the isomorphism class of \( M(D) \) unchanged.

(i) Move \( U \) from one end (left or right) of any gluing edge to the other end of that edge, replacing it with \( U^{-1} \).

(ii) Move \( U \) from one end (top or bottom) of any vertical bar to the other end, replacing it with \( U^{-1} \).

In addition, if \( D \) is one of the standard diagrams with a pair of reduction edges, \( U \) can be moved unchanged from the top of any vertical bar to the top of any other vertical bar (or from bottom to bottom) without changing the isomorphism class of \( M(D) \).

For example, in the first of the diagrams in (2.6.2) — diagram \( D_{BBd} \) — we can move \( U^{-1} \) to the top of its attached vertical bar, replacing it with \( U \). The resulting diagram is now nonstandard, although the \( \Lambda \)-module that it represents is unchanged.

**Remark 2.10 (Efficient vs. standard diagrams).** Our standard diagrams handle some indecomposable \( \Lambda \)-modules inefficiently. This includes all of the indecomposable finitely generated \( \Gamma \)-modules. One might expect the standard diagram for \( \Gamma/m^j \) \((1 \leq j \leq \infty)\) to be the second diagram in (2.10.1), but it is the first diagram that is standard.

\[
\begin{array}{c}
\Gamma/m^j: \\
\includegraphics[width=0.5\textwidth]{standard.png} \\
\text{(standard)} \\
\end{array}
\]

\[
\begin{array}{c}
\Gamma/m^j: \\
\includegraphics[width=0.5\textwidth]{not_standard.png} \\
\text{(not standard).}
\end{array}
\]

To see that the second diagram is not standard, it suffices to note that it does not appear among the three standard “block size one” diagrams in (2.4.1). (Set \( d = 1 \), the smallest value of \( d \) that makes sense in each of these diagrams.) The first diagram in (2.10.1) is a nonreduced diagram; to see that it actually describes \( \Gamma/m^j \) is a simple verification using the definition (2.2.4) of top-gluing.

To see the general principle involved, first note that the only places that the label 1 can occur in a standard diagram are the left-hand end of the three diagrams in (2.4.1) [the “block size 1 only” diagrams], and the right-hand end of the first of these. It is then easy to see the following.

(2.10.2) If a vertical bar with length label 1 occurs in a standard diagram \( D \) in (2.4.1) with an attached top-gluing edge, then the diagram \( D' \) that results from deleting the vertical bar and attached edge satisfies \( M(D') \cong M(D) \). But \( D' \) is not a standard diagram.

Making these deletions gives what we call the **efficient** representation of the module involved, as opposed to the “standard” representation of that module.

If the only purpose of this paper were to describe finitely generated \( \Lambda \)-modules, we would use the efficient representation. A minor disadvantage of
this would be that we would have to increase the number of types of standard diagrams by including, for example, diagrams whose left-most gluing edge is a bottom-gluing edge. But this minor disadvantage would be more than offset by the fact that not having an irrelevant vertical bar makes $M(D')$ easier to visualize.

The real advantage of keeping the “irrelevant” edge has to do with our Theorem 2.11 about morphisms in mod-$\Lambda$: We do not know whether the theorem applies when standard representations are replaced by efficient ones. Note that if $D$ is the disjoint union of diagrams $D_1, \ldots, D_m$ then $M(D) = M(D_1) \oplus \ldots \oplus M(D_m)$. Therefore (by the theorems stated so far in this section) every finitely generated $\Lambda$-module is isomorphic to $M(D)$ where $D$ is a disjoint union of standard diagrams. We can now state our main theorem on morphisms in mod-$\Lambda$, which shows that all such morphisms arise from homomorphisms of modules over the DVR $\Gamma$.

**Theorem 2.11** (\(\Lambda\)-homomorphisms versus \(\Gamma\)-homomorphisms). Let $D'$ and $D$ be disjoint unions of standard diagrams, and let $f: M(D') \rightarrow M(D)$ be a $\Lambda$-homomorphism. Then there is a $\Gamma$-homomorphism $f^{**}: X(D') \rightarrow X(D)$ whose restriction $f^*$ to $S(D')$ is a $\Lambda$-homomorphism: $S(D') \rightarrow S(D)$ that takes $K(D') \rightarrow K(D)$ and induces $f$. Moreover, if $f$ is one-to-one or onto, then so is any such $f^*$.

As we shall see in the proof of this theorem, the natural surjection $S(D) \twoheadrightarrow M(D)$ is a separated cover of $M(D)$, that is, our “best approximation” to $M(D)$ by a $\Lambda$-submodule of some $\Gamma$-module. See \S4 for properties of separated covers.

**Remark 2.12** (Indecomposability of $U U^{-1}$). Condition (2.6.3)(i) on the indecomposability of $U U^{-1}$ under similarity, and Condition (2.6.4)(ii) that $U$ not be similar to $U^{-1}$, fit in with Theorem 2.8 in a particularly interesting way.

Let $U$ be an invertible $m \times m$ matrix over $F$, and suppose that $U$ is indecomposable under similarity. Thus $U$ can be the companion matrix of any power of any irreducible polynomial (other than $x$) in $F[x]$.

By (2.6.4)(ii), $U$ can be the blocking matrix of a cycle diagram $D_{CY}$ in (2.6.2) except if $U \sim U^{-1}$ and the label sequences $\{i_\nu\}$ and $\{j_\nu\}$ fail to satisfy a certain nonsymmetry condition. When these restrictions are satisfied, it is the similarity class of $U$ that is an isomorphism invariant of $M(D_{CY})$. However it is not obvious what similarity classes are excluded here.

A “Hilbert Theorem 90” for matrices, due to Ballantine [see Lemma 8.11], states that a matrix satisfies the condition $U \sim U^{-1}$ if and only if there is a matrix $V$ such that $VV^{-1} = U$. According to Condition (2.6.3)(i), such matrices $V$ are precisely the matrices that are allowed to be blocking matrices for the remaining three diagrams that have blocking matrices [see (2.6.2)].
Moreover, according to our structure Theorem 2.8, it is the similarity class of $V \bar{V}^{-1} = U$ that is an isomorphism invariant of $M(D)$ in these cases — precisely the similarity classes that are excluded from $M(D_{Cy})$!

To see what similarity classes of matrices we are talking about, let $U = C(f)$ be the companion matrix of a polynomial $f \in F[x]$, where $f = x^m + a_{m-1}x^{m-1} + \cdots + a_0$. Since $U$ is invertible in this discussion, we have $a_0 \neq 0$. Let $g$ be the monic polynomial $g = (x^m/a_0)f(1/x)$. Then it follows easily, with the help of the Cayley-Hamilton Theorem, that $U^{-1} \sim C(g)$, and hence

$$U \sim \bar{U}^{-1} \iff f = (x^m/a_0)\bar{f}(1/x).$$

(2.12.1)

Using this, it is easy to see that for any positive integer $m$, there exist $m \times m$ matrices $U$ satisfying conditions (2.12.1). For example, use $f(x) = (x+1)^m$.

3. Indecomposable modules, strictly split case.

In this section we describe (but do not prove) how to construct all indecomposable finitely generated $\Lambda$-modules from indecomposable (necessarily uniserial) finitely generated $\Gamma$-modules, in the strictly split case. This is similar to — but simpler than — the unsplit case. This structure was determined previously, in [L1, L3]. Therefore our focus, in this short section, is to make the terminology of that description consistent with the other terminology of the present paper. This will be especially important when we deal with the nonlocal situation, in the fourth paper of this series. We also describe, very briefly, how homomorphisms of $\Lambda$-modules arise from homomorphisms of $\Gamma$-modules. For proofs, see §10.

Notation 3.1. Throughout this section $(\Lambda, m, k)$ is a strictly split Dedekind-like ring, as in Notation 1.1. Thus the normalization of $\Lambda$ is $\Gamma = \Gamma_1 \oplus \Gamma_2$ where each $(\Gamma_\nu, m_\nu, k)$ is a DVR and $m = m_1 \oplus m_2$. Following the style of Notation 2.1, we denote by $\rho$ not only the map $\Gamma \to \Gamma = \Gamma/m = k \oplus k$ in pullback diagram (1.1.1), but also the maps $\Gamma_\nu/m_\nu \to k$ induced by the original map $\rho$. We also use $\rho$ for direct sums of such maps. Moreover, we let $m_\nu^\infty = 0$, so that $\Gamma_\nu = \Gamma_\nu/m_\nu^\infty$.

Recall, from Notation 1.1, that we have chosen a standard $\Gamma_\nu$-generator $\pi_\nu$ of the maximal ideal $m_\nu$ of each $\Gamma_\nu$.

When $t$ is finite, our choice of generators $\pi_\nu$ yields natural $k$-linear isomorphisms $\sigma: k \cong m_\nu^{t-1}/m_\nu^t$ defined by $\sigma(\gamma_\nu + m_\nu) = \gamma_\nu \pi_\nu^{t-1} + m_\nu^t$. We usually regard this map as an identification, in which case we have $(\Gamma_\nu/m_\nu^t)\pi_\nu^{t-1} = m_\nu^{t-1}/m_\nu^t = k$, the standard copy of $k$ in $\Gamma_\nu/m_\nu^t$. This copy of $k$ is the $\Gamma_\nu$-socle of $\Gamma_\nu/m_\nu^t$. The following simple fact will be used many times.

(3.1.1) When $t \neq \infty$, $\neq 1$, the standard copy of $k$ in $\Gamma_\nu/m_\nu^t$ satisfies $\rho(k) = 0$.

Notation 3.2 (Diagrams). As in the unsplit case, our constructions of indecomposable finitely generated $\Lambda$-modules begin with a nonempty direct
sum $X$ of nonzero uniserial $\Gamma$-modules and a diagram $D$ associated with $X$, after which we define a $\Lambda$-module $M(D)$ associated with $D$. The module $\Gamma$-module $X$ has the form:

$$X = X_1 \oplus X_2 \text{ where}$$

$$X_1 = (\Gamma_1/m_i^1)(m) \oplus \ldots \oplus (\Gamma_1/m_i^d)(m)$$

$$X_2 = (\Gamma_2/m_j^1)(m) \oplus \ldots \oplus (\Gamma_2/m_j^d)(m).$$

Here each $i_\nu$ and $j_\nu$ is a positive integer or $\infty$, and the exponent $(m)$ denotes “direct sum of $m$ copies of.” Note that the same number $md$ of indecomposable summands occurs in both $X_1$ and $X_2$ in (3.2.1). We call $m$ the block size of $D$ and of $X$. For examples of diagrams, see (3.3.1). We now give the set of rules for forming and interpreting diagrams.

Let $X$ be given, as in (3.2.1). Each block of summands $(\Gamma_\nu/m_i^\nu)(m)$ is represented, in our diagrams, by a vertical bar with a length label $t$ written over it and the index $\nu$ written below it, as shown in (3.2.2). Let $D_0$ denote this diagram.

$$D_0 : \begin{array}{ccc} i_1 & j_1 & i_d & j_d \\ 1 & 2 & 1 & 2 \end{array}.$$

We define $D$ to be any diagram that can be formed from $D_0$ by a finite number of applications of the following two operations, each of which attaches an “edge” to a pair of vertical bars.

$$\begin{array}{c} \text{top-glue} \\ \begin{array}{cc} i & j \\ 1 & 2 \end{array} \begin{array}{cc} (U) \\ 1 & 2 \end{array} \end{array} \quad \begin{array}{c} \text{top-glue} \\ \begin{array}{cc} i & j \\ 1 & 2 \end{array} \begin{array}{cc} (U) \\ 1 & 2 \end{array} \end{array}.$$

Top-glue. Choose a vertical bar with a 1 below it, and a vertical bar with a 2 below it, and suppose that neither of the tops of these bars has an edge attached; then connect the tops of these bars by an edge, as shown in (3.2.3). Then label the left or right end of this edge with an invertible $m \times m$ matrix $U$ over $k$, as shown. We view $U$ as being attached to the top of the corresponding vertical bar, as well as to the gluing edge.

In the case of the first top-glue diagram in (3.2.3), replace the (external) direct sum $(\Gamma_1/m_1^i)(m) \oplus (\Gamma_2/m_2^j)(m)$ by the $\Lambda$-submodule given by the following pullback.

$$\{(x, y) \in (\Gamma_1/m_1^i)(m) \oplus (\Gamma_2/m_2^j)(m) \mid \rho(x) = \rho(y) \cdot U\}.$$
(One applies $\rho$ to a tuple by applying $\rho$ to each entry.) The reason for considering $U$ to be attached to the $i$-labeled (rather than $j$-labeled) vertical bar is that, since $U$ is invertible, the set of ordered pairs $(\rho(x), \rho(y))$ that arise from (3.2.4) is the set of all pairs

$$\{(\alpha U, \alpha) \in k^{(m)} \oplus k^{(m)} \mid \alpha \in k^{(m)}\}.$$  

This will be consistent with our definition of bottom-gluing, below, and with the matrix pairs that will appear when we prove the main theorems of this section.

In the case of the second top-glue diagram, replace the condition $\rho(x) = \rho(y) \cdot U$ in (3.2.4) by $\rho(x) \cdot U = \rho(y)$. The effect of this is to move $U$ to the $j$-side of the equal sign in (3.2.5).

We note that, if $U = I$, we get the same module whether we attach $I$ to the $i$-side or the $j$-side of the top-gluing edge. In this case we usually do not explicitly display the matrix $I$.

To explain our view of what has happened in this operation, first note that the pullback in (3.2.4) contains $(\frac{m_1}{m_i})^{(m)} \oplus (\frac{m_2}{m_j})^{(m)}$; that is, it contains all but the topmost part of $(\frac{\Gamma_1}{m_i})^{(m)} \oplus (\frac{\Gamma_2}{m_j})^{(m)}$. We think of the residue module

$$\frac{(\Gamma_1/m_i)^{(m)}}{(m_1/m_i)^{(m)} \oplus (m_2/m_j)^{(m)}} = k^{(m)} \oplus k^{(m)}$$

(canonical isomorphism via $\rho$) as the “top” of $(\frac{\Gamma_1}{m_i})^{(m)} \oplus (\frac{\Gamma_2}{m_j})^{(m)}$.

If $U = I_m$, then forming the pullback keeps only half of $k^{(m)} \oplus k^{(m)}$ in (3.2.6), namely the set of ordered pairs of the form $(\alpha, \alpha)$. A general invertible matrix $U$ “twists” the half of $k^{(m)} \oplus k^{(m)}$ that we are keeping.

Let $S(\mathcal{D})$ be the $\Lambda$-submodule of $X$ that results from whatever top-gluing operations that have been done. (The various matrix labels $U$ do not need to be the same.) Our discussion of which parts of the original summands of $X$ remain intact by top-gluing shows:

(3.2.7) $S(\mathcal{D})$ contains the standard copy of $k^{(m)}$ in every summand $(\frac{\Gamma_\nu}{m_\nu})^{(m)}$ of $X$ for which $t \neq \infty, \neq 1$.

Our remaining operations on diagrams are shown symbolically below. As before we do not usually display $U$ if $U = I$, and we regard $U$ as being attached to the appropriate vertical bar as well as to the bottom-gluing edge.

\[
\begin{array}{c|c}
\text{bottom-glue} & \text{bottom-glue} \\
\hline
\begin{array}{c}
\begin{array}{c}
2 \\
1 \\
\end{array}
\end{array} & \begin{array}{c}
\begin{array}{c}
2 \\
1 \\
\end{array}
\end{array} \\
\end{array}
\]
**Bottom-glue.** Choose a vertical bar with 1 below it and a vertical bar with 2 below it, neither with length label 1 or ∞, and without any edge attached to the bottom of either of them. Connect the bottoms of these bars with an edge, in the form of an elongated equal sign (as shown), and label the left or right end of this edge with an invertible \(m \times m\) matrix \(U\) over \(k\), as shown in (3.2.8). Since neither length label is infinite, \((\Gamma^2/m_2^j(m) \oplus (\Gamma^1/m_1^i(m)\) contains our standard copy of \(k(m) \oplus k(m)\). In the case of the first diagram in (3.2.8), form the \(\Lambda\)-submodule (3.2.5) of \(k(m) \oplus k(m)\), and call it the **bottom-gluing module** associated with the diagram. In the case of the second diagram in (3.2.8), move \(U\) to the \(i\)-side of the comma in (3.2.5), and again call the resulting set of ordered pairs the **bottom-gluing module** associated with the diagram.

Let \(D\) be the diagram that results from any such bottom- and top-gluing operations, and let \(K(D)\) be the (necessarily direct) sum of the bottom-gluing modules thus formed. Then let

\[
M(D) = S(D)/K(D)
\]

which we call the **\(\Lambda\)-module associated with \(D\)**. For (3.2.9) to make sense we need to prove that \(K(D) \subseteq S(D)\), and this follows from (3.2.7).

In order to see what passing modulo \(K(D)\) does, first consider the case that every \(U\) equals \(I_m\). Then every bottom-gluing amalgamates one standard copy of \(k(m) \oplus k(m)\) with another. When \(U \neq I_m\), basically the same thing happens (since \(U\) is invertible), but it is twisted by \(U\).

The \(\Lambda\)-modules \(M(D)\) thus formed are not always indecomposable. Connectivity of diagram \(D\) is clearly a necessary condition for indecomposability, but is not sufficient.

We now define the two types of connected diagrams whose associated \(\Lambda\)-modules turn out to be indecomposable. We reserve the term **standard diagrams** for the diagrams defined in 3.3. These definitions make use of the pair of **label sequences** in decomposition (3.2.1):

\[
I = \{i_1, i_2, \ldots, i_d\} \quad J = \{j_1, j_2, \ldots, j_d\}.
\]

Thus every \(i_\nu\) and \(j_\nu\) is a positive integer or ∞.

**Definitions 3.3** (Standard diagrams). Our first standard diagram — a “deleted cycle” diagram — occurs only with block size 1. For deleted cycle diagrams, the matrix labels \(U\) are all one-by-one identity matrices and are therefore not displayed. The second diagram — a “block cycle” diagram — can occur with arbitrary block size. Each block cycle diagram explicitly displays only one invertible \(m \times m\) matrix label \(U\) over \(k\), called the **blocking matrix** of the diagram, while all remaining matrix labels (not displayed) are \(m \times m\) identity matrices. The placement of the blocking matrix in the
A diagram is somewhat arbitrary. (See Proposition 3.6 about moving $U$.)

$D_{DCy}$:

\[
\begin{array}{cccccc}
  & i_1 & j_1 & i_2 & j_2 & \ldots & i_d & j_d \\
1 & 2 & 1 & 2 & \ldots & 1 & 2
\end{array}
\]

(3.3.1)

$D_{BCy}$:

\[
\begin{array}{cccccc}
1 & (U) & 2 & 1 & 2 & \ldots & 1 & 2
\end{array}
\]

We now complete these definitions.

**Deleted cycle diagrams, $D_{DCy}$:** The following conditions must be satisfied.

(3.3.2) (i) The block size is $m = 1$.

(ii) Only labels $i_1$ and $j_d$ can equal $\infty$ or 1.

**Block cycle diagrams, $D_{BCy}$:** The following conditions must be satisfied.

(3.3.3) (i) The blocking matrix $U$ is indecomposable under similarity.

(ii) Length labels $\infty$ and 1 cannot occur.

(iii) The sequence of pairs $\{(i_1, j_1), \ldots, (i_d, j_d)\}$ is repetition-free.

(Recall, from (2.6.1), that a finite sequence $W$ is repetition-free if there is no strictly shorter sequence $W'$ such that $W$ consists of repetitions of $W'$.)

When condition (iii) is not satisfied, a larger block size can be used, and the module $M(D)$ might be decomposable.

**Examples 3.4.** The first two diagrams below are the standard diagrams whose associated $\Lambda$-modules are $\Lambda$ and $k$, respectively.

(3.4.1) $\Lambda : \begin{array}{ccc} \infty & \infty & \begin{array}{cc} 1 & 1 \end{array} \end{array}$

$k : \begin{array}{ccc} 1 & 1 \end{array}$ (nonstandard) : \begin{array}{cc} 1 & 1 \end{array} also: \begin{array}{c} 1 \end{array}

The third and fourth diagrams above are nonstandard diagrams whose associated $\Lambda$-modules are again $k$. These are the “efficient” diagrams for $k$, as discussed in Remark 3.7.

Let $I, J$ be as in (3.2.10); and let $I', J'$ be another pair of sequences, each again of length $d$. We say that $I, J$ and $I', J'$ are equal modulo simultaneous cyclic permutations if there is an integer $c$ such that $I' = \nu^c(I)$ and $J' = \nu^c(J)$. (See Notation 2.3.)

**Theorem 3.5.** If $\Lambda$ is strictly split Dedekind-like, then every indecomposable finitely generated $\Lambda$-module is isomorphic to $M(D)$ for some deleted cycle or block cycle diagram $D$, and all such modules are indecomposable.
Let $D, D'$ be diagrams of these types. (If either diagram has a blocking matrix, assume that this blocking matrix is located as shown in standard diagram (3.3.1).) Then $M(D) \cong M(D')$ if and only if one of the following holds.

(i) Both $D$ and $D'$ are deleted cycle diagrams, and $D = D'$, (that is, they have the same label sequences).
(ii) Both $D$ and $D'$ are block cycle diagrams, they have the same pairs of label sequences modulo simultaneous cyclic permutations, and their blocking matrices are similar.

The placement of the blocking matrix in block cycle diagrams is somewhat arbitrary. The precise rules for moving the blocking matrix are the subject of our next result.

**Proposition 3.6** (Moving $U$). Let $D$ be any block cycle diagram, except that the blocking matrix $U$ is located at an arbitrary position in the diagram. Then each of the following operations on $D$ leaves the isomorphism class of $M(D)$ unchanged.

(i) Move $U$ from one end (left or right) of any gluing edge to the other end of that edge, replacing it with $U^{-1}$.
(ii) Move $U$ from one end (top or bottom) of any vertical bar to the other end, replacing it with $U^{-1}$.

**Remark 3.7** (Efficient vs. standard diagrams). The discussion given in Remark 2.10 applies here also, with only one very minor change: The indecomposable $\Gamma$-modules are now $\Gamma_{i}/m_{i}$, rather than the modules $\Gamma/m_{i}$ displayed in (2.10.1). Thus, the standard and efficient diagrams for (say) the $\Lambda$-modules $\Gamma_{i}/m_{i}$ are, respectively, the first and second diagrams displayed below.

\[
\begin{array}{c}
\text{\(i\)} & \Gamma_{1}/m_{1} \text{ (standard)} & \begin{array}{c}
\begin{array}{c}
1 \\
2 \\
i
\end{array}
\end{array}
& \Gamma_{1}/m_{1} \text{ (efficient)} : \begin{array}{c}
\begin{array}{c}
1
\end{array}
\end{array}
\end{array}
\]

A special situation exists for the case $i = 1$. As already noted, the $\Lambda$-module $k$ corresponds to the second through fourth diagrams displayed in (3.4.1); and the third and fourth are both efficient diagrams for $k$. This holds because the two nonisomorphic $\Gamma$-modules $\Gamma_{1}/m_{1}$ and $\Gamma_{2}/m_{2}$ (both isomorphic to $k$ as rings) become isomorphic when considered as $\Lambda$-modules.

The way that homomorphisms of $\Lambda$-modules arise from homomorphisms of $\Gamma$-modules is given in our final result. (We do not know the extent to which this theorem holds if nonstandard diagrams are used.)

**Theorem 3.8.** Let $D'$ and $D$ be disjoint unions of block cycle and deleted cycle diagrams, and let $f: M(D') \to M(D)$ be a $\Lambda$-homomorphism. Then
there is a $\Gamma$-homomorphism $f^{**}: X(D') \to X(D)$ whose restriction $f^*$ to $S(D')$ is a $\Lambda$-homomorphism: $S(D') \to S(D)$ that takes $K(D') \to K(D)$ and induces $f$. Moreover, if $f$ is one-to-one or onto, then so is any such $f^*$.

4. Separated covers and almost functorial property.

Notation 4.1. The rings considered in this section are much more general than in the rest of this paper. In particular, rings in this section are not necessarily commutative.

In the fixed notation in this section we assume that we have the following commutative diagram of rings:

\[
\begin{array}{ccc}
\Lambda & \subseteq & \Gamma \quad (\mathfrak{r}\Gamma \text{ noetherian}) \\
\downarrow & & \downarrow \\
\Lambda & \subseteq & \Gamma \quad (\Lambda \text{ semisimple artinian}, \quad \Lambda \text{ finitely generated})
\end{array}
\]

(4.1.1)

where these rings satisfy the following conditions (i)–(v).

(i) The ring $\Gamma$ is left noetherian;

(ii) the map $\rho: \Gamma \to \Gamma$ and its restriction $\rho: \Lambda \to \Lambda$ are both surjective ring homomorphisms with the same kernel $C$;

(iii) the ring $\Lambda$ is semisimple artinian;

(iv) the $\Lambda$-module $\Gamma$ is finitely generated; and

(v) $\Lambda = \{ x \in \Gamma \mid \rho(x) \in \Lambda \}$, the pullback of this diagram.

Thus, diagram (4.1.1) is a conductor square defining $\Lambda$ as a subring of $\Gamma$, and $C$ is a conductor ideal for $\Lambda$ and $\Gamma$. Since the hypotheses on $\Lambda$ are not left-right symmetric, the term module means “left module” unless otherwise stated. We include some reminders of this near the beginning of this section.

Lemma 4.2. The ring $\Lambda$ is left noetherian.

Proof. First note that $\Lambda\Gamma$ is finitely generated since $\Lambda(\Gamma/\Lambda)$ is finitely generated. Let $L$ be any left ideal of $\Lambda$. Since $\mathfrak{r}\Gamma$ is noetherian, $\mathfrak{r}(C\Lambda)$ is finitely generated. Since $\Lambda\Gamma$ is finitely generated, so therefore is $\Lambda(C\Lambda)$. Thus it suffices to show that $\Lambda(L/C\Lambda)$ is finitely generated. Since $C$ is a 2-sided ideal of $\Lambda$, this last $\Lambda$-module is a module over the artinian ring $\Lambda = \Lambda/C$; and it is a submodule of the finitely generated $\Lambda$-module $(\Lambda\Gamma)/(C\Lambda)$, hence is itself finitely generated. □

The purpose of this section is to show how all finitely generated left $\Lambda$-modules and their homomorphisms can be built from $\Gamma$-modules and homomorphisms.

Definition 4.3. We call a (left) $\Lambda$-module $S$ separated — $\Gamma$-separated if additional precision is required — if $S$ is a $\Lambda$-submodule of some left $\Gamma$-module, say $X$. In this situation $\Gamma S$ denotes the $\Gamma$-submodule of $X$ generated by $S$. Unless $X$ is specified, $\Gamma S$ is not well-defined, not even up to isomorphism.
However, there is always a canonical choice for $\Gamma S$, namely $\Gamma \otimes_{\Lambda} S$. This follows from the following simple fact, which we record as a lemma for future use.

For examples of nonseparated modules, see Examples 4.16.

**Lemma 4.4.** The $\Lambda$-module $S$ is separated (if and) only if the natural map $S \rightarrow \Gamma \otimes_{\Lambda} S$ is one-to-one.

**Proof.** Let $X$ be any $\Gamma$-module containing $S$. Then we can consider the composite map $S \rightarrow \Gamma \otimes S \rightarrow X$, where the second arrow denotes the map $\omega \otimes s \rightarrow \omega s$. The nontrivial half of the lemma holds because the composite map is the identity on $S$. $\square$

When we write $\Gamma S = \Gamma \otimes S$, for a $\Lambda$-module $S$, we mean that the natural map from the right-hand side to the left-hand side is a bijection, and $\otimes = \otimes_{\Lambda}$. One easily checks that the following simple properties hold.

**Lemma 4.5.** Let $S, S'$ be separated $\Lambda$-modules, regarded as $\Lambda$-submodules of $\Gamma S = \Gamma \otimes S$ and $\Gamma S' = \Gamma \otimes S'$, respectively. Then:

(i) Every $\Lambda$-homomorphism $f: S \rightarrow S'$ can be uniquely extended to a $\Gamma$-homomorphism $\Gamma S \rightarrow \Gamma S'$ (namely to $1 \otimes f$).

(ii) If $B$ is any left ideal of $\Gamma$ contained in $\Lambda$, then $BS$ is a $\Gamma$-submodule of $S$.

_Caution._ Statement (ii) becomes false without the hypothesis that $S$ is separated. The point is that we have $\gamma(bs) = (\gamma b)s$ provided that both sides are defined, but unless $S$ is contained in some $\Gamma$-module, the left-hand side is undefined.

**Definition 4.6.** We define a separated cover of a (left) $\Lambda$-module $M$ to be a surjective $\Lambda$-module homomorphism $\phi: S \twoheadrightarrow M$ in which $S$ is separated and “as close as possible to $M$” in the following sense. If $S \twoheadrightarrow S' \twoheadrightarrow M$ is any factorization of $\phi$, and $S'$ is a separated $\Lambda$-module, then the map $S \rightarrow S'$ is one-to-one (and therefore an isomorphism). Thus, passing from $S$ to $S'$ gets no closer to $M$.

The starting point of our theory is the following triviality, since $\Lambda$ is left noetherian.

**Proposition 4.7.** Every finitely generated (left) $\Lambda$-module has a separated cover.

**Proof.** Let $\phi': F \twoheadrightarrow M$ be a homomorphism of any finitely generated separated (e.g. free) $\Lambda$-module onto the given $\Lambda$-module $M$.

Since the $\Lambda$-module $F$ is noetherian, it has a submodule $H$ that is maximal with respect to the property that $H \subseteq \ker(\phi')$ and $F/H$ is a separated $\Lambda$-module. Then the $\Lambda$-module homomorphism $\phi: S = F/H \twoheadrightarrow M$ is easily seen to be a separated cover of $M$. $\square$
Remarks 4.8. (i) If $\Lambda$ is strictly split Dedekind-like, then we can drop the “finitely generated” hypothesis in the previous lemma. In fact, this is true in the much more general (nonlocal) setting described in [L], and for the noncommutative integral group ring studied in [K]. But we do not know whether it is true in the present setting, (in particular, for unsplit Dedekind-like rings).

(ii) The remaining results in this section do not have finite generation hypotheses; that is, when separated covers of infinitely generated modules exist, they behave as described in the rest of this section.

Lemma 4.9. Let $\phi: S \to M$ be a $\Lambda$-module surjection, with $S$ separated. Then the following two conditions are equivalent.

(i) $\phi$ is a separated cover of $M$.

(ii) (a) $\ker(\phi)$ has no nonzero $\Gamma$-submodules (and hence $C\ker(\phi) = 0$, so that $\ker(\phi)$ is canonically a $\overline{\Lambda}$-module); and

(b) $\ker(\phi) \subseteq CS$.

Proof. Let $X = \Gamma \otimes S$, which contains $S$ by Lemma 4.4.

(ii) $\Rightarrow$ (i). Consider any factorization $\phi: S \to S' \to M$ with $S'$ separated, and suppose that $\ker(\theta)$ contains a nonzero element $s$. Then $s \in \ker(\phi) \subseteq CS$, so there is an expression $s = \sum_i c_i s_i$ with each $c_i \in C$ and $s_i \in S$.

On the other hand, $\ker(\phi)$ contains no nonzero $\Gamma$-submodules. Therefore, for some $\omega \in \Gamma$, the element $\omega s$ of $X$ is not an element of $\ker(\phi)$. But $\omega s = \sum_i (\omega c_i) s_i$, an element of $S$ since $C$ is an ideal of $\Gamma$ contained in $\Lambda$. Therefore $\phi(\omega s)$ is defined, and is nonzero since $\omega s \notin \ker(\phi)$. It follows that $\theta(\omega s) \neq 0$. However, $\theta$ can be viewed as a $\Gamma$-homomorphism $\Gamma S \to \Gamma S'$ by Lemma 4.5. This yields the contradiction $0 \neq \theta(\omega s) = \omega \theta(s) = 0$, proving (i).

(i) $\Rightarrow$ (ii)(a). Let $H$ be any $\Gamma$-submodule of $\ker(\phi)$. Then there is a factorization $\phi: S \to S/H \to M$. Since $S/H$ is a separated $\Lambda$-module, it is a $\Gamma$-module containing $S/H$. Hence, by the definition of separated cover, the map $S \to S/H$ must be one-to-one; that is, $H = 0$ as desired.

Since $C\ker(\phi)$ is a $\Gamma$-submodule of $\ker(\phi)$, the previous paragraph shows that $C\ker(\phi) = 0$, as claimed in (ii)(a).

Note that we have not yet used the standing hypothesis that the ring $\overline{\Lambda}$ is semisimple. The consequence of the semisimplicity of $\overline{\Lambda}$ that we need below is that every $\overline{\Lambda}$-module is semisimple.

(i) $\Rightarrow$ (ii)(b). Suppose, by way of contradiction, that $\ker(\phi) \not\subseteq CS$. Since $C\ker(\phi) = 0$, $\ker(\phi)$ is a $\overline{\Lambda}$-module and hence is a semisimple $\Lambda$-module. Therefore, it has a simple submodule $U$ that is not contained in $CS$. Let $\nu$ be the natural homomorphism of $S$ onto $S/CS$. Since the $\Lambda$-module $S/CS$ is annihilated by $C$, it is a $\overline{\Lambda}$-module and hence is semisimple. Therefore there is a projection map $\pi: S/CS \to \nu(U)$. Moreover, since $U$ is simple and not contained in $CS$, $\nu$ is one-to-one on $U$. Therefore $\nu^{-1}: \nu(U) \to U$ is a
well-defined Λ-module map. The composition of the maps
\[ S \xrightarrow{\nu} S/CS \xrightarrow{\pi} \nu(U) \xrightarrow{\nu^{-1}} U \]
is a map \( S \to U \) that equals the identity on \( U \). Therefore \( U \) is a direct summand of \( S \), say \( S = U \oplus T \). Then \( S/U \cong T \) is again a separated Λ-module. Since \( U \subseteq \ker(\phi) \) there is a factorization \( \phi: S \to S/U \to M \). Since \( \phi \) is a separated cover, we obtain the contradiction \( U = 0 \), proving that statement (ii)(b) holds. □

**Lemma 4.10.** Let \( \phi: S \to M \) be a separated cover. Then \( \phi \) is a minimal epimorphism (i.e., if \( T \) is a Λ-submodule of \( S \) such that \( \phi(T) = M \), then \( T = S \)).

**Proof.** Let \( T \) be as above. Then \( T + \ker(\phi) = S \). Since \( \ker(\phi) \) is a \( \overline{\Lambda} \)-module [Lemma 4.9], it is semisimple. Therefore we have \( \ker(\phi) = (T \cap \ker(\phi)) \oplus K \) for some submodule \( K \), and therefore \( T \oplus K = S \). This yields a factorization \( \phi: S \to S/K \cong T \to M \) with \( S/K \cong T \) separated, showing that \( K = 0 \) and therefore \( T = S \). □

**Proposition 4.11.** If \( \phi: S \to M \) is a separated cover and \( M \) is finitely generated, then so is \( S \).

**Proof.** Choose pre-images \( s_1, \ldots, s_n \in S \) of some finite set of generators of \( M \). Then \( \phi(\sum_i \Lambda s_i) = M \). Since \( \phi \) is a minimal epimorphism we have \( \sum_i \Lambda s_i = S \), as desired. □

**Theorem 4.12** (Almost functorial property). Let \( f: N \to M \) be a Λ-module homomorphism, and let \( \phi', \phi \) be separated covers. Then \( f \) can be lifted to a Λ-homomorphism \( \theta \) such that the following diagram commutes.

\[ \begin{array}{ccc}
S' & \xrightarrow{\theta} & S \\
\downarrow{\phi'} & & \downarrow{\phi} \\
N & \xrightarrow{f} & M.
\end{array} \]

If \( f \) is one-to-one or onto, then any such \( \theta \) has the same property.

**Proof.** Since \( S'/CS' \) is a module over the semisimple artinian ring \( \overline{\Lambda} \), it has a decomposition \( S'/CS' = \oplus_i \Lambda \overline{\pi}_i \) where each \( \Lambda \overline{\pi}_i \cong \overline{\Lambda} \overline{e}_i \) for some idempotent element \( \overline{e}_i \in \overline{\Lambda} \). We can choose each \( \overline{\pi}_i \) such that

\[ \overline{e}_i \to \overline{\pi}_i \quad \text{under the isomorphism} \quad \overline{\Lambda} \overline{e}_i \cong \overline{\Lambda} \overline{\pi}_i. \]

Choose a pre-image \( e_i \in \Lambda \) of each \( \overline{e}_i \) and \( x'_i \in S' \) of \( \overline{\pi}_i \). Let \( \{c_j s'_j\} \) (with \( c_j \in C \)) be a set of generators of the Λ-module \( CS' \). Then choose elements \( x_i, s_j \) related to the previously chosen elements as follows.

\[ f\phi'(x'_i) = \phi(x_i) \quad \text{and} \quad f\phi'(s'_j) = \phi(s_j). \]
Since $\tau_i$ is idempotent we have $e_i\tau_i = \tau_i$ and therefore the elements $e_i x'_i$ together with $c_j s'_j$ generate $S'$. If we can define a $\Lambda$-homomorphism $\theta$ that sends each $e_i x'_i \rightarrow e_i x_i$ and $c_j s'_j \rightarrow c_j s_j$ then we have lifted $f$. Let

$$y' = \sum_i \lambda_i e_i x'_i + \sum_j \delta_j c_j s'_j$$

where the sums are finite and each $\lambda, \delta \in \Lambda$. To see that $\theta$ is well-defined it suffices to show that $y' = 0 \Rightarrow y = 0$. Since $\ker(\phi)$ contains no nonzero $\Gamma$-submodules, by Lemma 4.9, it suffices to show that $\Gamma y \subseteq \ker(\phi)$. Thus we suppose that $y' = 0$ and choose $\omega \in \Gamma$. Our objective to show that $\phi(\omega y)$ is defined and equal to zero.

We must be careful about two things: Elements of $N$ and $M$ cannot be multiplied by $\omega$, if these $\Lambda$-modules are not contained in $\Gamma$-modules; and consequently the $\Lambda$-homomorphisms $\phi', \phi$ need not be extendable to $\Gamma$-homomorphisms.

Reading the statement $y' = 0$ modulo $C S'$ and using the expression in (4.12.4) for $y'$ yields $0 = \sum_i \lambda_i e_i x'_i$. Directness of the sum $\oplus_i \Lambda x'_i$ then yields $\lambda_i \tau_i x'_i = 0$ for all $i$. Hence, by the isomorphism in (4.12.2), we have $\lambda_i x'_i = 0$. Therefore each $\lambda_i e_i \in C$. Since $C$ is an ideal of $\Gamma$ contained in $\Lambda$, we have $\omega \lambda_i e_i x'_i \in C S' \subseteq S'$, and therefore $\phi'(\omega \lambda_i e_i x'_i)$ is defined. Similarly $\phi'(\omega \delta_j c_j s'_j)$ is defined. Therefore $y' = 0$ yields

$$0 = f \phi'(\omega y') = \sum_i f \phi'(\omega \lambda_i e_i x'_i) + \sum_j f \phi'(\delta_j c_j s'_j).$$

Since each $\omega \lambda_i e_i \in C \subseteq \Lambda$, relations (4.12.3) show that

$$f \phi'(\omega \lambda_i e_i x'_i) = \omega \lambda_i e_i f \phi'(x'_i) = \omega \lambda_i e_i \phi(x_i) = \phi(\omega \lambda_i e_i x_i).$$

Similarly we have $f \phi'(\omega \delta_j c_j s'_j) = \phi(\omega \delta_j c_j s_j)$. Making these replacements in (4.12.5) now shows that $\phi(\omega y) = 0$, completing the proof that $\theta$ is well-defined.

Suppose that $f$ is onto. Then commutativity of diagram (4.12.1) shows that $\phi \theta(S') = M$. Since $\phi$ is a minimal epimorphism [Lemma 4.10] we have $\theta(S') = S$, as claimed in the theorem.

On the other hand, suppose that $f$ is one-to-one. Then $f \phi'$ is a separated cover of $f(N)$ by $S'$. Therefore the factorization $f \phi' = \phi \theta: S' \twoheadrightarrow \theta(S') \rightarrow f(N)$ shows (by the definition of “separated cover”) that $\theta$ is one-to-one. □

**Corollary 4.13** (Uniqueness of separated cover). Let $\phi': S' \rightarrow M$ and $\phi: S \rightarrow M$ be separated covers of a $\Lambda$-module $M$. Then there is an isomorphism $\theta: S' \cong S$ such that $\phi' = \phi \theta$.

**Proof.** Take $N = M$ and $f = 1$ in (4.12.1). Then apply the almost functorial property. □
Corollary 4.14. Let $\phi: S \rightarrow M$ be a separated cover and $f: S' \rightarrow M$ any surjective $\Lambda$-homomorphism such that $S'$ is a separated $\Lambda$-module. Then $f$ factors through $\phi$; that is, there is a surjective $\Lambda$-homomorphism $\theta$ such that $f$ factors as follows:

\begin{equation}
  f : S' \xrightarrow{\theta} S \xrightarrow{\phi} M.
\end{equation}

Proof. The Corollary follows from Theorem 4.12 if we take $N = S'$ and note that the identity map $\phi'$ on $S'$ is a separated cover. \qed

Corollary 4.15. Let $\phi: S \rightarrow M$ be a separated cover such that $\ker(\phi) \neq 0$. Then $M$ is not a separated module.

Proof. If $M$ were a separated module, then the identity map on $M$ would be a projective cover of $M$. Therefore, by uniqueness of separated covers (Corollary 4.13), we would have $\ker(\phi) = 0$. \qed

Examples 4.16 (nonseparated modules). (i) For the two simplest examples, let $\Lambda$ be an unsplit Dedekind-like ring with normalization $\Gamma$. Then $m/m^2$ is $\Gamma$-isomorphic to $\Gamma/m = F$, and therefore this simple $\Gamma$-module has length 2 as a $\Lambda$-module. Therefore there is a $\Lambda$-module $K$ strictly contained in $m$ and strictly containing $m^2$.

The natural maps $\Gamma/m^2 \rightarrow \Gamma/K$ and $\Lambda/m^2 \rightarrow \Lambda/K$ are separated covers [Lemma 4.9 with $C = m$] and have nonzero kernels. Therefore the $\Lambda$-modules $\Gamma/K$ and $\Lambda/K$ are not separated modules [Corollary 4.15].

(ii) Much more generally, let $M = M(D)$ be the $\Lambda$-module associated with any standard diagram. (See (2.4.1), (2.6.2), and (3.3.1).) We claim that, if any bottom-gluing or bottom-reduction actually occurs, then $M$ is not a separated module.

For the proof, let $S = S(D)$, in the notation of Subsections 2.2 and 3.2. Then the natural map $\phi: S(D) \rightarrow M(D)$ is a separated cover. (See Subsections 9.6 and 10.4.) The statement that bottom-gluing or bottom-reduction actually occurs is equivalent to the statement that $\ker(\phi) \neq 0$. Therefore, by Corollary 4.15, $M$ is not a separated module.

Remark 4.17. Before proceeding to the next section, we comment on the reason for the word “separated.” Let $\phi: S \rightarrow M$ be a separated cover. If $\Lambda$ is strictly split Dedekind-like, $\ker(\phi)$ is always an amalgamation relation. (See diagrams (3.3.1), in which the passage from $S$ to $M$ is always given by what we call “bottom-gluing” relations.) Thus a $\Lambda$-module is “separated” if no such amalgamation relation has been imposed. In the unsplit case, $\ker(\phi)$ is often, but not always an amalgamation. (See diagrams (2.4.1) and (2.6.2), in which the passage from $S$ to $M$ involves both bottom gluing and bottom reduction.) Thus our terminology is slightly misleading in the context of unsplit Dedekind-like rings.
5. Isomorphism as matrix problem, unsplit case.

The main results of this section use separated covers to transform the problem of describing isomorphism classes of Λ-modules into a matrix problem over the fields $k$ and $F$, in the unsplit case.

**Notation 5.1.** Throughout this section Λ is a pullback ring, as specified in diagram (1.1.1) and display (1.1.2). That is, the first few results apply to both the unsplit and strictly split case, but, beginning with Notation 5.6, we assume that $(Λ, m, k)$ is unsplit Dedekind-like with normalization $(Γ, m, F)$.

Recall: (i) For a Λ-submodule $S$ of a Γ-module $X$ we write $ΓS = Γ \otimes_Λ S$ to mean that the natural surjection of the right-hand side onto the left-hand side is a bijection. (ii) A separated Λ-module is any Λ submodule of some Γ-module. (iii) We write functions as right operators when they represent (or will represent) right multiplication by matrices.

**Lemma 5.2.** Let $S \neq 0$ be a separated, finitely generated Λ-module and $ΓS$ a Γ-module generated by $S$. Then $ΓS = Γ \otimes_Λ S$ if and only if there is a positive integer $n$ such that $S/mS \cong k(n)$ (free $k$-module of rank $n$) and $ΓS/mS \cong Γ(n)$ (free Γ-module of rank $n$).

**Proof.** Let $τ: Γ \otimes_Λ S \rightarrow ΓS$ be the natural surjection. Note that, since $m$ is an ideal of both rings Λ and Γ, we have $m(ΓS) = mS$.

Suppose first that $S/mS \cong k(n)$ and $ΓS/mS \cong Γ(n)$. The surjection $τ$ induces a surjection $τ: (Γ \otimes S)/m(Γ \otimes S) \rightarrow ΓS/mS$. We claim that $τ$ is a bijection. Tensoring the short exact sequence $m \hookrightarrow Γ \rightarrow Γ$ by $S$ and using right-exactness of the tensor product yields the following chain of isomorphisms.

$$\frac{Γ \otimes S}{m(Γ \otimes S)} \cong \frac{Γ}{m} \otimes_Λ S \cong \frac{Γ}{m} \otimes_Λ \frac{S}{mS} \cong Γ \otimes_k k(n) \cong Γ(n).$$

(5.2.1)

Since $τ$ is a Γ-module surjection from the left-hand side of (5.2.1) onto $ΓS/mS \cong Γ(n)$, and $Γ$ is a Γ-module of finite length, we see that $τ$ is a bijection, as claimed.

Now we lift this assertion to $τ$ itself. Take $x \in \ker(τ)$. Let $π$ be the image of $x$ in the left-hand side of (5.2.1). Since $π$ is an isomorphism, we have $π = 0$, that is, $x ∈ m(Γ \otimes S)$. Therefore there is an expression $x = \sum_i m_i \otimes s_i$ ($m_i ∈ m, s_i ∈ S$). Since $m ⊆ Λ$ we have $x = 1 \otimes (\sum_i m_i s_i)$. But $x ∈ \ker(τ)$ implies that $\sum_i m_i s_i = 0$. Therefore $x = 0$; and so $τ$ is an injection, hence a bijection.

Conversely, suppose that $ΓS = Γ \otimes S$. Then by right exactness of the tensor product, as in the first part of this proof, we have $ΓS/mS \cong Γ \otimes (S/mS)$. Since $S/mS$ is a module over the field $Λ/m = k$ we have $S/mS \cong k(n)$ for some $n$, and hence $Γ \otimes_Λ (S/mS) \cong Γ \otimes_k k(n) \cong Γ(n)$ as desired. □
Example 5.3. Caution. We have $\Gamma \cdot \Gamma = \Gamma$. However, when $\Lambda$ is unsplit or strictly split Dedekind-like, we do not have $\Gamma \cdot \Gamma = \Gamma \otimes_{\Lambda} \Gamma$. If this equality held, the previous lemma, with $S = \Gamma$, would require $\Gamma/m\Gamma$ to be $\Lambda$-isomorphic to $k^{(n)}$ for some $n$ and $\Gamma$-isomorphic to $\Gamma^{(n)}$ for this same $n$. This would imply that $k = \Gamma$, which never holds when $\Lambda \neq \Gamma$.

This fact will cause some inconvenience in the remainder of this paper.

Corollary 5.4. Let $X$ be a finitely generated $\Gamma$-module. Then $X$ has a $\Lambda$-submodule $S$ such that $X = \Gamma \otimes_{\Lambda} S$ if and only if $X/mX$ is a free $\Gamma$-module.

Remark. Note that if $\Lambda$ is unsplit Dedekind-like, the phrase “if and only if $X/mX$ is a free $\Gamma$-module” can be deleted from the statement of Corollary 5.4, because all modules over the field $\Gamma = F$ are free.

Proof. The “only if” assertion is an immediate consequence of Lemma 5.2.

Conversely, suppose that $X/mX \cong \Gamma^{(n)}$ for some integer $n$. Choose $n$ elements $x_1, \ldots, x_n \in X$ whose images in $\Gamma^{(n)}$ are a free $\Gamma$-basis of $\Gamma^{(n)}$, and let $S = \sum \lambda_i x_i$. Since $m \subseteq \ker(\gamma)$, Nakayama’s lemma shows that $\gamma S = X$.

Since $S/mS$ is a $k = \Lambda/m$-vector space generated by $n$ elements, we have $S/mS \cong k^{(m)}$ for some $m \leq n$. It now suffices, by Lemma 5.2, to show that $m = n$. So suppose that $m < n$. Then the $\Gamma$-module $\Gamma(S/mS) = \Gamma S/mS \cong \Gamma^{(n)}$ would be generated by $m < n$ elements, which is impossible for a free module of rank $n$ over an artinian ring.

Lemma 5.5. Let $Y$ be a $k$-subspace of some finitely generated $\Gamma$-module, and suppose $\Lambda \neq \Gamma$.

(i) If $\overline{\Gamma} Y = \overline{\Gamma} \otimes_k Y$, then $Y$ contains no nonzero $\overline{\Gamma}$-submodule.

(ii) If $Y$ contains no nonzero $\Gamma$-module and $\dim_k(\overline{\Gamma}) = 2$ (which holds whenever $\Lambda$ is Dedekind-like), then $\overline{\Gamma} Y = \overline{\Gamma} \otimes_k Y$.

Proof. (i) Let $d = \dim_k(\overline{\Gamma})$. Since $\Lambda \neq \Gamma$ we have $d > 1$. Let $G$ be the largest $\overline{\Gamma}$-submodule of $Y$. Then $Y = G \oplus V$ for some $k$-subspace $V$ of $Y$. Let $y, g, v$ be the $k$-dimensions of $Y, G, V$ respectively, so that $y = g + v$.

We have $\dim_k(\overline{\Gamma} Y) \leq \dim_k(G) + \dim_k(\overline{\Gamma} V) \leq g + dv$. Also, $\dim_k(\overline{\Gamma} \otimes Y) = dv$. Since $\dim_k(\overline{\Gamma} Y) = \dim_k(\overline{\Gamma} \otimes Y)$ by (i), we have $dv \leq g$. But then $d > 1$ implies that $g = 0$ as desired.

(ii) Let $\varepsilon \in \overline{\Gamma} - k$. Since $\dim_k(\overline{\Gamma}) = 2$ we have $\overline{\Gamma} = k + k\varepsilon$. Hence 1 and $\varepsilon$ are $k$-linearly independent and $\varepsilon^2 \in k + k\varepsilon$. We want to show that the natural surjection $\tau: \overline{\Gamma} \otimes_k Y \twoheadrightarrow \overline{\Gamma} Y$ is a monomorphism. Let $x = \sum \gamma_i \otimes y_i \in \ker(\tau)$. We may assume that the $y_i$ are linearly independent over $k$. Write $\gamma_i = \alpha_i + \beta_i \varepsilon$ ($\alpha_i, \beta_i \in k$).

Then $0 = \tau(x) = (\sum \alpha_i y_i) + \varepsilon(\sum \beta_i y_i)$ which we can write in the form $0 = u + \varepsilon v$ with $u, v \in Y$. But then $ku + kv = k\varepsilon v + kv$ is a $\overline{\Gamma}$-submodule of $Y$. By assumption this $\overline{\Gamma}$-submodule must be 0, and hence $u = v = 0$. 


Linear independence of the $y_i$ over $k$ therefore implies that every $\alpha_i$ and $\beta_i$ equals zero, and therefore $x = 0$, as desired.

\begin{notation}[Matrix setup, unsplit case] For the remainder of this section, let $(\Lambda, m, k)$ be an unsplit Dedekind-like ring with normalization $(\Gamma, m, F)$, as in Notation 1.1. Thus $\Gamma = F$, a 2-dimensional field extension of $k$.

The map $\rho: \Gamma \rightarrow F = \Gamma$ in pullback diagram (1.1.1) induces a $\Gamma$-linear map $\Gamma/m^t \rightarrow F$ for every $t$, which we again call $\rho$. In fact, we denote any direct sum of such maps — from the direct sum of any $n$ such modules to $F^{(n)}$ — by $\rho$. Moreover, we write $m^\infty = 0$, and therefore $\Gamma = \Gamma/m^\infty$.

When $t \neq \infty$, we have $F \cong m^{t-1}/m^t$ via the $\Gamma$-isomorphism $\gamma + m \rightarrow \gamma \pi^{t-1} + m^t$, where $\pi$ is the standard $\Gamma$-generator of $m$ mentioned in Notation 1.1. Thus — given this choice of $\pi$ — every element of $F$ has a standard image in $\Gamma/m^t$, which we often regard as an identification, and the set of all such standard images of elements of $F$ defines the standard copy of $F$ in $\Gamma/m^t$. We note:

\begin{equation}
(5.6.1) \text{When } t \neq \infty, \neq 1, \text{ the standard copy of } F \text{ in } \Gamma/m^t \text{ satisfies } \rho(F) = 0.
\end{equation}

Let $X$ be the nonzero (external) direct sum of $\Gamma$-modules displayed below. 

\begin{equation}
(5.6.2) \quad X = \oplus_{\nu=1}^n \Gamma/m^{t\nu}
\end{equation}

where every $t\nu$ is a positive integer or $\infty$. Since $\Gamma$ is a DVR, every finitely generated indecomposable $\Gamma$-module is isomorphic to such a module $X$.

Let $e_\infty$ be the number of summands $\Gamma/m^{t\nu}$ of $X$ in (5.6.2) of infinite length $t\nu$. The map $\rho$ yields the $\Gamma$-linear surjection to $F^{(n)}$ shown in (5.6.3).

\begin{equation}
(5.6.3) \quad \rho: X \rightarrow F^{(n)} \quad F^{(n-e_\infty)} \subseteq X.
\end{equation}

We call coordinate $\nu$ of $F^{(n)}$ the coordinate corresponding to the the $\nu^{th}$ summand $\Gamma/m^{t\nu}$ of $X$. Each of the $n - e_\infty$ summands $\Gamma/m^{t\nu}$ of finite length contains a standard copy of $F$. This yields the inclusion shown in (5.6.3). Thus every coordinate of $F^{(n-e_\infty)}$ has a corresponding summand $\Gamma/m^{t\nu}$ of $X$, namely the summand containing this copy of $F$.

A matrix setup $\mathcal{X}$ (for $\Lambda$ unsplit Dedekind-like) is any direct sum decomposition of the form $X$ in (5.6.2), together with the corresponding surjection and inclusion displayed in (5.6.3). The $\Gamma$-module associated with $\mathcal{X}$ is the module $X$ in (5.6.2).

\begin{notation}[Matrix pair $(A, B)$] Let $\mathcal{X}$ be a matrix setup (unsplit case). Let $A$ be an $n \times n$ matrix over $F = \Gamma$, and let $B$ be a $q \times (n - e_\infty)$ matrix over $F$, for some $q \geq 0$. We attach a length label to each column of $A$ and $B$ as follows. For $1 \leq \nu \leq n$ define the length label of column $\nu$ of $A$ to be the length $t\nu$ of the corresponding summand $\Gamma/m^{t\nu}$ of $X$. For $1 \leq j \leq n - e_\infty$ define the length label of column $j$ of $B$ to be the length of the summand of $X$ corresponding to coordinate $j$ of $F^{(n-e_\infty)}$. Thus every column of $B$
has a corresponding column in $A$, namely the column of $A$ with the same corresponding summand of $X$; and every column of $A$ whose length label is finite has a corresponding column in $B$.

We require $A$ and $B$ to have the following properties.

(5.7.1) $A$ is invertible; the rows of $B$ are linearly independent over $F$; and every column of $B$ whose length label is finite has a corresponding column in $B$.

Thus right multiplication by $A$ and $B$ define $k$-linear monomorphisms

$$A: k^{(n)} \rightarrow F^{(n)} \quad B: k^{(q)} \rightarrow F^{(n-e_\infty)}.$$  

We call $(A, B)$ a matrix pair associated with $X$.

Although we write the matrices of the ordered pair $(A, B)$ side by side, we think of $B$ as being written underneath $A$, with each column of $B$ written under its corresponding column in $A$, and the length label of each column of $A$ above that column. Thus columns of $A$ with length label $\infty$ have no corresponding column in $B$ written beneath them.

**Definition 5.8 (Associated $\Lambda$-module).** Suppose that $(A, B)$ is a matrix pair associated with the matrix setup $X$. Using the maps in (5.7.2), let $S(A)$ be the $\Lambda$-module

$$(5.8.1) \quad S(A) = \{ x \in X \mid \rho(x) \in \text{im}(A) = k^{(n)}A \}$$

and define the $\Lambda$-module $M(A, B)$ by

$$(5.8.2) \quad M(A, B) = S(A)/\text{im}(B) \quad \text{where} \quad \text{im}(B) = k^{(q)}B.$$  

(If $q = 0$ we interpret this to mean that $M(A, B) = S(A)$.) This definition makes sense — that is, $\text{im}(B) \subseteq S(A)$ — because $\rho(\text{im}(B)) = 0$ by (5.6.1) and the requirement about columns of zeros in (5.7.1). We call $M(A, B)$ the $\Lambda$-module associated with the matrix pair $(A, B)$ (with respect to the matrix setup $X$).

We note that $\Gamma \cdot S(A) = X$, since $A$ is invertible. Thus $S(A)$ is the pullback of the commutative square in diagram (5.8.3) below.

$$\begin{array}{ccc}
S(A) & \xrightarrow{\rho} & \Gamma \cdot S(A) = X \\
\downarrow & & \downarrow \\
k^{(n)}A & \subset & F^{(n)}
\end{array}$$

Note that $\ker(\rho) = mX = mS(A)$. (The second equality holds because $m\Gamma = m$.)

**Definition 5.9 (Display operation).** A display operation on a matrix pair $(A, B)$ (associated with a matrix setup $X$) is a permutation of the columns of $A$, together with the same permutation of their length labels and their corresponding columns of $B$ (so that corresponding columns of $A$ and $B$...
remain corresponding columns after being moved). Thus each display operation corresponds to the effect on \((A, B)\) of rearranging the summands \(\Gamma / m^t\) in decomposition (5.6.2) of \(X\). For future reference we record the following consequence of this observation.

(5.9.1) Let \((A', B')\) — including its length labels — be obtained by performing a display operation on \((A, B)\). Then \(M(A', B') \cong M(A, B)\), where the module on the left is computed with respect to the correspondingly altered matrix setup.

We will need display operations to display our matrix pairs in canonical form.

**Theorem 5.10.** Let \(X\) be a matrix setup (unsplit case) with associated \(\Gamma\)-module \(X\), and let \((A, B)\) be a matrix pair associated with \(X\). Also, let \(S = S(A)\). Then:

(i) \(\Gamma S = \Gamma \otimes_A S\).

(ii) The natural surjection \(S \twoheadrightarrow M(A, B)\) is a separated cover of \(M(A, B)\).

Moreover,

(iii) Every finitely generated \(A\)-module is isomorphic to \(M(A, B)\) for some \((A, B)\) with respect to some matrix setup \(X\).

(iv) Let \(X\) and \(X'\) be the \(\Gamma\)-modules associated with matrix setups \(X\) and \(X'\) respectively, and let \(M(A, B)\) and \(M(A', B')\) be \(\Lambda\)-modules computed with respect to \(X\) and \(X'\) respectively. If \(M(A, B) \cong M(A', B')\) as \(\Lambda\)-modules then \(X \cong X'\) as \(\Gamma\)-modules.

**Proof.** (i) This follows from Lemma 5.2 and (5.8.3).

(ii) By Lemma 4.9 (with \(C = m\)) it suffices to show: (a) \(\text{im}(B)\) contains no nonzero \(\Gamma\)-submodules, and (b) \(\text{im}(B) \subseteq mS\). Consider \(\Gamma \cdot \text{im}(B)\), the \(\Gamma\)-module generated by the rows of \(B\). We claim that \(\Gamma \cdot \text{im}(B) = \Gamma \otimes_k \text{im}(B)\).

Since the right-hand side maps onto the left-hand side, it suffices to show that both sides have the same \(F\)-dimension. Since the rows of \(B\) are \(F\)-linearly independent, by (5.7.1), the \(F\)-dimension of the left-hand side is the number of rows of \(B\). On the other hand, \(F\)-independence of the rows of \(B\) implies that they form a \(k\) basis of \(\text{im}(B) = k^{(q)}B\), and hence an \(F\)-basis of \(F \otimes_k \text{im}(B)\); and hence the \(F\)-dimension of \(F \otimes_k \text{im}(B)\) again equals the number of rows of \(B\), proving the claim.

The claim, together with Lemma 5.5(i), shows that \(\text{im}(B)\) contains no nonzero \(\Gamma\)-submodules, establishing (a).

To establish (b) it suffices to show that \(\text{im}(B)\) is contained in the sum of all those summands \(\Gamma / m^t\) of \(X\) whose length is finite but greater than 1. The “finite” part is part of the definition of a matrix setup; and the “greater than 1” part follows from the fact that every column of \(B\) with length label 1 consists of zeros, by (5.7.1).
(iii) Let $M$ be any finitely generated $A$-module. Then $M$ has a separated cover $\phi: S \to M$ [Proposition 4.7]. Since $S$ is a separated $\Gamma$-module, it can be regarded as a $\Lambda$-submodule of the $\Gamma$-module $X = \Gamma S = \Gamma \otimes_{\Lambda} S$ [Lemma 4.4]. Since $\Gamma$ is a DVR and $\Gamma X$ is finitely generated, we can take $X$ to be as shown in (5.6.2). This gives us modules $S$ and $\Gamma S = X$ to use in the top row of the commutative square in (5.8.3) that we are building. We will soon attach a matrix $A$ to $S$.

Decomposition (5.6.2) of $X$, together with the map $\rho$ in (5.6.3) yield an $F$-linear identification $\Gamma S / mS = F^n$. Since $\Gamma S = \Gamma \otimes_{\Lambda} S$ we have a $\Lambda$-isomorphism (equivalently, $k$-isomorphism) $S / mS \cong k^n$ [Lemma 5.2]. In addition, the relation $\Gamma S = \Gamma \otimes_{\Lambda} S$, reduced modulo $mS$ yields $F(S/mS) = F \otimes_k (S/mS)$. Therefore the $n$-dimensional $k$-submodule $S/mS$ of $F^n = \Gamma S / mS$ is $k$-generated by an $F$-basis of $F^n$. This yields an invertible $n \times n$ matrix $A$ over $F$ such that $S/mS = k^n A$.

To complete the proof that $S = S(A)$ it suffices to prove that $S$ is the pullback of this diagram, and for this it suffices to show that $S \supseteq \ker(\rho) = mX$. But $mX = m\Gamma S = mS$, which is contained in $S$ since $m \subseteq \Lambda$.

It now suffices to prove that $\ker(\phi) = \im(B)$ for a suitable $B$. Since $\phi$ is a separated cover, Lemma 4.9 shows that $\ker(\phi)$ is a $k$-submodule of $X$; that is, $m \cdot \ker(\phi) = 0$. Thus, in the current notation,

$$(5.10.1) \quad \ker(\phi) \subseteq \{ t_{\nu}^{-1} m^t_{\nu} \mid t_{\nu} \neq \infty \} = F^{(n-e_\infty)}.$$ 

We claim that, for some nonnegative integer $q$, $\ker(\phi)$ has a $k$-basis consisting of $q$ elements of $F^{(n-e_\infty)}$ that are linearly independent over $F$. For this, it suffices (by a dimension argument) to show that $F \cdot \ker(\phi) = F \otimes_k \ker(\phi)$. But $\phi$ is a separated cover, so Lemma 4.9 shows that $\ker(\phi)$ has no nonzero $F$-submodules. Since $\Lambda$ is unsplit Dedekind-like, we have $\dim_k(F) = 2$, so the desired equality follows by Lemma 5.5(ii).

In view of the claim, there is a $q \times (n-e_\infty)$ matrix $B$ over $F$ whose rows are $F$-linearly independent and form a $k$-basis of $\ker(\phi)$. One last appeal to Lemma 4.9 shows that $\ker(\phi) \subseteq mS = m\Gamma S = mX$. This shows that (5.10.1) can be refined to

$$\ker(\phi) \subseteq \{ t_{\nu}^{-1} m^t_{\nu} \mid t_{\nu} \neq \infty, \neq 1 \} \subseteq F^{(n-e_\infty)}$$

from which it follows that all columns of $B$ with length label 1 must be zero. Thus we now have $\ker(\phi) = k^{(n)} B$ were $B$ is as in (5.7.1), as desired.

(iv) By uniqueness of the separated cover [Corollary 4.13], the $\Lambda$-isomorphism class of $S(A)$ is determined by that of $M(A, B)$. Statement (iv) now follows from the fact that, in any matrix setup, $X = \Gamma \otimes_{\Lambda} S(A)$, by statement (i) above. \qed
The final step in transforming the problem of classifying isomorphism classes of finitely generated $\Lambda$-modules into a matrix problem is to transform the condition $M(A, B) \cong M(A', B')$ into a purely matrix-theoretic statement about the two matrix pairs $(A, B)$ and $(A', B')$. In view of statements (iii) and (iv) of the previous theorem, we can restrict our attention to the situation that both matrix pairs arise from the same matrix setup.

The notation $A[\text{cols } j]$ denotes the submatrix of $A$ consisting of the $j$-labeled columns.

Theorem 5.11 (Matrix operations, unsplit case). Let $(A, B)$ and $(A', B')$ be matrix pairs associated with a matrix setup $\mathcal{X}$ (where $\Lambda$ is unsplit Dedekind-like). Then $M(A, B) \cong M(A', B')$ if and only if $(A', B')$ can be obtained from $(A, B)$ by a finite sequence of the following operations.

(i) (a) Left multiply $A$ by an invertible matrix over $k$.
   (b) Left multiply $B$ by an invertible matrix over $k$.

(ii) For any length label $j$ right-multiply $A[\text{cols } j]$ by an invertible matrix $Q$ over $F$ and, if $j \neq \infty$, simultaneously right-multiply $B[\text{cols } j]$ by $Q$.

(iii) (a) For any $i > j$ add an $F$-scalar multiple of a $i$-labeled column of $A$ to a $j$-labeled column of $A$. ("Sweep toward smaller lengths in $A".)
   (b) For any $i < j$ add any $F$-scalar multiple of a $i$-labeled column of $B$ to a $j$-labeled column of $B$. ("Sweep toward larger lengths in $B".)

We call these matrix operations "$k$-$F$ sweeping-similarity" operations.

Proof. The matrix operations in the theorem are stated in a way that is convenient for the use of the theorem and our eventual statement of the canonical form of $(A, B)$, but not for the proof of the present theorem. Therefore we begin the proof by changing notation.

Note that it is possible to perform a display operation, rearranging the summands $\Gamma/m^t$ of $X$ in descending order; that is, such that the decomposition (5.6.2) of $X$ takes the form

\begin{equation}
X = (\Gamma/m^\infty)^{(e_\infty)} \oplus (\Gamma/m^t)^{(e_t)} \oplus \ldots \oplus (\Gamma/m^i)^{(e_i)} \oplus \ldots \oplus (\Gamma/m)^{(e_1)}
\end{equation}

where $\infty > t > \cdots > i > \cdots > 1$

where any particular block, e.g. the $\infty$ block, might not actually be present. (Recall that the notation $(e_t)$ denotes the multiplicity of $\Gamma/m^t$ in $X$.)

After doing this display operation we can restate the stated sweeping-similarity operations as relations of the form

\begin{equation}
A' = P_1AQ_1 \quad B' = P_2BQ_2
\end{equation}

in which $P_1$ and $P_2$ are invertible matrices over the field $k$, and $Q_1$ and $Q_2$ are block triangular invertible matrices over the field $F$. The block triangular forms of the matrices $Q_1$ and $Q_2$ are due to the fact that all sweeping in $A$ is toward smaller lengths while all sweeping in $B$ is toward greater lengths.
We describe this block triangular form precisely in the next subsection, and then state and prove the theorem in the altered notation.

**Notation 5.12** ($\mathcal{X}$ triangular form, unsplit case). Let $\mathcal{X}$ be a matrix setup whose summands are arranged in decreasing order, as in (5.11.1). As usual, let $n$ denote the number of individual summands $\Gamma/m^j$ of $X$ each counted as often as it occurs. Recall that each column of the $n$-column matrix $A$ has a length label. Since this label sequence consists of $n$ terms, we can use it to label both the rows and columns of any $n \times n$ matrix, and we can use the noninfinite labels to label both the rows and columns of any $(n - e_{\infty}) \times (n - e_{\infty})$ matrix. This partitions such matrices into blocks. If, for example, $Q$ is a matrix being partitioned in this way, $Q[\text{rows } i, \text{cols } j]$ denotes the $e_i \times e_j$ submatrix consisting of the intersection of the $i$-labeled rows with the $j$-labeled columns of $Q$.

We say that an $n \times n$ matrix $Q_1$ is $\mathcal{X}$ upper triangular if $Q_1[\text{rows } i, \text{cols } j] = 0$ when $i < j$. (In interpreting this, remember that the length labels occur in decreasing order.) For example, see the matrix $Q_1$ in (5.12.1), noting that we have suppressed multiplicities, writing each distinct length label only once. Similarly, we say that an $(n - e_{\infty}) \times (n - e_{\infty})$ matrix $Q_2$ is in $\mathcal{X}$ lower triangular if $Q_2[\text{rows } i, \text{cols } j] = 0$ when $i > j$. (Again see the example in (5.12.1).) Finally, we call any block of the form $Q[\text{rows } i, \text{cols } i]$ a main-diagonal block.

\[
(5.12.1) \quad Q_1 = \begin{bmatrix} 6 & 4 & \cdots & \cdots & \cdots \\ 0 & Q_1[6,6] & 6 & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & \cdots & 0 \\
\end{bmatrix} \quad \text{and} \quad Q_2 = \begin{bmatrix} 6 & 4 & \cdots & \cdots & \cdots \\ 0 & Q_2[6,6] & 6 & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & \cdots & 0 \\
\end{bmatrix}.
\]

We reformulate and prove Theorem 5.11 in the new notation.

**Theorem 5.13.** Let $(A, B)$ and $(A', B')$ be matrix pairs associated with a matrix setup $\mathcal{X}$ (A unsplit Dedekind-like) whose summands occur in descending order, as in (5.11.1). Then $M(A, B) \cong M(A', B')$ if and only if there exist matrix relations $A' = P_1AQ_1$ and $B' = P_2BQ_2$ in which the following conditions hold.

(i) $P_1$ and $P_2$ are invertible, with entries in $k$.
(ii) $Q_1$ is invertible and $\mathcal{X}$ upper triangular, with entries in $F$.
(iii) $Q_2$ is invertible and $\mathcal{X}$ lower triangular, with entries in $F$.
(iv) $Q_2[\text{rows } i, \text{cols } i] = Q_1[\text{rows } i, \text{cols } i]$ whenever $i \neq e_{\infty}$; that is, every diagonal block of $Q_2$ equals the corresponding diagonal block of $Q_1$.

**Remark.** The matrices in (5.12.1) illustrate the triangular forms of the matrices $Q_1$ and $Q_2$ in this theorem. In addition to what is displayed in (5.12.1),
the matrices in the theorem satisfy \( Q_2[\text{rows } i, \text{cols } i] = Q_1[\text{rows } i, \text{cols } i] \) for \( i \neq \infty \).

**Proof.** Let \( X \) be the \( \Gamma \)-module associated with \( X \). Also, let \( K(A, B) \) denote the kernel of the natural homomorphism \( S(A) \twoheadrightarrow M(A, B) \). The matrices \( P_i \) and \( Q_i \) in this theorem arise from an automorphism \( \tau \) of \( X \). Therefore, as in the rest of this series of papers, \( \tau \) acts on the right.

Claim 1: \( M(A, B) \cong M(A', B') \) if and only if there is a \( \Gamma \)-automorphism \( \tau \) of \( X \) such that \( (S(A))\tau = S(A') \) and \( (K(A, B))\tau = K(A', B') \) (as illustrated below).

\[
\begin{align*}
K(A, B) & \subset S(A) \twoheadrightarrow M(A, B) \\
(\cong) \downarrow \tau & (\cong) \downarrow \tau & (\cong) \downarrow \sigma \\
K(A', B') & \subset S(A') \twoheadrightarrow M(A', B').
\end{align*}
\]

First let \( \sigma : M(A, B) \cong M(A', B') \) be given. Since \( S(A) \twoheadrightarrow M(A, B) \) is a separated cover [Theorem 5.10(ii)], the almost functorial property of separated covers [Theorem 4.12] yields a \( \Lambda \)-isomorphism \( \tau \) such that both squares in \( (5.13.1) \) commute. Also, since \( X = \Gamma S(A) = \Gamma \otimes_\Lambda S(A) \) [Theorem 5.10(i)], we can extend \( \tau \) to a \( \Gamma \)-automorphism of \( X \), namely \( 1 \otimes \tau \). The converse assertion of the claim is obvious.

The effect of Claim 1 is that we have no further need for \( M(A, B) \) and \( M(A', B') \) in the rest of this proof. We replace the condition \( M(A, B) \cong M(A', B') \) by the condition that the automorphism \( \tau \) exists.

Now suppose that the \( \tau \) exists. To see how \( P_i \) and \( Q_i \) arise, we build the following commutative cube. Let the back (inner) square be the pullback diagram that defines \( A \), and the front (outer) square be the pullback diagram that defines \( A' \). Then insert \( \tau \) in the two places shown.

\[
\begin{align*}
S(A') & \subset X \\
S(A) & \twoheadrightarrow X \\
\rho & \\
k^{(n)}A' & \subset F^{(n)}
\end{align*}
\]

The top square commutes, by the hypothesis on \( \tau \) in this half of the proof. Since \( \rho \) is a surjection and \( \ker(\rho) = mX \), there is a unique (necessarily invertible) matrix \( Q_1 \) over \( F \) such that right multiplication by \( Q_1 \) makes the right-hand square commute. Since \( \tau \) takes \( S(A) \) to \( S(A') \), and the images of \( S(A) \) and \( S(A') \) under \( \rho \) are \( k^{(n)}A \) and \( k^{(n)}A' \) respectively, the left-hand square is a restriction of the right-hand square, and therefore commutes.
Therefore \(k^{(n)}AQ_1 = k^{(n)}A'\). Since \(A'\) is invertible, this can be rewritten \(k^{(n)}AQ_1(A')^{-1} = k^{(n)}\). Thus right multiplication by \(AQ_1(A')^{-1}\) is a \(k\)-linear automorphism of \(k^{(n)}\), and hence equals right multiplication by a unique invertible matrix over \(k\) that we call \(P_1^{-1}\). But then \(AQ_1(A')^{-1} = P_1^{-1}\), which is equivalent to \(A' = P_1AQ_1\), as desired.

To see how \(P_2\) and \(Q_2\) arise, we build the next diagram, starting with its top and bottom rows, as shown.

\[
\begin{array}{ccc}
k(q) & \xrightarrow{B} & F^{(n-e_{\infty})} \\ \downarrow P_2^{-1} & & \downarrow Q_2 \\ k(q) & \xrightarrow{B'} & F^{(n-e_{\infty})} \end{array} \subseteq X.
\]

(5.13.3)

Since \(F^{(n-e_{\infty})}\) is the \(\Gamma\)-socle of \(X\), the \(\Gamma\)-automorphism \(\tau\) takes \(F^{(n-e_{\infty})}\) isomorphically onto itself. Therefore the restriction of \(\tau\) to this submodule induces an isomorphism (given by an invertible matrix \(Q_2\)) from \(F^{(n-e_{\infty})}\) to \(F^{(n-e_{\infty})}\) making the right-hand square of (5.13.3) commute.

We have \((K(A, B))\tau = K(A', B')\) by (5.13.1). Since \(K(A, B) = k(q)B\) and \(K(A', B') = k(q)B'\), this and the second square in (5.13.3) yield \(k(q)BQ_2 = k(q)B'\). Since right multiplication by \(B'\) is one-to-one, this defines a one-to-one \(k\)-linear map \(k(q) \rightarrow k(q)\) which is therefore also surjective and equals right multiplication by a unique invertible matrix over \(k\) that we call \(P_2^{-1}\), making the left-hand square in (5.13.3) commute. Commutativity of the first square in (5.13.3) is precisely the desired relation \(B' = P_2BQ_2\).

To complete the proof of this half of the theorem we need to show that the existence of \(\tau\) implies the stated block triangular forms of \(Q_1\) and \(Q_2\). Recall the structure of \(X\) (below), where \(n = \sum_i e_i\).

\[
\begin{array}{c}
X = \Gamma^{(e_{\infty})} \oplus (\Gamma/m^t)^{(e_1)} \oplus \ldots \oplus (\Gamma/m^t)^{(e_i)} \oplus \ldots \oplus (\Gamma/m)^{(e_1)} \\
\downarrow \tau = (\tau_{\mu\nu}) \\
X = \Gamma^{(e_{\infty})} \oplus (\Gamma/m^t)^{(e_1)} \oplus \ldots \oplus (\Gamma/m^t)^{(e_i)} \oplus \ldots \oplus (\Gamma/m)^{(e_1)}. 
\end{array}
\]

(5.13.4)

We can view \(\tau\) as right multiplication by a matrix \((\tau_{\mu\nu})\) each of whose entries is a \(\Gamma\)-homomorphism

\[
(5.13.5) \quad \tau_{\mu\nu}: \Gamma/m^i \rightarrow \Gamma/m^j \quad (\mu \in \text{block } i, \nu \in \text{block } j)
\]

where “\(\mu \in \text{block } i\)” signifies that summand \(\mu\) of \(X\) is one of the \(e_i\) summands equal to \(\Gamma/m^i\).

The rest of the proof uses the fact that, because \(\Gamma\) is a DVR, each \(\Gamma/m^i\) is a uniserial \(\Gamma\)-module, and has finite length when \(t \neq \infty\). Since \(\rho\) is a ring homomorphism of \(\Gamma\) onto \(F\) and has kernel \(m\), we identify \(\rho\) with the natural homomorphism \(\Gamma \rightarrow \Gamma/m\). We note that, since \(\tau_{\mu\nu}\) is \(\Gamma\)-linear, \(\tau_{\mu\nu}\) equals multiplication by an element of \(\Gamma\). That is, there exists \(\gamma \in \Gamma\) such that, for
all \( x \in \Gamma \):

\[
(5.13.6) \quad (x + m^i)\tau_{\mu\nu} = x \cdot (1 + m^i)\tau_{\mu\nu} = x(\gamma + m^j).
\]

Moreover, \( \tau_{\mu\nu} \) induces the endomorphism of the \( F \)-vector space \( F = \Gamma/m \) given by multiplication by \( \rho(\gamma) = \gamma + m \). Finally, when \( i \neq \infty \), the homomorphism \( \tau_{\mu\nu} \) must take the unique minimal \( \Gamma \)-submodule \( m^{i-1}/m^i \) (our standard copy of \( F \)) to a module of length at most 1. We break up the remaining details into three cases.

Case 1: \( i < j \) (below the main-diagonal blocks). Here \( \Gamma/m^i \) is uniserial of finite length \( i \) and hence its image under \( \tau_{\mu\nu} \) is uniserial of length at most \( i \). Therefore \( \text{im}(\tau_{\mu\nu}) \subseteq m/m^j \), and so the induced map \( \Gamma/m \to \Gamma/m \) equals zero. In other words, the \((\mu,\nu)\)-entry of \( Q_1 \) is zero whenever \((\mu,\nu)\) belongs to a block \( Q_1[\text{rows } i, \text{cols } j] \) that lies below the main-diagonal blocks, as claimed in statement (ii).

Case 2: \( \infty \neq i > j \) (above the main-diagonal blocks). Here \( \tau_{\mu\nu} \) cannot be one-to-one, and hence takes the unique simple \( \Gamma \)-submodule \( m^{i-1}/m^i \) of \( \Gamma/m^i \) to zero. Therefore, in this case, the \((\mu,\nu)\)-entry of \( Q_2 \) equals zero as claimed in (iii).

Case 3: \( i = j \neq \infty \) (main-diagonal block). We have already noted that the endomorphism of \( \Gamma/m \) induced by \( \tau_{\mu\nu} \) equals multiplication by \( \rho(\gamma) \). But since \( \tau_{\mu\nu} \) equals multiplication by \( \gamma \in \Gamma \) on \( \Lambda/m^i \), it follows that \( \tau_{\mu\nu} \) induces an endomorphism of the socle \( m^{i-1}/m^i \) of \( \Gamma/m \), again given by multiplication by the same element \( \rho(\gamma) \), as desired.

This completes the proof of the “only if” part of the theorem.

Conversely, suppose that \( A' = P_1AQ_1 \) and \( B' = P_2BQ_2 \) and (i)–(iv) hold. We want to prove that a \( \Gamma \)-automorphism \( \tau \) of \( X \) exists satisfying Claim 1, that is, such that \((S(A))\) \( \tau = S(A') \) and \((K(A,B))\) \( \tau = K(A',B') \). We define the map \( \tau \) by defining each \( \tau_{\mu\nu} \) in (5.13.5), considering two cases.

Case A. \( i \geq j \) (on or above main diagonal). Let \( x \) be the \((\mu,\nu)\)-entry of \( Q_1 \). Then \( x = \rho(\gamma) \) for some \( \gamma \in \Gamma \). Since \( i \geq j \), multiplication by \( \gamma \) followed by reduction modulo \( m^j \) is a well-defined \( \Gamma \)-homomorphism that we use for \( \tau_{\mu\nu} \) in (5.13.5).

Case B. \( i < j \) (below main diagonal). Using our standard \( \Gamma \)-generator \( \pi \) of \( m \), multiplication by \( \pi^{j-i} \) yields a \( \Gamma \)-embedding of \( \Gamma/m^i \) into \( \Gamma/m^j \). Let \( x \) be the \((\mu,\nu)\)-entry of \( Q_2 \), and let \( \gamma \) be any element of \( \Gamma \) such that \( \rho(\gamma) = x \). Then multiplication by \( \pi^{j-i} \cdot \gamma \) is a well-defined \( \Gamma \)-linear map that we use for \( \tau_{\mu\nu} \) in (5.13.5).

Claim 2: With the given definition of \( \tau \), the right-hand square in diagram (5.13.2) commutes. It suffices to check commutativity of each smaller square, in which \( \tau \) is replaced by some \( \tau_{\mu\nu} \) in (5.13.5) and \( Q_1 \) is replaced by its
\((\mu, \nu)\)-entry. Except if the block containing \((\mu, \nu)\) is strictly below the main-diagonal blocks, this is true by the definition in Case A. For the below-diagonal blocks, the map induced by \(\tau_{\mu\nu}\) is zero regardless of our choice of \(\tau_{\mu\nu}\), as shown in the proof of Case 1 above; so the claim, for below-diagonal blocks, follows from the block upper triangular form of \(Q_1\).

**Claim 3:** The endomorphism \(\tau\) is an automorphism. The map induced by \(\tau\) modulo \(m\) — i.e., right multiplication by \(Q_1\) — is a surjection. Therefore, by Nakayama’s lemma, \(\tau\) itself is a surjection. Since \(X\) is noetherian, it follows that \(\tau\) is also an injection.

**Claim 4:** The endomorphism \(\tau\) induces right multiplication by \(Q_2\); that is, the right-hand square of diagram (5.13.3) commutes. As in the proof of Claim 2, it suffices to check each \(\tau_{\mu\nu}\) individually. For \((\mu, \nu)\) strictly below the main-diagonal blocks this is true by the definition in Case B. For \((\mu, \nu)\) strictly above the main-diagonal blocks, the map induced by \(\tau_{\mu\nu}\) is zero regardless of our choice of \(\tau_{\mu\nu}\), as shown in the proof of Case 2 above. This leaves only the case that \((\mu, \nu)\) belongs to some main-diagonal block, say \([\text{rows } i, \text{cols } i]\).

Let \(\gamma\) and \(x = \rho(\gamma)\) be as in Case A of our definition. Then \(x\) is the \((\mu, \nu)\)-entry of \(Q_1\), and hence, by condition (iv) in the statement of the theorem, \(x\) is the \((\mu, \nu)\)-entry of \(Q_2\). The desired commutativity of the second square of diagram (5.13.3) is now immediate.

Now we are ready to prove that \((S(A))\tau = S(A')\). First note that, in the bottom left of diagram (5.13.2), right multiplication by \(Q_1\) takes \(k^{(n)}A\) onto \(k^{(n)}A'\) because \(A' = P_1AQ_1\) and \(P_1\) is an invertible matrix over \(k\). Therefore the bottom square commutes. The inner and outer squares commute, being pullback squares by definition (5.8.3), and the right-hand square commutes by Claim 2. Therefore the upper left map \(\tau\) takes the pullback \(S(A)\) of the inner square onto the pullback \(S(A')\) of the outer square, as claimed.

Finally, we show that \((K(A, B))\tau = K(A', B')\). The left-hand square in (5.13.3) commutes since \(B' = P_2BQ_2\), and the right-hand square commutes by Claim 4. Therefore \(\tau\) takes \(\text{im}(B) = K(A, B)\) onto \(\text{im}(B') = K(A', B')\).

\[\square\]

**Notation 5.14.** Let \((A, B)\) and \((A', B')\) be matrix pairs associated with matrix setups \(\mathcal{X}\) and \(\mathcal{X}'\) respectively (unsplit case) [Notation 5.7]. We write \((A, B) \cong (A', B')\) to indicate that each pair can be obtained from the other by means of display operations [Definition 5.9] and \(k\)-F sweeping-similarity operations [Theorem 5.11].

It is easy to see that, when such an isomorphism holds, we can do the display operations before the sweeping-similarity operations. After doing the display operations and changing notation, we have \(\mathcal{X} = \mathcal{X}'\). Moreover, when \(\mathcal{X} = \mathcal{X}'\) we have \((A, B) \cong (A', B')\) if and only if each pair can be obtained from the other by \(k\)-F sweeping-similarity alone.
The following result summarizes much of the content of this section.

**Theorem 5.15.** In the unsplit case, every finitely generated \( \Lambda \)-module is isomorphic to some \( M(A, B) \) associated with some matrix setup. Moreover, \( M(A, B) \cong M(A', B') \) if and only if \( (A, B) \cong (A', B') \).

**Proof.** The first assertion is Theorem 5.10(iii). Next, suppose that \( M(A, B) \cong M(A', B') \). Then \( X \cong X' \), where \( X \) and \( X' \) are the \( \Gamma \)-modules associated with the matrix setups given by the row of length labels in \( (A, B) \) and in \( (A', B') \) [Theorem 5.10(iv)]. Consider the decomposition (5.6.2) of \( X \), whose summands are various modules \( \Gamma/m^t \) (1 \( \leq t \leq \infty \)), and its analog for \( X' \), which we call (5.6.2)'. For each \( t \), let \( X(t) \) denote the subsum, in the decomposition of \( X \), consisting of all summands \( \Gamma/m^t \). Since the Krull-Schmidt Theorem holds for finitely generated modules over principal ideal domains, \( X \) and \( X' \) have the same summands \( \Gamma/m^t \), except for order of occurrence. In particular, \( X(t) = X'(t) \) for every \( t \). Therefore some permutation of the indecomposable summands of \( X \) transforms them to the indecomposable summands of \( X' \). This permutation of summands in (5.6.2) and (5.6.2)' defines a display operation on \( (A, B) \). After performing this display operation we have \( X = X' \), whence Theorem 5.11 states that \( k-F \) sweeping-similarity operations transform \( (A, B) \) to \( (A', B') \), as desired.

The converse implication holds because neither display operations [(5.9.1)] nor sweeping-similarity operations [Theorem 5.11] change the isomorphism class of \( M(A, B) \). \( \square \)

**Reminder.** This is a good point to remind readers of the existence of the terminological index in §13.

### 6. Isomorphism as matrix problem, strictly split case.

The main results of this section use separated covers to transform the problem of describing isomorphism classes of \( \Lambda \)-modules into a matrix problem, for the case that \( \Lambda \) is a strictly split Dedekind-like ring. Since this problem was solved in [L1, L3] (and in [NR], corrected in [NRSB]), we omit most details, which are very similar to the corresponding details in the unsplit case considered in the previous section. The emphasis here is on phrasing these results in the terminology of the present paper, in such a way that it will be useful later in this paper and in the forthcoming paper on the nonlocal situation.

The principal differences between results in the split and unsplit cases are due to the fact that, in the split case, \( \Gamma \) is the direct sum of two DVRs, while in the unsplit case, \( \Gamma \) is a single DVR. Therefore, in the split case, there are twice as many matrices to manipulate as in the unsplit case. But in the split case only one field is involved, namely \( k \).
**Notation 6.1.** Throughout this section \((\Lambda, m, k)\) is a strictly split Dedekind-like ring, as in Notation 1.1. Thus the normalization of \(\Lambda\) is \(\Gamma = \Gamma_1 \oplus \Gamma_2\) where each \((\Gamma_i, m_i, k)\) is a DVR and \(m = m_1 \oplus m_2\). Following the style of Notation 5.6, we use the notation \(\rho\) not only for the map \(\Gamma \rightarrow \overline{\Gamma} = \Gamma/m = k \oplus k\) in pullback diagram (1.1.1), but also for the maps \(\Gamma_i/m_i \rightarrow k\) induced by the original map \(\rho\), and for direct sums of such maps. Moreover, we let \(m_\infty = 0\), so that \(\Gamma_i = \Gamma_i/m_\infty\).

**Example 6.2.** Note that the two direct summands \(k\) of \(\overline{\Gamma}\) are isomorphic rings but *not* isomorphic \(\Gamma\)-modules (because their annihilators are different). Thus the field \(k\) has two \(\Gamma\)-module structures. When necessary we distinguish between these nonisomorphic \(\Gamma\)-modules by using the notation \((k, 0)\) and \((0, k)\). However — and this is very important in the rest of this paper — we have \((k, 0) \cong (0, k)\) as \(\Lambda\)-modules (equivalently, as \(k\)-vector spaces).

**Notation 6.3** (Matrix setup, strictly split case). Let \(t \neq \infty\). Then, making the identifications \(k = \Gamma_i/m_i\) \((i = 1, 2)\) via \(\rho\), we have \(k \cong m_i^{t-1}/m_i^t\) \((i = 1, 2)\) as \(k\)-vector spaces via \(\gamma_i + m_i \rightarrow \gamma_i \pi_i^{t-1} + m_i^t\), where \(\pi_i\) is the standard \(\Gamma\)-generator of \(m_i\) mentioned in Notation 1.1. Thus every element of \(k\) has a *standard image* in each \(m_i^{t-1}/m_i^t\), which we often regard as an identification; and the set of all such standard images defines the *standard copy* of \(k\) in \(\Gamma_i/m_i^t\).

Let \(X\) be any of the nonisomorphic \(\Gamma\)-modules \((k, 0), (0, k), (k, k) = (k, 0) \oplus (0, k)\). Each of these has a \(\Lambda\)-submodule \(S\) that is \(\Lambda\)-isomorphic to \(k\) and such that \(X = \Gamma S\), namely \(X\) itself in the first two cases, and the diagonal submodule \(\text{diag}(k) = \{(x, x) \in k \oplus k\}\) in the third case. However only \(S = \text{diag}(k)\) satisfies \(\Gamma S = \Gamma \otimes \Lambda S\) [Lemma 5.2].

This shows that the notation \(\Gamma S\) is not well-defined unless one specifies the \(\Gamma\)-module \(X\) inside of which it is computed.

**Notation 6.3** (Matrix setup, strictly split case). Let \(t \neq \infty\). Then, making the identifications \(k = \Gamma_i/m_i\) \((i = 1, 2)\) via \(\rho\), we have \(k \cong m_i^{t-1}/m_i^t\) \((i = 1, 2)\) as \(k\)-vector spaces via \(\gamma_i + m_i \rightarrow \gamma_i \pi_i^{t-1} + m_i^t\), where \(\pi_i\) is the standard \(\Gamma\)-generator of \(m_i\) mentioned in Notation 1.1. Thus every element of \(k\) has a *standard image* in each \(m_i^{t-1}/m_i^t\), which we often regard as an identification; and the set of all such standard images defines the *standard copy* of \(k\) in \(\Gamma_i/m_i^t\).

Let \(X\) be the nonzero (external) direct sum of \(\Gamma\)-modules displayed below.

\[(6.3.1) \quad X = X_1 \oplus X_2 \quad \text{where} \quad X_1 = \bigoplus_{\nu=1}^n \Gamma_1/m^{s_\nu}_1 \quad \text{and} \quad X_2 = \bigoplus_{\nu=1}^n \Gamma_2/m^{t_\nu}_2\]

where each \(s_\nu\) and \(t_\nu\) is a positive integer or \(\infty\). Since each \(\Gamma_i\) is a DVR, every finitely generated \(\Gamma_i\)-module is isomorphic to a module of the form \(X_i\) \((i = 1, 2)\). However, note that the *same number* \(n\) of indecomposable summands occurs in both \(X_1\) and \(X_2\) in (6.3.1). The reason for this restriction is the following fact, which we will have many occasions to use.

\[(6.3.2) \quad \text{The } \Gamma\text{-module } Y = (\bigoplus_{\nu=1}^n \Gamma_1/m^{s_\nu}_1) \oplus (\bigoplus_{\nu=1}^n \Gamma_2/m^{t_\nu}_2) \text{ has a } \Lambda\text{-submodule } S \text{ such that } Y = \Gamma \otimes \Lambda S \text{ if and only if } m = n.\]
This follows immediately from Corollary 5.4 since — in the notation of Example 6.2 — the right hand side, when reduced modulo \( m = m_1 \oplus m_2 \), becomes \( \Gamma \)-isomorphic to \( (k,0)^{(m)} \oplus (0,k)^{(n)} \), while \( \Gamma / m \cong (k,k) \).

The map \( \rho \) yields a pair of \( \Gamma \)-linear surjections to \( k^{(n)} \) shown in the first part of (6.3.3).

(6.3.3) \( p: X_i \to k^{(n)} \ (i = 1, 2) \quad k^{(n-d_\infty)} \subseteq X_1 \quad k^{(n-e_\infty)} \subseteq X_2 \)

where \( k \) really denotes \((k,0)\) when \( i = 1 \), and \((0,k)\) when \( i = 2 \). We call coordinate \( \nu \) of \( k^{(n)} \) the coordinate corresponding to the the \( \nu^{th} \) summand \( \Gamma_1 / m_{1}^{s_{\nu}} \) of \( X_1 \). Then an analogous definition applies to \( X_2 \).

Let \( d_\infty, e_\infty \), respectively, be the number of indecomposable summands of \( X_1, X_2 \) of infinite length. Each of the \( n - d_\infty \) summands of finite length contains a standard copy of \( k \). This yields the first inclusion shown in (6.3.3).

Thus every coordinate of \( k^{(n-d_\infty)} \) has a corresponding summand \( \Gamma_1 / m_{1}^{s_{\nu}} \) of \( X_1 \), namely the summand containing this copy of \( k \). Analogous statements apply to the second displayed inclusion.

A matrix setup \( \mathcal{X} \) (for \( \Lambda \) strictly split Dedekind-like) is any three direct-sum decompositions of the form (6.3.1), together with the corresponding surjections and inclusions displayed in (6.3.3). The \( \Gamma \)-module associated with \( \mathcal{X} \) is the module \( X \) in (6.3.1). (This replaces the somewhat differently stated, but equivalent, “matrizing choices” in [L1, L3].)

**Notation 6.4** (Matrix 4-tuple \((A_1, A_2, B_1, B_2)\); associated \( \Lambda \)-module). Let \( \mathcal{X} \) be a matrix setup (strictly split case). Let \( A_1 \) and \( A_2 \) be \( n \times n \) matrices over \( k \), and let \( B_1 \) and \( B_2 \) be \( q \times (n - d_\infty) \) and \( q \times (n - e_\infty) \) matrices, respectively, over \( k \) (for some \( q \geq 0 \)).

We attach a length label to each column of these matrices as in the unsplit case: Define the length label of column \( \nu \) of \( A_1 \) to be the length \( s_{\nu} \) of its corresponding direct summand \( \Gamma_1 / m_{1}^{s_{\nu}} \) of \( X_1 \), and define the length label of each column of \( B_1 \) to be the length of its corresponding summand of \( X_1 \). Thus every column of \( B_1 \) has a corresponding column in \( A_1 \), namely the column with the same corresponding summand (of finite length) of \( X_1 \). Analogous definitions apply to \( A_2, B_2, X_2 \).

We require these matrices to satisfy:

(6.4.1) \( A_1 \) and \( A_2 \) are invertible; the rows of each \( B_i \) are linearly independent over \( k \); and the columns of each \( B_i \) with length label 1 consist of zeros.

Consequently, right multiplication by \((A_1, A_2)\) and by \((B_1, B_2)\) are \( k \)-linear monomorphisms:

(6.4.2) \( (A_1, A_2): k^{(n)} \hookrightarrow k^{(n)} \oplus k^{(n)} \quad (B_1, B_2): k^{(q)} \hookrightarrow k^{(n-d_\infty)} \oplus k^{(n-e_\infty)} \).

We call \((A_1, A_2, B_1, B_2)\) a matrix 4-tuple associated with \( \mathcal{X} \).

As in the unsplit case, we write these four matrices side by side, but we always think of each \( B_i \) as being written underneath \( A_i \), with each column
of $B_i$ written under the corresponding column of $A_i$ (and no column of $B_i$ written under columns of $A_i$ with infinite length labels). (See (7.1.1) for an example of a 4-tuple written out in this form.)

Using the maps in (6.4.2), let $S(A_1, A_2)$ be the $\Lambda$-module
\[(6.4.3)\quad S(A_1, A_2) = \{ x \in X \mid \rho(x) \in \text{im}(A_1, A_2) = k^{(n)}(A_1, A_2) \}\]
and define the $\Lambda$-module $M(A_1, A_2, B_1, B_2)$ by
\[(6.4.4)\quad M(A_1, A_2, B_1, B_2) = S(A_1, A_2) / \text{im}(B_1, B_2)\]
where $\text{im}(B_1, B_2) = k^{(q)}(B_1, B_2)$. If $q = 0$ we interpret this to mean that $M(A_1, A_2, B_1, B_2) = S(A_1, A_2)$. We call $M(A_1, A_2, B_1, B_2)$ the $\Lambda$-module associated with $(A_1, A_2, B_1, B_2)$ (with respect to the matrix setup $X$).

We note that $\Gamma \cdot S(A_1, A_2) = X$, since $A_1$ and $A_2$ are invertible. Thus $S(A_1, A_2)$ is the pullback of the commutative square in diagram (6.4.5) below.

\[
\begin{array}{c}
S(A_1, A_2) \quad \subset \quad \Gamma \cdot S(A_1, A_2) = X \\
\downarrow \rho \quad \quad \downarrow \rho \\
k^{(n)}(A_1, A_2) \quad \subset \quad (k, k)^{(n)}.
\end{array}
\]

The similarity between (6.4.5) and (5.8.3) can be enhanced by noting that we can consider the $n \times 2n$ matrix $(A_1, A_2)$ over $k$ to be an invertible $n \times n$ matrix over $\Gamma = k \oplus k$, and just calling it $A$. Although this is sometimes a helpful point of view, it will more often be useful to focus on the individual matrices $A_1$ and $A_2$. However, we cannot consider $(B_1, B_2)$ to be a matrix over $\Gamma$ because the number of columns of $B_1$ need not equal the number of columns of $B_2$.

**Definition 6.5** (Display operation, split case). We define a display operation on a matrix 4-tuple $(A_1, A_2, B_1, B_2)$ (associated with a matrix setup $X$) to be a permutation of the columns of each $A_i$, together with the same permutation of their length labels and corresponding columns of $B_i$ (so that corresponding columns of $A_i$ and $B_i$ remain corresponding columns after being moved).

Thus each display operation corresponds to the effect on $(A_1, A_2, B_1, B_2)$ of rearranging the summands $\Gamma_i/m^u_i$ in decompositions (6.3.1) of each $X_i$. For future reference we record the following consequence of this:

\[(6.5.1)\quad \text{Let } (A'_1, A'_2, B'_1, B'_2) \quad \text{— including its length labels — be obtained by performing a display operation on } (A_1, A_2, B_1, B_2). \quad \text{Then } M(A'_1, A'_2, B'_1, B'_2) \cong M(A_1, A_2, B_1, B_2), \quad \text{where the module on the left is computed with respect to the correspondingly altered matrix setup.}\]
We will need display operations to display our matrix pairs in canonical form.

**Theorem 6.6.** Let $\mathcal{X}$ be a matrix setup (strictly split case) with associated $\Gamma$-module $X$, and let $(A_1, A_2, B_1, B_2)$ be a matrix 4-tuple associated with $\mathcal{X}$. Also, let $S = S(A_1, A_2)$. Then:

(i) $\Gamma S = \Gamma \otimes_\Lambda S$.

(ii) The natural surjection $S \to M(A_1, A_2, B_1, B_2)$ is a separated cover.

Moreover,

(iii) Every finitely generated $\Lambda$-module is isomorphic to $M(A_1, A_2, B_1, B_2)$ for some matrix 4-tuple $(A_1, A_2, B_1, B_2)$ with respect to some matrix setup $\mathcal{X}$.

(iv) Let $X$ and $X'$ be the $\Gamma$-modules associated with matrix setups $\mathcal{X}$ and $\mathcal{X}'$ respectively, and let $M(A_1, A_2, B_1, B_2)$ and $M(A'_1, A'_2, B'_1, B'_2)$ be $\Lambda$-modules associated with $\mathcal{X}$ and $\mathcal{X}'$ respectively. If $M(A_1, A_2, B_1, B_2) \cong M(A'_1, A'_2, B'_1, B'_2)$ as $\Lambda$-modules then $X \cong X'$ as $\Gamma$-modules.

**Proof.** The proof is a minor modification of that of Theorem 5.10, so we omit most details except where the two situations differ.

(i) Use Lemma 5.2, remembering that $\ker \rho = mS = mX$.

(ii) The proof of Theorem 5.10(ii) works in the current context except for the proof of statement (a): $\text{im}(B_1, B_2)$ contains no nonzero $\Gamma$-submodules. If $(x, y)$ is an element of any $\Gamma$-module, then that module also contains $(x, 0)$ and $(0, y)$ (since $\Gamma = k \oplus k$). But since each of $B_1$ and $B_2$ has linearly independent rows, right multiplication by these matrices is a one-to-one map. Therefore $(x, 0) \in \text{im}(B_1, B_2)$ implies $x = 0$, and the analogous statement holds for $(0, y)$.

(iii) Let $M$ be a finitely generated $\Lambda$-module. As in the proof of Theorem 5.10(iii) we take a separated cover $\phi: S \to M$, where $S$ is a $\Lambda$-submodule of $X := \Gamma S = \Gamma \otimes_\Lambda S$. We need to prove that $S = S(A_1, A_2)$ and $\ker(\phi) = \text{im}(B_1, B_2)$ for appropriate matrices $A_1, A_2, B_1, B_2$. In fact, we do this with $A_1$ and $A_2$ equal to identity matrices. (However, we will not always have each $A_i = I$, later in this paper.)

Since the $\Lambda$-module $S$ is finitely generated [Proposition 4.11], so is the $\Gamma$-module $X = \Gamma \otimes_\Lambda S$. Since $\Gamma_1, \Gamma_2$ are DVRs, we can take $X$ to be as displayed as in (6.3.1), provided that we can show that $X_1$ and $X_2$ have the same number of indecomposable direct summands. This is done in (6.3.2). The inclusion $S \subset X$ gives us modules to use in the top row of pullback square (6.4.5) that we are building. We will soon attach matrices to $S$.

For some $n$ we have $S/mS \cong k^{(n)}$ and $X/mX \cong \Gamma^{(n)}$ as $\Lambda$- and $\Gamma$-modules, respectively, by Lemma 5.2. We use the inclusion map $k^{(n)} \subset \Gamma^{(n)}$ for the bottom row. Define the right hand vertical map to be our usual map $\rho$, and define the left-hand vertical map to be the restriction of $\rho$ to $S$. Since
\( mS = m\Gamma S = mX \), we have that \( S \) is the pullback of this diagram and, in fact, \( S = S(I_n, I_n) \) as desired.

Now we show that \( \ker(\phi) \) has the required form, \( \im(B_1, B_2) \). By Lemma 4.9, \( \ker(\phi) \) is a \( k \)-module. Choose a \( k \)-linear identification \( \ker(\phi) = k(q) \), for some \( q \). The projection of \( \ker(\phi) \) in each \( X_i \) \( (i = 1, 2) \) is again a \( k \)-submodule of \( X_1 \). Therefore the projection of \( \ker(\phi) \) in \( X_1 \) is contained in \( (k, 0)^{(n-d_{\infty})} \). Moreover, Lemma 4.9 together with the fact that we are in the strictly split case yields \( \ker(\phi) \subseteq mS = mX = m_1X_1 \oplus m_2X_2 \). Since \( m_1 \) annihilates the \( \Gamma_1 \)-module \( \Gamma_1/m_1 \), we see that the projection of \( \ker(\phi) \) in any indecomposable summand of \( X_1 \) of length 1 is zero. An analogous statement, with \( d \) replaced by \( e \), holds for the projection in \( X_2 \). This gives us two \( q \)-rowed matrices \( B_1, B_2 \) over \( k \) such that \( \ker(\phi) = k(q)(B_1, B_2) \), and every column with length label 1 consists of zeros. Moreover, right multiplication by \((B_1, B_2)\) from \( k(q) \) to \( \ker(\phi) \) is one-to-one.

All that remains to be shown now is that each of these matrices has linearly independent rows; that is, right multiplication by each of \( B_1 \) and \( B_2 \) is one-to-one. Suppose that (say) \( xB_1 = 0 \). Then \( x \cdot (B_1, B_2) = (0, x \cdot B_2) \) is contained in the \( \Gamma \)-socle of \( X_2 \). Note that every \( \Lambda \)-submodule of \( X_2 \) is a \( \Gamma \)-module. Therefore the \( \Lambda \)-module \( kx \cdot (B_1, B_2) = (0, x \cdot B_2) \) of \( \ker(\phi) \) is a \( \Gamma \)-submodule. By Lemma 4.9, it follows that \( x \cdot (B_1, B_2) = (0, 0) \). Since right multiplication by \((B_1, B_2)\) is one-to-one, we now have \( x = 0 \), as desired. This completes the proof of (iii).

(iv) This is exactly the same as the proof of Theorem 5.10(iv). \( \square \)

The final main result of this section determines when \( M(A_1, A_2, B_1, B_2) \cong M(A_1', A_2', B_1', B_2') \) in terms of the matrices involved. Because of Theorem 6.6(iv) we can restrict our attention to the situation that both matrices arise from the same associated matrix setup.

**Theorem 6.7** (Matrix Operations, strictly split case). Suppose that \((A_1, A_2, B_1, B_2)\) and \((A_1', A_2', B_1', B_2')\) are matrix 4-tuples over \( k \) associated with the matrix setup \( \mathcal{X} \) (where \( \Lambda \) is strictly split Dedekind-like with residue field \( k \)). Then \( M(A_1, A_2, B_1, B_2) \cong M(A_1', A_2', B_1', B_2') \) if and only if \((A_1, A_2, B_1, B_2)\) can be obtained from \((A_1, A_2, B_1, B_2)\) by a finite sequence of the following operations.

(i) (a) Simultaneously left multiply \( A_1 \) and \( A_2 \) by an invertible matrix over \( k \).

(b) Simultaneously left multiply \( B_1 \) and \( B_2 \) by an invertible matrix over \( k \).

(ii) (a) For any length label \( j \) right-multiply \( A_1[\text{cols } j] \) by an invertible matrix \( Q \) over \( k \) and, if \( j \neq \infty \), simultaneously right-multiply \( B_1[\text{cols } j] \) by \( Q \).
(b) For any length label \( j \) right-multiply \( A_2[\text{cols } j] \) by an invertible matrix \( Q \) over \( k \) and, if \( j \neq \infty \), simultaneously right-multiply \( B_2[\text{cols } j] \) by \( Q \).

(iii)(a) For any \( i > j \) add a \( k \)-scalar multiple of an \( i \)-labeled column of \( A_1 \) to a \( j \)-labeled column of \( A_1 \). (“Sweep toward smaller lengths in \( A_1 \).”)

(b) For any \( i > j \) add a \( k \)-scalar multiple of an \( i \)-labeled column of \( A_2 \) to a \( j \)-labeled column of \( A_2 \). (“Sweep toward smaller lengths in \( A_2 \).”)

(c) For any \( i < j \) add a \( k \)-scalar multiple of an \( i \)-labeled column of \( B_1 \) to a \( j \)-labeled column of \( B_1 \). (“Sweep toward larger lengths in \( B_1 \).”)

(d) For any \( i < j \) add a \( k \)-scalar multiple of an \( i \)-labeled column of \( B_2 \) to a \( j \)-labeled column of \( B_2 \). (“Sweep toward larger lengths in \( B_2 \).”)

We call these matrix operations “\( k-k \) sweeping-similarity” operations.

Proof. We omit the proof, which is similar to the proof of Theorem 5.11 and is given in detail in [L3, Theorems 4.6, 4.7], albeit it with many changes of notation. (See the next remark.) □

Remark 6.8 (Comparison with notation in [L3]). The translation of isomorphism of modules to a matrix problem is carried out in [L3, §4] for commutative rings called “Dedekind-like.” In the local case, this is what we call “split Dedekind-like” in [KL1], and is slightly more general than what we call “strictly split Dedekind-like” in the present paper. But in the complete local case, the two are the same. Many of the difficulties one has in reading [L3] stem from the fact that the complete local case was not completely detached from the global case, as in the present series of papers.

The four matrices that we call \( A_1, A_2, B_1, B_2 \) are called \( C^{-1}, D^{-1}, A, B \) respectively in [L3], and matrices act via left multiplication there, rather than via right multiplication, as in the present paper. Instead of making the direct module-versus-matrix connection, as done in the present paper, [L3] first connects a given module with a diagram \( \mathcal{D} \) — not the same thing that is called a diagram in the present paper — that gives a detailed picture of the separated cover (called a “separated representation” there) of the module [L3, §2] and then converts isomorphism of diagrams to a matrix problem [L3, §4]. The nature of the matrices in a matrix setup is given in [L3, Proposition 4.5], and our present Matrix Operations Theorem 6.7 is given in [L3, 4.6, 4.7]. In the strictly split local case, the case considered in the present paper, one can ignore the “global condition” in [L3, 4.7] since it becomes trivial locally. We return to this global condition, in greater generality, in [KL4].

Notation 6.9. Let \((A_1, A_2, B_1, B_2)\) and \((A'_1, A'_2, B'_1, B'_2)\) be matrix 4-tuples associated with matrix setups \( \mathcal{X} \) and \( \mathcal{X}' \) respectively (split case) [Notation 6.4]. We write \((A_1, A_2, B_1, B_2) \cong (A'_1, A'_2, B'_1, B'_2)\) to indicate that each
4-tuple can be obtained from the other by means of display operations [Definition 6.5] and $k$-k sweeping-similarity operations [Theorem 6.7].

It is easy to see that, when such an isomorphism holds, we can do the display operations before the sweeping-similarity operations. After doing the display operations and changing notation, we have $X = X'$. Moreover, when $X = X'$, we have $(A_1, A_2, B_1, B_2) \cong (A'_1, A'_2, B'_1, B'_2)$ if and only if each 4-tuple can be obtained from the other by $k$-k sweeping-similarity alone.

As in the unsplit case [Theorem 5.15], one proves:

**Theorem 6.10.** In the strictly split case, every finitely generated $\Lambda$-module is isomorphic to some $M(A_1, A_2, B_1, B_2)$ associated with some matrix setup. Moreover, $M(A_1, A_2, B_1, B_2) \cong M(A'_1, A'_2, B'_1, B'_2)$ if and only if $(A_1, A_2, B_1, B_2) \cong (A'_1, A'_2, B'_1, B'_2)$.

7. Solution of matrix problem, strictly split case.

In this section we describe the canonical form to which matrix 4-tuples can be reduced, using display operations [Definition 6.5] and sweeping-similarity operations [Theorem 6.7]. As in the previous section, this was solved in [L1, L3], although there is a better proof in [KL0]. Therefore, our object in the present section is to focus on establishing consistency with the notation of the present series, and connecting with the proof in [KL0].

Our canonical forms of indecomposable 4-tuples are of two types that we call “deleted cycle” and “block cycle” 4-tuples.

Throughout this section $(\Lambda, m, k)$ is strictly split Dedekind-like with normalization $\Gamma$, and all matrices have entries in $k$.

**Definition 7.1** (Direct sum of 4-tuples). We define the direct sum

$$(A_1, A_2, B_1, B_2) = (A'_1, A'_2, B'_1, B'_2) \oplus (A''_1, A''_2, B''_1, B''_2)$$

of matrix 4-tuples to be the expected matrix 4-tuple: $A_1 = \begin{bmatrix} A'_1 & 0 \\ 0 & A''_1 \end{bmatrix}$ with length labels those of $A'_1$ followed by those of $A''_1$, and the remaining three matrices in this 4-tuple defined similarly.

For example, consider the following matrix 4-tuple, where the numbers $3, 6, 3, \infty, 5$ are the length labels of the columns of $A_1$ as well as the corresponding columns of $B_1$, and an analogous comment applies to $A_2$ and
This 4-tuple is the direct sum of the two 4-tuples whose $A$-matrices are identity matrices of appropriate sizes and whose $B$-matrices, including their length labels, are $(B'_1, B'_2) = (3 6 3 \infty 5, 0 1 0 0 0, 0 0 1 0 0, 0 0 0 1 0, 0 0 0 0 1)$ and $(B''_1, B''_2) = (\infty 5 6 4 1, 0 1 0 0 0, 0 0 1 0 0, 0 0 0 1 0, 0 0 0 0 1)$. (7.1.2)

Note that $B''_1$ and $B''_2$ are $1 \times 1$ matrices, with no actual columns labeled by $\infty$.

**Definitions 7.2 (Canonical forms)**. We call our first canonical form a deleted cycle 4-tuple. It is determined by:

(7.2.1) A pair of label sequences $i_1, i_2, \ldots, i_d$ and $j_1, j_2, \ldots, j_d$ ($d \geq 1$), where each term is a positive integer or $\infty$, and length labels $\infty$ and 1 cannot occur except possibly as $i_1$ or $j_d$.

The deleted cycle matrix 4-tuple determined by these sequences is defined as follows. Set $A_1 = A_2 = I_d$, with the length labels of $A_1$ and $A_2$ given respectively by the two sequences in (7.2.1). The matrices $B_1$ and $B_2$ are as displayed in (7.2.2).

(7.2.2) $(B_1, B_2) = \left( \begin{array}{ccc} i_1 & \cdots & j_d \\ Z & I_{d-1} & Z' \end{array} \right)$.

Here, the dots over the first $I_{d-1}$ indicate that its columns are labeled by $i_2, \ldots, i_d$ and the dots over the second $I_{d-1}$ indicate that its columns are labeled by $j_1, \ldots, j_{d-1}$. Column $Z$ consists of zeros if the label $i_1$ is finite, but the column is not present if $i_1 = \infty$. Similarly, $Z'$ is a column of zeros if the label $j_d$ is finite, but is not present if $j_d = \infty$.

If, for example, $i_1$ is finite and $j_d$ infinite, then the sizes of $B_1$ and $B_2$ are $(d - 1) \times d$ and $(d - 1) \times (d - 1)$ respectively, and $Z'$ indicates that the $\infty$-labeled column of $A_2$ has no corresponding column in $B_2$. 

$B_2$. 

\[
\begin{bmatrix}
3 & 6 & 3 & \infty & 5 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]
As a second example, the 4-tuple displayed in (7.1.1) is the direct sum of the two deleted cycle 4-tuples whose $B$-matrices are displayed in (7.1.2).

We call our other canonical form a block cycle 4-tuple. It is determined by:

(7.2.3) (i) A pair of label sequences $i_1, i_2, \ldots, i_d$ and $j_1, j_2, \ldots, j_d$ ($d \geq 1$), where each term is a positive integer not equal to 1 (never $\infty$!), and such that the sequence of pairs $(i_1, j_1), \ldots, (i_d, j_d)$ is repetition-free.

(ii) An invertible matrix $L$ (over $k$), indecomposable under similarity, called the blocking matrix. The number of rows of $L$ is called the block size of the matrix 4-tuple.

(Recall, from (2.6.1), that a finite sequence $W$ is repetition-free if there is no strictly shorter sequence $W'$ such that $W$ consists of repetitions of $W'$.)

We display in (7.2.4) the block cycle 4-tuple determined by the data (7.2.3), and then explain it.

(7.2.4)

Let the block size be $m$. Thus the size of $L$ is $m \times m$, and each matrix in the 4-tuple has size $dm \times dm$. Matrices $A_1$ and $A_2$ have the displayed block diagonal form, with all blocks of size $m \times m$, the upper left-hand block of $A_1$ equals $L$, and all other diagonal blocks equal $I_m$. Each term of each label sequence appears $m$ times as a length label (although we display it only once here), thus labeling all $m$ columns of the block over which it is displayed. The matrix $B_1$ has identity blocks on the super-diagonal, one identity block in the lower left-hand corner, and zeros elsewhere. Thus — as suggested by the name “block cycle” — $B_1$ is a block permutation matrix whose associated permutation is a cycle. We also have $B_2 = I_{dm} = A_2$. We emphasize that $\infty$ and 1 never occur in the label sequences of a block cycle 4-tuple.
Recall [Notation 6.9] that two matrix 4-tuples are called isomorphic if one can be transformed into the other by means of display operations and k-k sweeping-similarity operations. Let $I, J$ and $I', J'$ be two pairs of finite sequences, each sequence of the same length, say $d$. Recall that $I, J$ and $I', J'$ are said to be equal modulo simultaneous cyclic permutations if there is an integer $c$ such that $I' = \nu^c(I)$ and $J' = \nu^c(J)$. (See Notation 2.3.) Finally, we call a matrix 4-tuple indecomposable if it is not isomorphic to the direct sum of two matrix 4-tuples of strictly smaller size.

**Theorem 7.3.**

(i) Every deleted cycle 4-tuple is indecomposable. Two deleted cycle 4-tuples are isomorphic if and only if they are equal — which holds if and only if they have the same pair of label sequences.

(ii) Every block cycle 4-tuple is indecomposable. Two block cycle 4-tuples are isomorphic if and only if their pairs of label sequences are equal modulo simultaneous cyclic permutations and their blocking matrices are similar.

(iii) A deleted cycle 4-tuple is never isomorphic to a block cycle 4-tuple.

Since only the similarity class of $L$ is important, $L$ can always be chosen to be the companion matrix of some power $g(X)^{e}$ of some irreducible polynomial $g(X) \neq X$ over $k$. Recall that “display operations” were defined in 6.5.

**Theorem 7.4.** Every matrix 4-tuple is isomorphic to a direct sum of deleted cycle and block cycle 4-tuples. Moreover, two direct sums of deleted cycle and block cycle 4-tuples are isomorphic if and only if they contain the same deleted cycle summands and isomorphic block cycle summands, with the same multiplicities but ignoring order of occurrence.

**Proof.** These two theorems are a special case of [KLO, 1.3-1.7]. So we limit ourselves to a few comments relating the notation here to the notation there.

First note that the sweeping-similarity operations in Theorem 6.7 form a set of sweeping-similarity operations, as described in [KLO, p. 68], with one change: All sweeping in that paper is done to the right. But this is just a matter of notation, only affecting the way in which the problem is displayed. In our situation, no actual row-sweeping occurs; that is, all rows of $A_1$ belong to a single row-block (as defined in [KLO]), and the same is true of the other three matrices.

Second, note that the present paper uses a slightly more generous definition of “display operation” than was given in [KLO, (2.5.2)]. In that earlier paper, unlike the present paper, display operations were not allowed to change the order of two columns with the same label. However, a pair of columns in $A_i$ and the corresponding pair of columns in $B_i$ may be simultaneously permuted by means of a sweeping-similarity operation. Therefore the more generous definition of “display operations” does not change what
can be done by combining display operations with sweeping-similarity operations.

Having made these observations, one then applies the aforementioned theorems from [KL0].


Throughout this section we assume that \((\Lambda, m, k)\) is unsplit Dedekind-like with normalization \((\Gamma, m, F)\). In this section we actually use the assumption (not invoked earlier in the paper) that the 2-dimensional field extension \(F\) of \(k\) is separable. [See (1.1.3).] Our objective is to find a canonical form for indecomposable matrix pairs \(p = (A, B)\) (entries in \(F\)). Since this is the most complicated section in the paper, we begin with a brief summary of what will be done.

Since \(F\) is Galois of dimension 2 over \(k\), every matrix \(C\) over \(F\) has a natural conjugate \(\overline{C}\) (with the same length labels as \(C\) if \(C\) is part of a matrix pair or 4-tuple). Given a matrix pair \(p\) over the field \(F\), our plan is to associate with \(p = (A, B)\) the matrix 4-tuple \(f(p) = (A, A, B, B)\). Our solution of the problem then falls into four parts.

(i) Show that \(p \cong p'\) if and only if \(f(p) \cong f(p')\) (Theorem 8.3). This reduces isomorphism in the unsplit case to isomorphism in the strictly split case, whose canonical forms were summarized in the previous section. But our 4-tuples now have entries in \(F\) instead of \(k\).

(ii) For each self-conjugate indecomposable 4-tuple \(g\) we describe a pair \(p\) such that \(f(p) \cong g\), and for each non-self-conjugate indecomposable 4-tuple \(g\) we describe a pair \(p\) such that \(f(p) \cong g \oplus \overline{g}\). We obtain seven types of canonical forms \(p\) in this way, and it is easy to see that each is an indecomposable pair.

Our proof that every self-conjugate indecomposable 4-tuple has the form \(f(p)\), for some matrix pair \(p\), uses a “Hilbert Theorem 90” for matrices, due to Ballantine [Lemma 8.11].

(iii) We show that the list of canonical forms in (ii) is complete. The critical theorem is that a 4-tuple \(g\) is a “package” — that is, \(g \cong f(p)\) for some pair \(p\) — if and only if \(g \cong \overline{g}\). The proof of this seems to require the list of seven canonical forms listed in (ii).

An easy consequence of all of this is that the Krull-Schmidt Theorem holds for matrix pairs.

(iv) The final main theorem in this section describes the precise extent to which our canonical forms are uniquely determined by their isomorphism class.

The results of the present section are purely matrix-theoretic results involving any pair of fields \(k\) and \(F\), with \(F\) separable of degree 2 over \(k\).
In other words, \( \Lambda \) and \( \Gamma \) are really irrelevant. In fact we use this observation at one critical point in the proof of part (i). But by keeping \( \Lambda \) and \( \Gamma \) in the background (rather than omitting them) we avoid the necessity of introducing new terminology.

**Notation 8.1.** Since \( F \) is separable of dimension 2 over \( k \), there is precisely one nonidentity automorphism of \( F \) that equals the identity on \( k \). We denote the image of \( \alpha \in F \), under this automorphism, by \( \overline{\alpha} \), the conjugate of \( \alpha \).

Let \( p = (A, B) \) be a matrix pair associated with a matrix setup \( \mathcal{X} \) (un-split case) [Notation 5.7], and recall that every column of \( A \) and \( B \) has an attached length label, a positive integer or \( \infty \). Let \( f(p) = (A, \overline{A}, B, \overline{B}) \), the matrix 4-tuple in which “bar” indicates that we take the conjugate of every entry of the matrix. The length labels of \( f(p) \) are obtained by keeping the original length labels of \( A \) and \( B \), and giving \( A \) and \( B \) the same length labels as \( A \) and \( B \), respectively. We call \( f(p) \) the matrix 4-tuple associated with \( p \). It is straightforward to verify that the length labels of \( f(p) \) are a legitimate set of length labels for some matrix setup in the strictly split case [Notation 6.3], and \( f(p) \) is a legitimate matrix 4-tuple associated with that setup [Notation 6.4].

Recall [Notation 5.14] that, for matrix pairs \( p \) and \( p' \), we write \( p \cong p' \) if \( p' \) can be obtained from \( p \) by display operations and \( k-F \) sweeping-similarity operations [Theorem 5.11].

Let \( g = (C_1, C_2, D_1, D_2) \) and \( g' = (C'_1, C'_2, D'_1, D'_2) \) be matrix 4-tuples with entries in \( F \) (associated with suitable matrix setups for a strictly split Dedekind-like ring with residue field \( F \)). We write \( g \cong g' \) if \( g' \) can be obtained from \( g \) by display operations and \( F-F \) sweeping-similarity operations, the *the analog over \( F \) of the \( k-k \) sweeping-similarity operations of Theorem 6.7*.

The *conjugate* of the matrix 4-tuple \( g = (C_1, C_2, D_1, D_2) \) is the 4-tuple \( \overline{g} = (C_2, \overline{C_1}, D_2, \overline{D_1}) \) obtained by interchanging \( C_1 \) with \( C_2 \), *together with their length labels*, doing the same with \( D_1 \) and \( D_2 \), and then taking the conjugates of all four matrices. We call the matrix 4-tuple \( g \) self-conjugate if \( g \cong \overline{g} \). We call the matrix 4-tuple \( g \) a *package* if \( g \cong f(p) \) for some matrix pair \( p \). We call a package \( g \) indecomposable if it is not the direct sum of two packages.

From the definitions, it is evident that *every package is self-conjugate*. The converse of this is also true, as we show in Theorem 8.15.

We call \( F \otimes_k F \) a *left \( F \)-algebra* when we use the scalar product \( \gamma \cdot (\alpha \otimes \beta) = \gamma \alpha \otimes \beta \). The notation \( F[[X]] \) denotes the ring of formal power series in an indeterminate \( X \) over \( F \).

**Lemma 8.2.**  (i) \( F \otimes_k F \cong F \oplus F \) as left \( F \)-algebras via the map defined by \( \tau(\alpha \otimes \beta) = (\alpha \beta, \alpha \overline{\beta}) \).
(ii) $F \otimes_k F[[X]] \cong F[[X]] \oplus F[[X]]$ as left solf-gebras via the extension
\[\tau'(\alpha \otimes f) = (\alpha f, \alpha \overline{f})\] of the map in (i), where $\overline{f}$ is the result of replacing every coefficient of $f$ by its conjugate over $k$.

Proof. (i) This well-known fact is easily proved by first choosing any $\varepsilon \in F - k$ and noting that $F = k[\varepsilon]$ since $F$ has dimension 2 over $k$. Since, in addition, $F$ is separable over $k$, the minimal polynomial of $\varepsilon$ over $k$ is $m(X) = (X - \varepsilon)(X - \overline{\varepsilon})$ and has distinct linear factors. Therefore $F \otimes_k F \cong F \otimes (k[[X]]/(m(X))) \cong F[[X]]/(m(X)) \cong F[[X]]/(X - \varepsilon) \oplus F[[X]]/(X - \overline{\varepsilon})$, and the result follows easily.

(ii) Note that $\tau'$ is obviously a homomorphism of left $F$-algebras; we need to show that $\tau'$ is a bijection. Since $F \otimes_k F$ is an $F$-$F$-bimodule, we can form the following tensor product and map:
\[\tau \otimes 1: (F \otimes_k F) \otimes_F F[[X]] \to (F \oplus F) \otimes_F F[[X]].\]
Then we can use the identification $F \otimes_F F[[X]] = F[[X]]$ to identify $\tau'$ with $\tau \otimes 1$. Since $\tau$ is an isomorphism, by (i), and $\tau'$ is the tensor product of two isomorphisms, $\tau'$ is itself an isomorphism. \qed

**Theorem 8.3.** Let $p$ and $p'$ be matrix pairs. Then $p \cong p'$ if and only if $f(p) \cong f(p')$.

Proof. This proof is an adaptation a familiar “extension of scalars” argument. The main difficulty to be surmounted is that $\Lambda$ need not be an algebra over a field.

First note that the matrix assertions in the theorem ultimately depend only on the pair of fields $k$ and $F$. Hence we can choose $\Lambda$ to be any convenient unsplit Dedekind-like ring such that the bottom row of its pullback diagram (1.1.1) is the inclusion $k \subset F$. Let $\Gamma = F[[X]]$ and $\Lambda = k + xF[[x]]$. Then $\Lambda$ is an unsplit Dedekind-like ring with maximal ideal $m = xF[[X]]$ and residue field $k$, and the normalization of $\Lambda$ is $\Gamma = (F[[X]], m, F)$. This is displayed in the first pullback square in (8.3.1) below, where $\ker \rho = m$.

\[
\begin{array}{ccc}
\Lambda & \subset & \Gamma \\
\downarrow \rho & & \downarrow \rho \\
k & \subset & F \\
\end{array}
\quad (8.3.1)
\begin{array}{ccc}
F \otimes_k \Lambda & \stackrel{\iota'}{\to} & F \otimes_k \Gamma = \Gamma \oplus \Gamma \\
\downarrow 1 \otimes \rho & & \downarrow 1 \otimes \rho \\
F & \stackrel{\iota'}{\to} & F \otimes_k F = F \oplus F.
\end{array}
\]

To obtain the second square in (8.3.1), tensor the first square with $F$ over $k$, as shown, making the identifications that we proceed to explain. Identify the $F$ in the lower left corner with $F \otimes_k k$ in the natural way, and let the two vertical maps be as shown. The two horizontal maps are simply $1 \otimes i$ ($i$ the relevant inclusion map). Then the diagram commutates.

The equalities in the upper right corner and lower right corner refer to the isomorphisms $\tau'$ and $\tau$, respectively, of Lemma 8.2, both of which we regard as identification. Then the lower horizontal map $i$ is the composition of two
identifications and $1 \otimes i$, and as such $\iota(\alpha) = (1 \otimes i)(\alpha \otimes 1) = \alpha \otimes 1 = (\alpha, \alpha)$ for each $\alpha \in F$; that is, $\iota$ is the diagonal inclusion of $F$ into $F \oplus F$.

This square is a pullback diagram because tensoring a pullback diagram by a flat module again yields a pullback diagram (use the idea in e.g. [K, Proposition 2.10] or [L3, Lemma 6.1]). This shows that $F \otimes_k \Lambda$ is the strictly split Dedekind-like ring whose pullback diagram is the second square displayed in (8.3.1).

Now we have a pair $\Lambda, \Gamma$ to use in a standard extension of scalars argument. Since the argument involves a lot of details, we describe only the main steps. Let $M, N$ be finitely generated $\Lambda$-modules. Then we claim:

$$(8.3.2)\quad M \cong N \text{ (as } \Lambda\text{-modules) } \iff F \otimes_k M \cong F \otimes_k N \text{ (as } F \otimes_k \Lambda\text{-modules)}. $$

For the proof of the nontrivial ($\iff$) assertion, first note that, since $F$ is a 2-dimensional $k$-vector space, the second isomorphism implies that $M \oplus M \cong N \oplus N$ as $\Lambda$-modules. Then use the Krull-Schmidt Theorem for strictly split Dedekind-like rings [Lemma 1.3].

Let $\Lambda' = F \otimes_k \Lambda$ and $\Gamma' = F \otimes_k \Gamma = \Gamma \oplus \Gamma$. When it is necessary to distinguish between the two direct summands of $\Gamma'$, we write $\Gamma' = \Gamma_1 \oplus \Gamma_2$. (Note that $\Gamma_1 \cong \Gamma_2$ as $F$-algebras and $\Gamma$-modules, but not as $\Gamma'$-modules, being annihilated by $(0, \Gamma)$ and $(\Gamma, 0)$, respectively.) The length labels of our given matrix pair $p$ determine the $\Gamma$-module $X$ in some matrix setup $\mathcal{X}$ for $\Lambda$, and the length labels of $f(p)$ determine the $\Gamma'$-module $X' = X_1 \oplus X_2$ in some matrix setup $\mathcal{X}'$ for $\Lambda'$. In fact, one verifies with the help of Lemma 8.2(ii), that for every indecomposable summand $Y = \Gamma/m^i$ ($1 \leq i \leq \infty$) in the definition of $X$, we have $F \otimes_k Y = Y_1 \oplus Y_2$, where $Y \cong Y_1 \cong Y_2$ as $\Gamma$-modules, (but not as $\Gamma'$-modules).

It now suffices to show that

$$(8.3.3)\quad F \otimes_k M(A, B) \cong M(A, \overline{A}, B, \overline{B}) \text{ (as } F \otimes_k \Lambda\text{-modules)} $$

where $p = (A, B)$ and hence $f(p) = (A, \overline{A}, B, \overline{B})$. For it then follows from (8.3.2) that $M(p) \cong M(p')$ if and only if $M(f(p)) \cong M(f(p'))$, after which our two theorems on modules versus matrices 5.15 and 6.10 (unsplit and split cases, respectively) imply that $p \cong p'$ if and only if $f(p) \cong f(p')$.

The proof of (8.3.3) consists of reviewing the definitions of the maps in the definitions of $M(A, B)$ and $M(A, \overline{A}, B, \overline{B})$ and noting that, for every map $\theta$ that arises in the definition of $M(A, B)$, the map $1 \otimes \theta$ is the direct sum of two corresponding maps in the definition of $M(A, \overline{A}, B, \overline{B})$, and for every entry $\beta$ of $A$ (respectively $B$), the tensor product $1 \otimes \beta$ corresponds to a pair of entries $(\beta, \overline{\beta})$ of the matrices $A$ and $\overline{A}$ (respectively $B$ and $\overline{B}$), under the map $\tau$ of Lemma (8.2)(i).

Lemma 8.4. Given $\varepsilon \in F - k$, let $Y$ be an invertible $m \times m$ matrix over the field $F$, and consider the conjugate pair of $2m \times 2m$ matrices $T$ and $\overline{T}$.
below.

\[ T = \begin{bmatrix} Y & I_m \\ \varepsilon Y & -\varepsilon I_m \end{bmatrix}, \quad \bar{T} = \begin{bmatrix} \bar{Y} & I_m \\ \varepsilon \bar{Y} & -\varepsilon I_m \end{bmatrix}. \]

Then there is an invertible matrix \( P \) over \( F \) such that \( PT \) and \( P\bar{T} \) equal the following two matrices, respectively.

\[ \begin{bmatrix} Y & 0 \\ 0 & I_m \end{bmatrix}, \quad \begin{bmatrix} 0 & I_m \\ Y & 0 \end{bmatrix}. \]

(8.4.2)

(We refer to the matrix \( T \) as a “basic tile.”)

**Proof.** We prove the lemma by performing simultaneous row operations on the matrices in (8.4.1). Subtracting \( \varepsilon \) times the first \( m \) rows from the last \( m \) rows in both matrices yields:

\[ \begin{bmatrix} Y & I_m \\ 0 & \delta I_m \end{bmatrix}, \quad \begin{bmatrix} \bar{Y} & I_m \\ \delta \bar{Y} & 0 \end{bmatrix} \]

(8.4.3)

where \( \delta = \varepsilon - \varepsilon \neq 0 \), since \( F \) is separable over \( k \). Dividing the last \( m \) rows by \( \delta \) in both matrices in (8.4.3) and then subtracting the last \( m \) rows from the first \( m \) rows yields the desired matrices (8.4.2). □

The following simple lemma helps to motivate our systematic enumeration of canonical forms for matrix pairs (but does not show that the resulting list is complete).

**Lemma 8.5.** Let \( p \) be a matrix pair such that \( f(p) \) is either an indecomposable self-conjugate 4-tuple, or else \( f(p) \cong g \oplus \bar{g} \) where \( g \) is indecomposable and non-self-conjugate. Then \( p \) is indecomposable (equivalently, \( f(p) \) is an indecomposable package).

**Proof.** Only the case \( f(p) \cong g \oplus \bar{g} \) is nontrivial. So suppose that \( p \cong q \oplus r \), and hence \( f(p) \cong f(q) \oplus f(r) \). Since \( g \) is indecomposable, so is \( \bar{g} \). Since the Krull-Schmidt Theorem holds for matrix 4-tuples [Theorem 7.4], we must therefore have, say, \( f(q) \cong g \). But, since \( f(q) \) is a package, it is obviously self-conjugate. Therefore \( g \cong \bar{g} \), a contradiction. □

The following notation will be fundamental throughout this section.

**Notation 8.6 (Sequence manipulation).** Given the finite sequence \( I = \{x_1, \ldots, x_d\} \), we let \( \mu \) be the mirror image permutation that reverses the order of \( I \); that is, \( \mu(I) = \{x_d, x_{d-1}, \ldots, x_1\} \). We let \( \nu \) be the unit forward rotation, that is, the cyclic permutation defined by \( \nu(I) = \{x_2, x_3, \ldots, x_d, x_1\} \). A cycle is any cyclic permutation of the form \( \nu^t \), where \( t \) is an integer. Thus, \( \nu^t(I) = \{x_{t+1}, x_{t+2}, \ldots, x_{t-1}, x_t\} \) (reading subscripts modulo \( d \)). We frequently use the following, easily verified relation.

\[ \nu^t \mu = \mu \nu^{-t}. \]

(8.6.1)
For finite sequences $I$ and $J$ we use the notation $\{I, J\}$ for the concatenation of $I$ and $J$, that is, the sequence consisting of the terms of $I$ followed by the terms of $J$.

**Definition 8.7** (Pairs in canonical form arising from deleted cycle 4-tuples). For each pair $p = (A, B)$ that we describe here, we shall show that $f(p)$ is isomorphic to the direct sum of one or two deleted cycle 4-tuples [Theorem 8.9]. The name that we use for each individual form is chosen to coincide with the name of the corresponding type of module, described in §2.

All blank spaces in our matrix displays — except ellipsis dots — represent zeros; and $\varepsilon$ denotes any element of $F - k$ (hence $F = k[\varepsilon]$).

The data from which each matrix pair is constructed is a pair of label sequences

\[ I = \{i_1, i_2, \ldots, i_d\} \quad \text{and} \quad J = \{j_1, j_2, \ldots, [j_d]\} \tag{8.7.1} \]

in which each term is a positive integer or $\infty$. The term $j_d$ is in brackets because, in some types, it is not present.

**Remark**. Note that we use a pair of label sequences to define the deleted cycle matrix pairs below, even though a matrix pair $(A, B)$ needs only a single label sequence for labeling the columns of $A$ (and hence the corresponding columns of $B$). As we shall see below, the two label sequences $I$ and $J$ are woven together to form a single label sequence for the matrix pairs to be defined. The reason for splitting into two sequences $I$ and $J$ in this definition is that two label sequences are needed by matrix 4-tuples. (See Theorem 8.9 below.)

**Nonreduced pair**. Label $j_d$ occurs. The label sequences must satisfy:

\[ (8.7.2) \]

(i) Only labels $i_1$ and $j_d$ can equal $\infty$ or 1.

(ii) The pair of label sequences $I$ and $J$ must be unsymmetrical, in the sense that $J \neq \mu(I)$ (the mirror image of $I$).

The form of these matrices is displayed in (8.7.3), below, in which the row of length labels at the top, comes from the pair of label sequences. Each column of $Z$’s is a column of zeros except if the length label is $\infty$, in which case that column is not present in $B$ (because no column of $B$ is allowed to have an infinite length label). Note that $A$ is a direct sum of $2 \times 2$ basic tiles (8.4.1), while $B$ is a modified such direct sum. The basic tiles that appear in $A$ always overlap the tiles that appear in $B$, as shown.

We shall show, for nonreduced pairs $p$, that $f(p)$ is isomorphic to the direct sum of two non-self-conjugate deleted cycle 4-tuples [Theorem 8.9]. The reason for nonsymmetry condition (8.7.2)(ii) is that, when it is violated, $f(p)$ becomes isomorphic to the direct sum of two self-conjugate 4-tuples, and $p$ itself is isomorphic to the direct sum of two isomorphic indecomposable pairs [Proposition 8.19].
\begin{equation}
A = \begin{bmatrix}
i_1 & j_1 & i_2 & j_2 & i_3 & \ldots & j_{d-1} & i_d & j_d \\
1 & 1 & \epsilon & \epsilon & \epsilon & \ldots & \epsilon & \epsilon & \epsilon \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
1 & \epsilon & \epsilon & \epsilon & \epsilon & \ldots & 1 & 1 & \epsilon \\
\end{bmatrix}
\end{equation}

\begin{equation}
B = \begin{bmatrix}
Z & 1 & 1 & \ldots & Z \\
Z & \epsilon & \epsilon & \ldots & Z \\
Z & 1 & 1 & \ldots & Z \\
Z & \epsilon & \epsilon & \ldots & Z \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
Z & \ldots & 1 & 1 & Z \\
Z & \ldots & \epsilon & \epsilon & Z \\
\end{bmatrix}
\end{equation}

**Bottom-reduced pair.** Label \( j_d \) occurs. We require the label sequences to satisfy:

\begin{equation}
\text{(8.7.4) Only label } i_1 \text{ can equal } \infty \text{ or } 1.
\end{equation}

(There is no symmetry restriction.) The matrices are the same as in the nonreduced case, except that \( B \) has the following additional row at its bottom.

\begin{equation}
\text{(8.7.5) } (B : ) \begin{bmatrix}
i_1 & j_1 & i_2 & j_2 & i_3 & \ldots & i_d & j_d & Z \end{bmatrix}
\end{equation}

where \( Z \) denotes zero except if \( i_1 = \infty \), in which case that column of \( B \) is not present. As usual, all entries in this row that are not explicitly shown equal 0. In the extreme case \( d = 1 \), the new bottom row consists only of \( Z \) (if it is present) and 1.

**Top-reduced pair.** In this case, label \( j_d \) is not present. We require the label sequences to satisfy (8.7.4), and there is no symmetry restriction. The matrix pair takes the form given in (8.7.6), below, in which the previously discussed conventions apply. In the extreme case \( d = 1 \), \( A \) is the one-by-one identity matrix (i.e., no basic tiles occur) and \( B \) is the one-by-one zero matrix, except if \( i_1 = \infty \), in which case the matrix \( B \) does not occur.
We shall show [Theorem 8.9] that, if $p$ is either of the reduced pairs, then $f(p)$ is a self-conjugate deleted cycle 4-tuple. First, we need to determine which deleted cycle 4-tuple is isomorphic to the conjugate of a given deleted cycle 4-tuple.

**Lemma 8.8.** Let $g$ and $h$ be the deleted cycle 4-tuples whose label sequences are:

\[(8.8.1)\]

$g$: $H = \{i_1, i_2, \ldots, i_e\}$ and $K = \{j_1, j_2, \ldots, j_e\}$

$h$: $\mu(K) = \{j_e, j_{e-1}, \ldots, j_1\}$ and $\mu(H) = \{i_e, i_{e-1}, \ldots, i_1\}$

Then $g \simeq h$. In particular, the deleted cycle 4-tuple $g$ is self-conjugate if and only if $K = \mu(H)$.

**Proof.** We have $g = (I_e, I_e, [Z I_{e-1}], [I_{e-1} Z])$ with length labels as displayed. Therefore we have $\bar{g} = (I_e, I_e, [I_{e-1} Z], [Z I_{e-1}])$, and the label sequences of $\bar{g}$ are the pair $K$ and $H$. The 4-tuple $\bar{g}$ is not a deleted cycle 4-tuple. We transform $\bar{g}$ to the deleted cycle 4-tuple $h$.

Let $h'$ be the 4-tuple obtained from $\bar{g}$ by the display operation that replaces the pair of label sequences by $\mu(K)$ and $\mu(H)$, respectively. This reverses the order of columns of each $A_i$, together with their corresponding columns in the $B$-matrices. Then $h' \cong \bar{g}$ and the label sequences are now as desired. The first and second matrices of $h'$ are now the matrix $R_e$ obtained by reversing the columns of $I_e$. The third and fourth matrices are now $[Z R_{e-1}]$ and $[R_{e-1} Z]$. Therefore, left multiplying the first and second
matrices of \( h' \) by \( R_e^{-1} \) changes these two matrices to \( I_e \) and left-multiplying the third and fourth matrices by \( R_{e-1}^{-1} \) changes them to \([Z \ I_{e-1}] \) and \([I_{e-1} \ Z] \) respectively. In other words, we have now changed \( h' \) to \( h \), as desired.

The last statement now follows from Theorem 7.3. \( \square \)

**Theorem 8.9.** Let \( p = (A,B) \) be any of the types of matrix pairs listed below, constructed from the pair of label sequences

\[
(8.9.1) \quad I = \{i_1, \ldots, i_d-1, i_d\} \quad \text{and} \quad J = \{j_1, \ldots, j_d-1, j_d\}.
\]

Then \( f(p) \) is as described below, where \( \mu \) is the mirror image permutation as in Notation 8.6.

(i) \( p \) is nonreduced. Then \( f(p) \) is isomorphic to \( g \oplus \overline{g} \) where \( g \) is the non-self-conjugate deleted cycle 4-tuple whose pair of label sequences is \( I \) and \( J \).

(ii) \( p \) is bottom-reduced. Then \( f(p) \) is isomorphic to the deleted cycle 4-tuple \( g \cong \overline{g} \) whose pair of label sequences is \( \{I, \mu(J)\} \) and \( \{J, \mu(I)\} \).

(iii) \( p \) is top-reduced. Then \( f(p) \) is isomorphic to the deleted cycle 4-tuple \( g \cong \overline{g} \) whose pair of label sequences is \( \{I, \mu(J)\} \) and \( \{J, \mu(I)\} \). (Recall that, in this case, the label \( j_d \) does not occur.)

In all cases, \( p \) is an indecomposable matrix pair (equivalently, \( f(p) \) is an indecomposable package). Conversely, if \( g \) is a non-self-conjugate deleted cycle 4-tuple, then there is some nonreduced matrix pair \( p \) with \( f(p) \cong g \oplus \overline{g} \); and if \( g \) is a self-conjugate deleted cycle 4-tuple, then there is some bottom reduced or top-reduced matrix pair \( p \) with \( f(p) \cong g \).

**Proof.** (ii) Let \( g \) and \( p \) be as stated. Our plan is to use display operations and \( F-F \) sweeping-similarity operations to transform \( f(p) \) to \( g \) (see Notation 6.9 and Theorem 6.10).

Choose a pair of rows of \( A \) and \( \overline{A} \) whose nonzero columns consist of a basic tile and its conjugate, as displayed in (8.4.1). The assertion of Lemma 8.4 is that we can transform this pair of tiles to the form displayed in (8.4.2), by simultaneously left-multiplying these two rows of \( A \) and \( \overline{A} \) by an invertible matrix \( P \) over \( k \). Since this can be accomplished by left-multiplying \( A \) and \( \overline{A} \) themselves by an invertible matrix \( P' \), it is a sweeping-similarity operation. For example, if the two rows that we are altering are rows 1 and 2, we can use the block diagonal matrix \( P' = \text{diag}(P, I, I, \ldots, I) \). Do this for every pair of basic tiles in \( f(p) \), calling the result \( h \).
The form of \( h \), when \( d = 3 \) for example, is displayed in (8.9.2) below, with our usual convention that nondisplayed entries equal zero.

(8.9.2)

\[
A_1 = \begin{bmatrix}
1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 \\
Z & 1 & 0 & 0 & 1 \\
Z & 0 & 1 & 0 & 1 \\
Z & 0 & 1 & 0 & 1 \\
Z & 1 & 0 & 1 & 0 \\
\end{bmatrix} \quad A_2 = \begin{bmatrix}
1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 \\
Z & 0 & 1 & 0 & 1 \\
Z & 1 & 0 & 1 & 0 \\
\end{bmatrix}
\]

\[
B_1 = \begin{bmatrix}
i_1 & j_1 & i_2 & j_2 & i_3 & j_3 \\
1 & 0 & 1 & 0 & 1 & 0 \\
Z & 1 & 0 & 0 & 1 & 0 \\
Z & 0 & 1 & 0 & 1 & 0 \\
Z & 0 & 1 & 0 & 1 & 0 \\
\end{bmatrix} \quad B_2 = \begin{bmatrix}
i_1 & j_1 & i_2 & j_2 & i_3 & j_3 \\
0 & 1 & 0 & 1 & 0 \\
Z & 0 & 1 & 0 & 1 \\
Z & 1 & 0 & 1 & 0 \\
Z & 1 & 0 & 1 & 0 \\
\end{bmatrix}
\]

Note that the columns of \( A_1 \) also occur in \( A_2 \), but in different locations. More precisely:

(8.9.3) For every index \( \kappa \), the \( i_\kappa \)-labeled column of \( A_1 \) equals the \( j_\kappa \)-labeled column of \( A_2 \), and the \( j_\kappa \)-labeled column of \( A_1 \) equals the \( i_\kappa \)-labeled copy of \( A_2 \).

Note that statement (8.9.3) remains true after arbitrary display operations. Let \( h' \) be the result of doing the display operation that arranges the label sequences as follows:

(8.9.4) \( \{I, \mu(J)\} \) and \( \{J, \mu(I)\} \).

Then (8.9.3) implies that the \( A_1 \) and \( A_2 \)-matrices of \( h' \) are equal (ignoring length labels). Call this matrix \( C \). It is invertible, in fact, a permutation matrix. Left-multiplying \( A_1 \) and \( A_2 \) by \( C^{-1} \) changes \( h' \) to the form \( h'' = (I, I, B''_1, B''_2) \).

Now consider the \( B \)-matrices. The \( B \)-matrices of \( h'' \) are the same as those in (8.9.2) except for having their columns permuted by a display operation. We can therefore find the form of any column of the \( B \)-matrices of \( h'' \) by finding the column with the same length label in the corresponding matrix in (8.9.2). Moreover, we can simultaneously left-multiply \( B''_1 \) and \( B''_2 \) by any permutation matrix. In other words, we can arbitrarily permute the order of occurrence of the rows of \( (B''_1, B''_2) \). Permute the rows of these two matrices so that the new \( B_2 \) takes the form \([I \ 0]\), with the newly relabeled columns. In our example \( B_2 \) is as shown below, and the length labels of \( B_1 \) are also shown in their new order. (The actual arrangement of the entries of \( B_1 \) is
yet to be determined, so we leave those entries unspecified.)

\[(8.9.5)\]

\[
B_1 = \begin{bmatrix}
i_1 & i_2 & i_3 & j_3 & j_2 & j_1 \\
* & * & * & * & * & * \\
* & * & * & * & * & * \\
* & * & * & * & * & * \\
* & * & * & * & * & *
\end{bmatrix}
\]

\[
B_2 = \begin{bmatrix}
j_1 & j_2 & j_3 & i_3 & i_2 & i_1 \\
1 & Z & 1 & Z & 1 & Z
\end{bmatrix}
\]

This can be done because, if we ignore the \(Z\)-column of \(B_2\) in (8.9.2), \(B_2\) becomes a permutation matrix.

It now suffices to show that the new \(B_1\), with its newly rearranged labels and rearranged rows, equals \([Z \ I]\). For this purpose, return to (8.9.2). The \(j_1\)-, \(j_2\)-, and \(j_3\)-labeled columns of \(B_2\) have their nonzero entries in the same rows as the \(i_2\)-, \(i_3\)-, and \(j_3\)-labeled columns, respectively, of \(B_1\). Therefore the same is true in the new \((B_1, B_2)\). In other words, the first, second, and third rows of the new \(B_1\) have their nonzero entries in columns 2,3,4 respectively, as desired. The remaining columns are treated similarly. Also, the \((i_1\text{-labeled})\) \(Z\)-column occurs in the first column of the new \(B_1\), as desired.

(iii) This is essentially the same as the proof of (ii), so we omit the details.

(i) This is similar to the proof of (ii). We sketch enough of the proof to show why we get a direct sum of two 4-tuples here, and to establish the following stronger statement that we need later.

\[(8.9.6)\] We have \(f(p) \cong g \oplus \bar{g}\), even without the nonsymmetry restriction (8.7.2)(ii) in the definition of a nonreduced pair.

Note that the only differences between a nonreduced pair and a bottom reduced pair are the presence of an extra row at the bottom of the latter, and the fact that nonreduced pairs are allowed to have the last column length-labeled by \(\infty\).

Begin by reducing all pairs of basic tiles in \(f(p)\), as in the proof of (ii), calling the result \(h\) as before. To visualize the result in the case \(d = 3\), delete the bottom row of the \(B\)-matrices in (8.9.2), and change the entries the last column of \(B_1\) and \(B_2\) to \(Z'\). Continue the reduction until \(A_1 = I = A_2\) (ignoring length labels) and — in the case \(d = 3\) — we reach the situation displayed in (8.9.5), the only difference being that the \(j_3\)-labeled column of \(B_2\) is now the \(Z'\)-column. (Note that this column is now in the middle of the matrix!) Similarly the \(i_1\)- and \(j_3\)-labeled columns of \(B_1\) now have entries \(Z\) and \(Z'\) respectively. The remaining entries of \(B_1\) are the same as in the bottom-reduced case, and this completes our proof-sketch of this case.

**Supplementary statements.** Indecomposability is a special case of Lemma 8.5. Completeness of our list follows easily from Lemma 8.8 and the definition of deleted cycle 4-tuples. \[\square\]
The next step is to do for block cycles what we just did for deleted cycles.

**Definition 8.10** (Pairs in canonical form arising from block cycle 4-tuples). The pairs arising in these definitions are constructed from a *blocking matrix* $U$ — always invertible, of size that we call the *block size* and usually write $m \times m$ — and a pair of *label sequences*:

\[(8.10.1) \quad I = \{i_1, i_2, \ldots, i_d\} \quad \text{and} \quad J = \{j_1, j_2, \ldots, [j_d]\},\]

in which each term is a positive integer not equal to 1 (never $\infty$). The term $j_d$ is in brackets because, in some types, it is not present. Moreover, *each label in (8.10.1) is the length label of $m$ consecutive columns of $A$ and $B$*, namely the $m$ columns in the block over which it appears in the diagrams that define the individual forms.

As in Definition 8.7, we weave together the two label sequences $I$ and $J$ to form a single label sequence for each of the matrix pairs defined below. The two label sequences $I$ and $J$ reappear when we consider the associated matrix 4-tuples in Theorem 8.14 below.

We shall show [Theorem 8.14] that if $p$ is any of the “reduced” pairs below, then $f(p)$ is a self-conjugate block cycle 4-tuple. In the remaining case, cycle pairs, $f(p)$ is the direct sum of two conjugate, but non-self-conjugate, block cycle 4-tuples.

*Bottom-bottom reduced pairs.* Label $j_d$ occurs. We require:

\[(8.10.2) \quad \text{(i) } UU^{-1} \text{ is indecomposable under similarity.} \]

\[(8.10.2) \quad \text{(ii) The concatenated sequence } \{I, \mu(J)\} \text{ is repetition-free; that is, it does not consist of repetitions of some strictly shorter sequence.} \]

The form of these matrices is displayed in (8.10.3), below. Here $I = I_m$ with $m$ the block size. The matrix $A$ is a direct sum of one or more basic tiles (defined in Lemma 8.4), always with $Y = I$. The matrix $B$ is the direct sum of basic tiles (one fewer than $A$) and two *exceptional tiles*, one equal to the inverse $U^{-1}$ of the blocking matrix and the other equal to $I$. (We use the inverse $U^{-1}$ rather than the blocking matrix $U$ in the matrix $B$ in order to achieve uniformity in later formulas, for example, for $L$ in the first three parts of Theorem 8.14.) The basic tiles in $B$ always overlap those in $A$, as in the canonical forms that come from deleted cycle 4-tuples. Here, as before, entries not shown (except for ellipsis dots) are zero. As mentioned above, each term of the label sequences is the length label of all $m$ rows in the block over which it appears.

The extreme case is $d = 1$, in which case $B$ contains no basic tiles.
Bottom-top reduced pairs. Label $j_d$ does not occur. Again we require (8.10.2); the pair is as displayed below.

$$A = \begin{bmatrix} i_1 & j_1 & i_2 & j_2 & i_3 & \ldots & i_{d-1} & j_{d-1} & i_d & j_d \\ I & I & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & I & I \\ I & I & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & I & I \\ I & I & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & I & I \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \ldots & \ldots & U^{-1} \\ I & \ldots & \ldots & \ldots & 0 \end{bmatrix}$$

(8.10.3)

$$B = \begin{bmatrix} I & I & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & I & I \\ I & I & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & I & I \\ I & I & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & I & I \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \ldots & \ldots & U \\ I & \ldots & \ldots & \ldots & 0 \end{bmatrix}$$

In the extreme case $d = 1$ no basic tiles occur.
Top-top-reduced pairs. These are the same as bottom-bottom-reduced pairs except that $A$ is interchanged with $B$ and the matrix $U^{-1}$ is replaced by the blocking matrix $U$. In particular, we require (8.10.2).

Cycle pairs. Label $j_d$ occurs. Here we require:

(8.10.5) (i) The blocking matrix $U$ is indecomposable under similarity.
(ii) The blocking matrix $U$ is not similar to $U^{-1}$ or, for all cycles $\nu'$, $J \neq \nu'\mu(I)$, where $\nu$ is the unit forward rotation and $\mu$ is the mirror image permutation [Notation 8.6].
(iii) The sequence of pairs $(i_1,j_1),\ldots,(i_d,j_d)$ is repetition-free.

These pairs have the form displayed in diagram (8.10.6) below. Thus, no exceptional tiles occur here. Note that both $A$ and $B$ are direct sums of basic tiles, but one of the tiles “wraps” from the end to the beginning of $B$. The blocking matrix $U$ always occurs in exactly one basic tile: The tile in the upper left corner of $A$.

\[
A = \begin{bmatrix}
U & I & & & & \\
U & I & & & & \\
I & I & \ddots & & \\
I & \ddots & \ddots & I & \\
\vdots & \ddots & \ddots & \ddots & I & \\
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
I & I & & & & \\
I & I & & & & \\
I & \ddots & \ddots & I & \\
I & \ddots & \ddots & \ddots & I & \\
I & \ddots & \ddots & \ddots & \ddots & I \\
\end{bmatrix}
\]

In order to prove that our list of canonical forms yields all isomorphism classes of self-conjugate 4-tuples, we need the following matrix version of Hilbert’s Theorem 90. It follows immediately from Ballantine [Ba, Theorem 2.1], using the fact that every matrix over a field is similar to its transpose.

Lemma 8.11 (“Theorem 90” for matrices). Let $F$ be a 2-dimensional separable extension field of a field $k$. Then the following two assertions about an invertible matrix $L$ over $F$ are equivalent.

(i) $L$ is similar to $L^{-1}$.
(ii) $L$ is similar to $UU^{-1}$ for some invertible symmetric matrix $U$ over $F$. 
Also, we need to determine the block cycle 4-tuple that is isomorphic to the conjugate of a given block cycle 4-tuple.

**Lemma 8.12.** Let $g$ and $h$ be the block cycle 4-tuples with blocking matrices $L$ and $L^{-1}$ respectively, and with label sequences:

\[
\begin{align*}
g &: \ H = \{i_1, i_2, \ldots, i_e\} \\
h &: \ \mu(K) = \{j_e, j_{e-1}, \ldots, j_1\} \quad \text{and} \quad \mu(H) = \{i_e, i_{e-1}, \ldots, i_1\}.
\end{align*}
\]

Then $g \cong h$. In particular, $g \cong \overline{g}$ if and only if $L$ is similar to $L^{-1}$ and $K = \nu'\mu(H)$ (equivalently, $H = \nu'\mu(K)$) for some cycle $\nu'$.

**Proof.** Let the size of $L$ be $m \times m$. We have $g = (A, I, P, I)$ where, as displayed in (7.2.4) with $d$ in place of the present $e$, $A$ is the $me \times me$ block diagonal matrix whose upper left main-diagonal block is $L$ and whose other main diagonal blocks equal $I_m$, and $P$ is a block-permutation matrix. Then $\overline{g} = (I, \overline{A}, I, P)$, where the length labels are moved in exactly the same way that the matrices were moved. Let $g'$ be obtained from $\overline{g}$ by reversing the order of the $m$-column blocks in all four matrices — together with their length labels — but keeping the columns within each $m$-column block in their original order in order not to disturb the blocking matrix. Since this is a display operation we have $g' \cong \overline{g}$ [Notation 6.9 and Theorem 6.10].

Let $\overline{R}$ be the block permutation matrix whose $e$ nonzero blocks each equal $I_m$, and are located on the diagonal that runs from the lower left corner to the upper right corner of $R$. Then we have $g' = (R, \overline{A}R, R, PR)$, and the label sequences for $g'$ are those of $h$, displayed in (8.12.1). Let $g''$ be the 4-tuple obtained by left-multiplying all four matrices of $g'$ by $R$. Then $\overline{g'} \cong g'$ by Theorem 6.7, and we have $g'' = (I, R\overline{A}R, I, PR)$.

Let $g'''$ be obtained by left multiplying the first and second matrices of $g''$ by $(R\overline{A}R)^{-1}$ and the third and fourth matrices by $(RP)R^{-1}$. Then $g''' \cong g''$ and we have $g''' = (R\overline{A}^{-1}R, I,RP^{-1}R, I)$.

Note that $RP^{-1}R = P$. (This is essentially formula (8.6.1).)

Next consider the form of $R\overline{A}R$. $\overline{A}$ is the block diagonal matrix whose diagonal blocks are $\overline{I}, I, I, \ldots, I$. Therefore $R\overline{A}^{-1}R$ is the block diagonal matrix whose diagonal blocks are $I, I, \ldots, I, \overline{L}^{-1}$. Thus, when $e = 5$, for example, our 4-tuple $g'''$ is as shown in (8.12.2) below.

Therefore, to complete the proof it now suffices to move $\overline{L}^{-1}$ to the upper left corner of $A_1$. This is accomplished by the following sequence of multiplications, which are $F\cdot F$ sweeping-similarity operations: Left multiply the last row block of $A_1$ and $A_2$ by $\overline{L}$; then right multiply the $i_1$-labeled columns of $A_2$ and $B_2$ by $\overline{L}^{-1}$; then left multiply the last row block of $B_1$ and $B_2$ by $\overline{L}$; then right multiply the $j_5$-labeled columns of $A_1$ and $B_1$ by $\overline{L}^{-1}$. The resulting matrix 4-tuple is the given $h$, and so $h \cong \overline{g}$.
Supplementary statement. Suppose that $g \cong \overline{g}$. By the part of the lemma already proved, $g$ is isomorphic to a block cycle 4-tuple with blocking matrix $\overline{L}^{-1}$ and pair of label sequences $\mu(K)$ and $\mu(H)$. But by the uniqueness theorem for block cycle 4-tuples [Theorem 7.3], the pair of label sequences of any block cycle 4-tuple isomorphic to $g$ must have the form $\nu^t(H)$ and $\nu^t(K)$ for some cycle $\nu^t$. This yields $\nu^t(K) = \mu(H)$, and hence $K = \nu^{-t} \mu(H) = \mu \nu^{-t}(H)$, as desired. Moreover, also by the uniqueness theorem for block cycle 4-tuples, the blocking matrix $\overline{L}^{-1}$ of $g$ must be similar to the blocking matrix $L$ of $g$.

Conversely, suppose that $L$ is similar to $\overline{L}^{-1}$ and $K = \nu^t \mu(H)$ for some cycle $\nu^t$. Then formula (8.6.1) yields $H = \nu^t \mu(K)$. Therefore, by the uniqueness theorem for block cycle 4-tuples [Theorem 7.3], the block cycle 4-tuple $g$, with blocking matrix $L$ and label sequences $H$ and $K$, is isomorphic to the block cycle 4-tuple with blocking matrix $L$ and label sequences $\nu^{-t}(H) = \mu(K)$ and $\nu^{-t}(K) = \mu(H)$, which is in turn isomorphic to $\overline{g}$, since $L$ is similar to $\overline{L}^{-1}$. □

The following lemma simplifies the way to visualize self-conjugate block cycle 4-tuples.

Lemma 8.13. Let $g$ be a self-conjugate block cycle 4-tuple. Then $g$ is isomorphic to a block cycle 4-tuple whose pair of label sequences has one of the forms:

(a) $H$ and $\mu(H)$, or
(b) $H$ and $\nu \mu(H)$

where $\mu$ is the mirror image permutation and $\nu$ is the unit forward rotation.

Let $e$ be the number of terms in $H$. If $e$ is even then only one of (a) and (b) applies to $g$. If $e$ is odd then (a) always applies.
Proof. By Lemma 8.12, \( g \) has a pair of label sequences of the form \( H \) and \( K = \nu^t \mu(H) \) for some cycle \( \nu^t \). We consider two cases, according to the parity of \( t \).

Case (a): \( t = 2s \) for some integer \( s \). Then \( K = \nu^{2s} \mu(H) = \nu^s \mu \nu^{-s}(H) \) and hence \( \nu^{-s}(K) = \mu \nu^{-s}(H) \). Therefore, by the uniqueness theorem for block cycle 4-tuples, possibility (a) holds if we replace \( H \) and \( K \) by the sequences \( \nu^{-s}(H) \) and \( \nu^{-s}(K) \), respectively.

Case (b): \( t = 2s + 1 \) for some integer \( s \). Proceed similarly to case (a).

Now suppose that \( e \) is odd. Then the cyclic group generated by \( \nu \) has odd order, and therefore contains a square root of \( \nu \). Say \( \nu = (\nu^u)^2 \). Then \( \nu^t = \nu^{2ut} \) and case (a) applies.

Finally, suppose (by way of contradiction) that \( e \) is even and both possibilities (a) and (b) hold. Thus some pair of isomorphic block cycle 4-tuples have label sequences of the form \( H \) and \( \mu(H) \), and \( H' \) and \( \nu \mu(H') \), respectively. By the uniqueness theorem for block cycle 4-tuples, there is a cycle \( \nu^t \) such that \( \nu^t(H) = H' \) and \( \nu^t \mu(H) = \nu \mu(H') \). Therefore, \( \nu^t \mu(H) = \nu \mu \nu^t(H) \). Simplifying, using (8.6.1), yields \( H = \nu^{2l-1}(H) \). Since \( e \) is even, \( \nu^{2l-1} \neq 1 \). Thus \( H \) consists of some number, say \( r \), of repetitions of some strictly shorter sequence, and hence so does \( \mu(H) \). But then the sequence of pairs of labels from \( H \) and \( \mu(H) \) is not repetition-free, contrary to the definition of a block cycle 4-tuple [see (7.2.3)(i)]. \( \square \)

Theorem 8.14. Let \( p = (A, B) \) be any of the types of matrix pairs listed below, with blocking matrix \( U \) and constructed from the pair of label sequences

\[
(8.14.1) \quad I = \{i_1, \ldots, i_{d-1}, i_d\} \quad \text{and} \quad J = \{j_1, \ldots, j_{d-1}, [j_d]\}.
\]

Then \( f(p) \) is as described below, where \( \mu \) is the mirror image permutation, and \( \nu \) is the unit forward rotation, as in Notation 8.6.

(i) \( p \) bottom-bottom reduced: Then \( f(p) \) is isomorphic to the self-conjugate block cycle 4-tuple whose blocking matrix is \( L = UU^{-1} \) and whose pair of label sequences is \( \{I, \mu(J)\} \) and \( \{J, \mu(I)\} \).

(ii) \( p \) bottom-top reduced: Then \( f(p) \) is isomorphic to the self-conjugate block cycle 4-tuple whose blocking matrix is \( L = UU^{-1} \) and whose pair of label sequences is \( \{I, \mu(J)\} \) and \( \{J, \mu(I)\} \). (Recall that, in this case, label \( j_d \) does not occur.)

(iii) \( p \) top-top reduced: Then \( f(p) \) is isomorphic to the self-conjugate block cycle 4-tuple whose blocking matrix is \( L = UU^{-1} \) and whose pair of label sequences is \( \{J, \mu(I)\} \) and \( \nu(\{I, \mu(J)\}) \).

(iv) \( p \) a cycle pair: Then \( f(p) \) is isomorphic to the direct sum \( g \oplus \overline{g} \) where \( g \) is the non-self-conjugate block cycle 4-tuple whose blocking matrix is \( L = U \) and whose pair of label sequences is \( I \) and \( J \).
In all cases, \( p \) is an indecomposable matrix pair (equivalently, \( f(p) \) is an indecomposable package). Conversely, if \( g \) is a non-self-conjugate block cycle 4-tuple, then there is some cycle pair \( p \) with \( f(p) \cong g \oplus \overline{g} \), and if \( g \) is a self-conjugate block cycle 4-tuple, then there is some bottom-bottom reduced, bottom-top reduced or top-top-reduced matrix pair \( p \) with \( f(p) \cong g \).

**Proof.** Write \( f(p) = (A, \overline{A}, B, \overline{B}) \). We do situation (i) in detail, and then the others more briefly. Situation (i) itself is a slightly more complicated version of what was done in the deleted cycle case, and we refer to that case for steps that are very similar.

(i) \( p \) is bottom-bottom reduced, as in (8.10.3). As in the deleted cycle situations we first reduce every pair of basic tiles, as described in Lemma 8.4, and call the resulting 4-tuple \( h \). The situation \( d = 3 \) is shown in (8.14.2).

\[
(8.14.2)
\]

\[
A_1 = \begin{bmatrix}
i_1 & j_1 & i_2 & j_2 & i_3 & j_3 \\
I & 0 & I & 0 & I & 0 \\
0 & I & 0 & I & 0 & I \\
I & 0 & I & 0 & I & 0 \\
0 & I & 0 & I & 0 & I \\
I & 0 & I & 0 & I & 0 \\
U^{-1} & 0 & U^{-1} & 0 & U^{-1} & 0
\end{bmatrix}
\]

\[
B_1 = \begin{bmatrix}
i_1 & j_1 & i_2 & j_2 & i_3 & j_3 \\
0 & I & 0 & I & 0 & I \\
I & 0 & I & 0 & I & 0 \\
0 & I & 0 & I & 0 & I \\
I & 0 & I & 0 & I & 0 \\
0 & I & 0 & I & 0 & I \\
0 & U^{-1} & 0 & U^{-1} & 0 & U^{-1}
\end{bmatrix}
\]

Note that statement (8.9.3) is valid here, too if we replace “column” by “\( m \)-column block,” and it obviously remains valid after arbitrary display operations. Apply the display operation — in block form — that changes the pair of label sequences in (8.14.2) to the desired sequences \( \{I, \mu(J)\} \) and \( \{J, \mu(I)\} \), and call the resulting 4-tuple \( h' \). By (8.9.3) we have \( h' = (C, C, B'_1, B'_2) \) for some \( C \). Left multiplying the \( A \)-matrices of \( h' \) by \( C^{-1} \) converts it to \( h'' = (I, I, B'_1, B'_2) \).

Next, change the pair \((U^{-1}, \overline{U}^{-1})\) to \((UU^{-1}, I) = (L^{-1}, I)\), by left-multiplying the row-block containing these matrices by \( \overline{U} \). Then right-multiply the \( j_d \)-labeled columns of \( A_1 \) and \( B_1 \) by \( L \), changing \( L^{-1} \) to \( I \) and changing the corresponding nonzero entry in the \( j_d \)-labeled column-block of \( A_1 \) to \( L \). The result of this is that, after a change of notation, \( h'' \) takes the form \((D, I, B_1, B_2)\) where \( D \) is a block-diagonal matrix one of whose diagonal
blocks equals $L$ and whose other diagonal entries equal $I$, and $B_1$ and $B_2$ are block-permutation matrices.

Now consider the pair of block-permutation matrices $(B_1, B_2)$. We can arbitrarily permute the row blocks of this pair by simultaneously left-multiplying $B_1$ and $B_2$ by a suitable permutation matrix. Multiply by $B_2^{-1}$, so that $B_2$ is replaced by an identity matrix. The result of all of the above transformations is illustrated below, in the case $d = 3$. (Ignore the entries of $B_1$ until we determine them.)

\[
\begin{aligned}
A_1 &= \begin{bmatrix} I & I & I & L & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
\end{bmatrix}, \\
A_2 &= \begin{bmatrix} I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
\end{bmatrix}, \\
B_1 &= \begin{bmatrix} I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
\end{bmatrix}, \\
B_2 &= \begin{bmatrix} I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
I & I & I & I & I & I \\
\end{bmatrix}.
\end{aligned}
\]

Next we find the locations of the nonzero entries of $B_1$. First look for the nonzero entries of the first three row-blocks. Consider column-blocks $j_1$, $j_2$, and $j_3$ of $B_2$ in (8.14.3). We ask: Which column-blocks of $B_1$ have their nonzero entries in the same rows as the nonzero entries of the $j_1$, $j_2$, $j_3$-labeled column-blocks of $B_2$? The answer — if we identify columns by their formal labels $j_\nu$ — is unchanged by simultaneous row permutations in $B_1, B_2$ and by arbitrary display operations. After consulting (8.14.2) we see that we want the $i_2, i_3$, and $j_3$-labeled column-blocks of $B_1$. Thus, returning to (8.14.3), we see that the first three row-blocks of $B_1$ have their nonzero entries in the second, third, and fourth column-blocks of $B_1$, respectively, as shown. By proceeding in this fashion, we eventually find that $B_1$ is the block form of the cyclic permutation matrix required for a block cycle 4-tuple.

In order to complete the proof, two things are needed: (a) Move $L$ to the upper left corner of $A_1$; and (b) check that the sequence of pairs of labels from $\{I, \mu(J)\}$ and $\{J, \mu(I)\}$ is repetition-free (as required in the definition of a block cycle 4-tuple).

(a) Move $L$. It suffices to slide $L$ up one level; that is, move it to the next higher $I$, for we can repeat this often enough to get $L$ to the right location. We accomplish this by the following four sweeping-similarity operations.
Right-multiply the column-block of $A_1$ and $B_1$ containing $L$ by $L^{-1}$, thus replacing the nonzero entry in that column-block of $B_1$ by $L^{-1}$. Then use a left multiplication of the row block of $(B_1, B_2)$ containing $L^{-1}$ to put $L$ into that row-block of $B_2$. Then use a right multiplication to put $L^{-1}$ into $A_2$, and finally a left multiplication to put $L$ into $A_1$.

(b) By hypothesis (8.10.2)(ii), the sequence $\{I, \mu(J)\}$ is repetition-free, which is formally stronger than the statement that the sequence of pairs of labels from $\{I, \mu(J)\}$ and $\{J, \mu(I)\}$ is repetition-free. (Actually, since the sequence $\{J, \mu(I)\}$ is just $\{I, \mu(J)\}$ written backwards, the two types of "repetition-free" are equivalent here.)

(ii) $p$ is bottom-top reduced, as in (8.10.4). We omit the details of this reduction, which is similar to that in (i).

(iii) $p$ is top-top reduced. Thus the matrix pair $p$ is obtained from the bottom-bottom reduced pair $p'$ in (8.10.3) by interchanging $A$ with $B$ and replacing $U^{-1}$ by $U$. All of steps in the reduction of $f(p')$ done in part (i) can be done with the roles of $A$ and $B$ reversed. Therefore, by the result of (i), $f(p)$ is isomorphic to the first matrix 4-tuple displayed in (8.14.4) with its pair of label sequences above it.

\[
\begin{align*}
\{I, \mu(J)\} & \rightarrow \{J, \mu(I)\} \rightarrow \{J, \mu(I)\} \rightarrow \nu^{-1}\{J, \mu(I)\} \rightarrow \{I, \mu(J)\} \\
P \quad I & \rightarrow \quad RPR \quad I & \rightarrow \quad I \quad I \quad \nu^{-1} \quad \{J, \mu(I)\} & \rightarrow \quad \{I, \mu(J)\} \\
D \quad I & \rightarrow \quad RDR \quad I & \rightarrow \quad RDRP \quad I.
\end{align*}
\]

Here $P$ is the block permutation matrix displayed as $B_1$ in (7.2.4). This matrix, acting on the right performs the inverse $\nu^{-1}$ of the unit forward column rotation. The matrix $D$ is a block diagonal matrix (see $A_1$ in (7.2.4)) whose upper left-hand block is $U^{-1}U$ and whose other diagonal blocks each equal $I_m$, by the result of part (i) with $U$ in place of $U^{-1}$.

Let $R$ ("reversal matrix") be the matrix formed by reversing the order of the columns of the $d \times d$ identity matrix and then replacing each nonzero entry by $I_m$ and each zero entry by a zero matrix.

To continue the reduction of $f(p)$, perform the display operation that reverses the order of the column blocks of $P$ and $D$, together with their length labels, while leaving the order of the columns within each block unchanged; and do the same with the two identity matrices. Then left-multiply all 4 matrices by $R$, obtaining the second displayed 4-tuple.

Note that $RPR = P^{-1}$ (this is essentially formula (8.6.1)). Therefore right-multiplying the left-hand pair of matrices and its row of column labels by $P$ is a display operation. It yields the third 4-tuple in (8.14.4).

Note that $RDR = D_1$, the block diagonal matrix obtained by moving its nonidentity block $U^{-1}U$ to the lower right-hand corner. Moreover, $D_1P = PD_2$, where $D_2$ is the block diagonal matrix obtained from $D_1$ by moving $U^{-1}U$ to a different block on the main diagonal. Right-multiplying this column-block of $PD_2$ and the identity matrix above it by $(U^{-1}U)^{-1}$ is a
sweeping-similarity operation and replaces the last 4-tuple in (8.14.4) by $(D_2^{-1}, I, P, I)$ without moving any length-labels. Note that $D_2^{-1}$ is the block diagonal matrix obtained from $D_2$ by inverting its nonidentity diagonal block $U^{-1}U$. This diagonal block of $D_2^{-1}$ therefore equals $U^{-1}$. Call this last 4-tuple $h'$. Then $h'$ is the block cycle 4-tuple in the statement of the theorem except that: (a) Its blocking matrix is in the wrong block of $A_1$; (b) the blocking matrix of $h$ is $U^{-1}U$ instead of $UU^{-1}$; and (c) the pair of label sequences is not yet exactly correct.

(a) As in the proof of part (i) [“Move $L$”] we can slide $U^{-1}U$ up to the upper left corner of $A_1$ without changing the isomorphism class of the 4-tuple. The resulting 4-tuple is now a block cycle 4-tuple (though not yet the desired one).

(b) Reversing factors in a product of two invertible matrices does not change its similarity class, and replacing the blocking matrix of a block cycle 4-tuple within its similarity class does not change the isomorphism class of the 4-tuple [Theorem 7.3(ii)]. Doing this yields the correct blocking matrix. Call this latest 4-tuple $h''$.

(c) Performing any rotation simultaneously to both of the label sequences of any block cycle matrix leaves the isomorphism class of the matrix unchanged [Theorem 7.3(ii)]. Note that the pair of label sequences of $h''$ is the same as that of the last 4-tuple displayed in (8.14.4). Therefore applying the rotation $\nu$ to these label sequences yields the desired sequences, and completes the proof of (iii).

(iv) $p$ is a cycle pair, as in (8.10.6). Unlike the previous cases, the reduction of basic tiles in this case uses the full strength of Lemma 8.4, because $U$ is located in a basic tile. Otherwise, the details are just a slight modification of the bottom-bottom reduced case, and we omit them. As in the nonreduced deleted cycle situation, the reduction in this case actually proves:

(8.14.5) We have $f(p) \cong g \oplus \overline{g}$, even without the nonsymmetry restriction (8.10.5)(ii) in the definition of a cycle pair.

Supplementary statements. Indecomposability is an immediate consequence of Lemma 8.5.

To see that our list is exhaustive, let $g$ be any block cycle 4-tuple, with blocking matrix $L$ and pair of label sequences $I'$ and $J'$.

Consider, first, the case that $g \cong \overline{g}$. Then $L$ is similar to $\overline{L}^{-1}$ [Lemma 8.12]. Therefore, by Ballantine’s Hilbert Theorem 90 [Lemma 8.11], there is a matrix $U$ such that $U\overline{U}^{-1}$ is similar to $L$. This gives us the blocking matrix $U$ to use in situations (i)–(iii) of the theorem. Again, since $g \cong \overline{g}$, we can choose $g$, within its isomorphism class, such that $J' = \mu(I')$ or $J' = \nu\mu(I')$ [Lemma 8.13]. Moreover, if the length $e$ of each of these sequences is odd,
then we can choose $J' = \mu(I')$. This yields three cases, corresponding to the enumeration of situations in the statement of the theorem:

**Case (i):** $e$ is even and $J' = \mu(I')$. Since $e$ is even we can define sequences $I$ and $J$ of length $d = e/2$ by $I' = \{I, \mu(J)\}$. Take $p$ to be the bottom-bottom reduced pair of (i), with blocking matrix $U$ and label sequences $I$ and $J$. Then the block cycle 4-tuple yielded by situation (i) has blocking matrix $L$ and label sequences $\{I, \mu(J)\} = I'$ and $\{J, \mu(I)\} = \mu(I') = J'$, as desired.

**Case (ii):** $e$ is odd and $J' = \mu(I')$. Since $e$ is odd, we can define sequences $I$ and $J$ of lengths that we call $d$ and $d - 1$ respectively, such that $I' = \{I, \mu(J)\}$. Take $p$ to be the bottom-top reduced pair of (ii), with blocking matrix $U$ and label sequences $I$ and $J$. Then the block cycle 4-tuple yielded by situation (ii) has blocking matrix $L$ and the desired label sequences $I'$ and $J'$, as in Case (i).

**Case (iii):** $e$ is even and $J' = \nu\mu(I')$. Define sequences $I$ and $J$ of length $d = e/2$ by $I' = \{J, \mu(I)\}$. Take $p$ to be the top-top reduced pair of (iii), with blocking matrix $U$ and label sequences $I$ and $J$. Then the block cycle 4-tuple yielded by situation (iii) has blocking matrix $L$ and label sequences $I' = \{J, \mu(I)\}$ and $\nu(\{I, \mu(J)\}) = \nu\mu(I') = J'$, as desired.

This leaves the case that $g$ is non-self-conjugate; and here we take $p$ to be the cycle pair of (iv). □

**Theorem 8.15.** Let $g$ be a matrix 4-tuple. The following are equivalent:

(i) $g$ is a package.

(ii) $g \cong \overline{g}$.

(iii) $g \cong \oplus_i f(p_i)$, where each $p_i$ is one of the seven canonical forms in Definitions 8.7 and 8.10.

**Proof.** We have already noted that (i) $\Rightarrow$ (ii) follows from the definitions. Also, (iii) $\Rightarrow$ (i) is immediate since each $f(p_i)$ is a package.

Thus, the only nontrivial implication is (ii) $\Rightarrow$ (iii), so suppose $g \cong \overline{g}$. By the structure theorems for decompositions of 4-tuples [Theorems 7.3 and 7.4] we have

\[(8.15.1) \quad g \cong \oplus_j g_j \quad \text{and therefore} \quad \overline{g} \cong \oplus_j \overline{g}_j\]

where each $g_j$ is either a deleted or block cycle 4-tuple, and the multiplicity of each isomorphism class of deleted or block cycle 4-tuple that occurs in the first decomposition is determined by the isomorphism class of $g$. Since the conjugate of every deleted cycle or block cycle 4-tuple is again (isomorphic to) a deleted or block cycle 4-tuple, respectively [Lemmas 8.8 and 8.12] and $g \cong \overline{g}$, it follows from Theorem 7.4 that, for each index $j$, $\overline{g}_j$ occurs (up to isomorphism) in the decomposition (8.15.1) of $g$ exactly as often as $g_j$. 


occurs. Therefore $g$ has a decomposition of the form

$$ g \cong \oplus_i h_i \quad (8.15.2) $$

where each $h_i$ is either a self-conjugate $g_j$ or the direct sum $g_j \oplus \overline{g}_j$ for some $g_j$ such that $g_j \not\cong \overline{g}_j$. Each $h_i$ is therefore isomorphic to $f(p_i)$ for some deleted or block cycle pair $p_i$ by Theorems 8.9 and 8.14, completing the proof. □

**Corollary 8.16.** Suppose $g$ is a matrix 4-tuple. Then $g$ is an indecomposable package if and only if $g$ is either:

(i) A self-conjugate, indecomposable 4-tuple; or

(ii) The direct sum of two conjugate, non-self-conjugate, indecomposable 4-tuples.

**Proof.** This follows from Theorem 8.15 and the fact that, for all of our canonical forms $p$, $f(p)$ has the form (i) or (ii) [Theorems 8.9 and 8.14]. □

**Corollary 8.17.** Every indecomposable matrix pair is isomorphic to one of the seven canonical forms in Definitions 8.7 and 8.10. Moreover, if

$$ \oplus_{i=1}^m p_i \cong \oplus_{j=1}^n q_j \quad (8.17.1) $$

where each $p_i$ and each $q_j$ is indecomposable, then $m = n$ and, after suitable renumbering, $p_i \cong q_i$ for each index $i$.

**Proof.** If $p$ is a matrix pair, then $f(p)$ is self-conjugate, so by Theorem 8.15, we can write $f(p) \cong \oplus_i f(p_i)$, where each $p_i$ is a one of our canonical forms. By Theorem 8.3, it follows that $p \cong \oplus_i p_i$, from which the first assertion of the corollary follows.

For the uniqueness claim, suppose that we are given an isomorphism of direct sums as in (8.17.1). Then by Theorem 8.3 we get that $\oplus_{i=1}^m f(p_i) \cong \oplus_{j=1}^n f(q_j)$. It suffices to show that some $q_i$, say $q_1$, satisfies $p_1 \cong q_1$ and $\oplus_{i \neq 1} p_i \cong \oplus_{i \neq 1} q_i$. Write each $f(p_i)$ and $f(q_j)$ as the direct sum of one or two indecomposable terms, as described in Theorems 8.9 and 8.14.

Let $x$ be one of the one or two indecomposable direct summands of $f(p_1)$. By the Krull-Schmidt Theorem for 4-tuples [Theorem 7.4], some indecomposable direct summand on the right-hand side, say of $f(q_1)$, must be isomorphic to $x$. If $x$ is not a package, then $x \oplus \overline{x}$ is the unique indecomposable package containing $x$ as a direct summand [by Corollary 8.16], and is therefore isomorphic to both $f(p_1)$ and $f(q_1)$. Otherwise $x$ itself is a package, and indecomposability of $f(p_1)$ and $f(q_1)$ again implies that $f(p_1) \cong f(q_1)$.

In either case the Krull-Schmidt Theorem for 4-tuples then implies that $\oplus_{i \neq 1} f(p_i) \cong \oplus_{i \neq 1} f(q_i)$, and therefore $p_1 \cong q_1$ and $\oplus_{i \neq 1} p_i \cong \oplus_{i \neq 1} q_i$ [Theorem 8.3], completing the proof. □
There remains the question of uniqueness of the seven canonical forms in Definitions 8.7 and 8.10. The next theorem establishes the extent of this uniqueness.

**Theorem 8.18.** Let $p$ and $q$ be matrix pairs in canonical form, with pairs of label sequences $I$ and $J$, and $I'$ and $J'$, respectively, and blocking matrices $U$ and $V$, respectively, if relevant. Then $p \cong q$ if and only if they are of the same type (i)-(vii) below, and the conditions listed for their type hold (with $\mu$ and $\nu$ as in Notation 8.6).

(i) (Nonreduced pair): Either $I' = I$ and $J' = J$, or $I' = \mu(J)$ and $J' = \mu(I)$.

(ii) (Bottom-reduced pair): $I' = I$ and $J' = J$.

(iii) (Top-reduced pair): $I' = I$ and $J' = J$.

(iv) (Bottom-bottom reduced pair): Either $I' = I$ and $J' = J$, or $I' = \mu(J)$ and $J' = \mu(I)$; and (in either case) $V V^{-1}$ is similar to $U U^{-1}$.

(v) (Bottom-top reduced pair): $I' = I$ and $J' = J$; and $V V^{-1}$ is similar to $U U^{-1}$.

(vi) (Top-top reduced pair): Either $I' = I$ and $J' = J$, or $I' = \mu(J)$ and $J' = \mu(I)$; and (in either case) $V V^{-1}$ is similar to $U U^{-1}$.

(vii) (Cycle pair): Either $I' = \nu^t(I)$ and $J' = \nu^t(J)$ (for some $t$) and $V$ is similar to $U$; or $I' = \nu^t(\mu(I))$ and $J' = \nu^t(\mu(I))$ (for some t) and $V$ is similar to $U^{-1}$.

**Proof.** All of the cases considered in this proof use our description of $f(p)$ in terms of $p$ [Theorems 8.9 and 8.14] and the fact that $p \cong q \iff f(p) \cong f(q)$ [Theorem 8.3].

Uniqueness of types: We claim that the type, (i)–(vii), of $p$ is determined by the isomorphism class of $p$. It suffices to show that this type is determined by the isomorphism class of $f(p)$. We use the following facts: $f(p)$ is the direct sum of one or two deleted cycle or block cycle 4-tuples; deleted cycle and block cycle 4-tuples are indecomposable [Theorem 7.3]; the Krull-Schmidt Theorem holds for direct sums of 4-tuples [Theorem 7.4]; and no deleted cycle 4-tuple is isomorphic to any block cycle 4-tuple [Theorem 7.3]. The following enumeration of cases therefore distinguishes among types of 4-tuples.

Types (i)-(iii) are the types for which $f(p)$ is a direct sum of deleted cycle 4-tuples. Among these types, Type (i) yields two summands while Types (ii) and (iii) each yield only one summand. In both of these last two types the first label sequence of $f(p)$ is $\{I, \mu(J)\}$. In Type (ii) this has even length because $I$ and $J$ have the same length, but in Type (iii) the length is odd because $j_d$ is missing from $J$.

Types (iv)-(vii) are the types built from block cycle 4-tuples. Among these, Type (vii) is a direct sum of two block cycle 4-tuples while (iv)-(vi)
each consist of a single block cycle 4-tuple. Among Types (iv)-(vi), the label sequences of \( f(p) \) have even length in Types (iv) and (vi) because \( I \) and \( J \) have the same length, and odd length in Type (v) because \( j_d \) is missing from \( J \). Thus, it remains to distinguish between Types (iv) and (vi), the most subtle of the distinctions.

Recall that both label sequences \( I \) and \( J \) of \( p \) have the same length, say \( d \), in each of situations (iv) and (vi). The label sequences of \( f(p) \) are

\[
\begin{align*}
&\{I, \mu(J)\} \text{ and } \{J, \mu(I)\} \quad \text{ in Type (iv)} \\
&\{J, \mu(I)\} \text{ and } \nu(\{I, \mu(J)\}) \quad \text{ in Type (vi)}.
\end{align*}
\]

After writing \( H = \{I, \mu(J)\} \) and \( K = \mu(H) \) these sequences become

\[
\begin{align*}
&H \text{ and } \mu(H) \quad \text{ in Type (iv)} \\
&K \text{ and } \nu\mu(K) \quad \text{ in Type (vi)}.
\end{align*}
\]

Note that \( H \) and \( K \) both have even length \( e = 2d \). Therefore, by Lemma 8.13, a self-conjugate block cycle 4-tuple with one of the pairs of label sequences displayed in (8.18.2) can never be isomorphic to one with the other pair of label sequences. This shows that no pair of Type (iv) can be isomorphic to a pair of Type (vi), and completes our proof of the uniqueness of types.

We now proceed to the remaining uniqueness properties of each individual type. In view of the uniqueness of types, we may assume, for the rest of this proof that \( p \) and \( q \) are of the same type (i)-(vii).

Types (i)-(iii). Here we frequently use, without explicitly mentioning it, that two deleted cycle 4-tuples are isomorphic if and only if their pairs of label sequences are the same [Theorem 7.3].

Type (i): Nonreduced. Here \( f(p) \cong g \oplus \overline{g} \) where \( g \) is the deleted cycle 4-tuple with label sequences \( I \) and \( J \), and hence \( \overline{g} \) is isomorphic to the deleted cycle 4-tuple with label sequences \( \mu(J) \) and \( \mu(I) \) [Lemma 8.8]. Similarly \( f(q) \cong h \oplus \overline{h} \) where the label sequences of \( h \) are \( I' \) and \( J' \), and \( \overline{h} \) is isomorphic to the deleted cycle 4-tuple with label sequences \( \mu(J') \) and \( \mu(I') \). If \( h \cong g \) then \( I' = I \) and \( J' = J \), the first possibility stated in part (i) of the theorem. Otherwise, since \( g, \overline{g}, h, \overline{h} \) are all indecomposable 4-tuples and the Krull-Schmidt Theorem holds for 4-tuples, we have \( h \cong \overline{g} \); and therefore \( I' = \mu(J) \) and \( J' = \mu(I) \), the second possibility stated in part (i) of the theorem.

Conversely, if \( p \) and \( q \) are nonreduced pairs with \( I' = I \) and \( J' = J \) — or with \( I' = \mu(J) \) and \( J' = \mu(I) \) — then reversing the above argument easily yields \( p \cong q \).

Types (ii) and (iii): Bottom-reduced and top-reduced, respectively. In both types \( f(p) \) is a single deleted cycle 4-tuple, with label sequences \( \{I, \mu(J)\} \) and \( \{J, \mu(I)\} \). A similar statement holds for \( f(q), I' \) and \( J' \).

Suppose \( p \cong q \), and hence \( f(p) \cong f(q) \). Since isomorphic deleted cycle 4-tuples have the same pairs of label sequences, we have \( \{I, \mu(J)\} = \{I', \mu(J')\} \), which is equivalent to \( I = I' \) and \( J = J' \), as desired.
Conversely, since \( f(p) \) and \( f(q) \) are each a single deleted cycle 4-tuple, if \( I' = I \) and \( J' = J \), then \( f(p) \cong f(q) \) and hence \( p \cong q \).

We now turn to the types involving block cycle 4-tuples. Let \( F \) and \( F' \) be such 4-tuples, where \( F \) has label sequences \( H \) and \( K \), and \( F' \) has label sequences \( H' \) and \( K' \). Recall [Theorem 7.3] that \( F \cong F' \) if and only both 4-tuples have similar blocking matrices and, for some \( t \),

\[
\nu^t(H) = H' \quad \text{and} \quad \nu^t(K) = K'.
\]

Type (vii): Cycle pairs. Here, \( f(p) \cong g \oplus \bar{g} \), where \( g \) is the block cycle 4-tuple with label sequences \( I \) and \( J \) and blocking matrix \( U \). Therefore \( \bar{g} \)

is isomorphic to the block cycle 4-tuple with label sequences \( \mu(J) \) and \( \mu(I) \)

and blocking matrix \( \overline{U}^{-1} \) [Lemma 8.12]. Similarly, \( f(q) \cong h \oplus \bar{h} \) where \( h \) has label sequences \( I' \) and \( J' \) and blocking matrix \( V \), and \( \bar{h} \) has label sequences \( J' \) and \( I' \) and blocking matrix \( \overline{V}^{-1} \).

As in the the analysis of Type (i) above, we have either \( h \cong g \) or \( h \cong \bar{g} \).

In the situation \( h \cong g \), we have \( I' = \nu^t(I) \) and \( J' = \nu^t(J) \) (for some \( t \)) and \( V \) is similar to \( U \) (see (8.18.3)). Similarly, in the situation \( h \cong \bar{g} \) we have \( I' = \nu^t\mu(J) \) and \( J' = \nu^t\mu(I) \) (for some \( t \)) and \( V \) is similar to \( \overline{U}^{-1} \). As in Type (i), the converse follows easily by reversing the argument.

Types (iv)–(vi): Bottom-bottom-reduced, bottom-top-reduced, and top-top-reduced, respectively. Here \( f(p) \) is isomorphic to a single block cycle 4-tuple. Its label sequences are

\[
(8.18.4) \quad \{I, \mu(J)\} \quad \text{and} \quad \{J, \mu(I)\} \quad \text{(Type (iv) or (v))}
\]

\[
\{J, \mu(I)\} \quad \text{and} \quad \nu(\{I, \mu(J)\}) \quad \text{(Type (vi))}
\]

and the blocking matrix is \( U\overline{U}^{-1} \) in all three types. Similar remarks apply to \( f(q) \).

Suppose that \( p \cong q \). Then \( f(p) \cong f(q) \) and hence their blocking matrices \( V\overline{V}^{-1} \) and \( U\overline{U}^{-1} \) are similar, as desired. Let \( f(p) \) have label sequences \( H \) and \( K \); and let \( f(q) \) have label sequences \( H' \) and \( K' \). Then (8.18.3) holds. Let \( d \) be the number of terms in the sequence \( I \). We now consider each of these three types separately.

Type (v): Bottom-top reduced. Here, the number of terms in \( \{I, \mu(J)\} \) is \( 2d - 1 \), because \( j_d \) is missing in \( J \). In view of (8.18.4), relations (8.18.3) become

\[
(8.18.5) \quad \nu^t(\{I, \mu(J)\}) = \{I', \mu(J')\} \quad \text{and} \quad \nu^t(\{J, \mu(I)\}) = \{J', \mu(I')\}.
\]

The second equation in (8.18.5) can be rewritten as \( \nu^t\mu(\{I, \mu(J)\}) = \mu(\{I', \mu(J')\}) \). By (8.6.1), we can substitute \( \nu^t\mu = \mu\nu^{-t} \), so after canceling \( \mu \), we obtain the equation \( \nu^{-t}(\{I, \mu(J)\}) = \{I', \mu(J')\} \). Combining this with the first equation in (8.18.5) then yields \( \nu^{-t}(\{I, \mu(J)\}) = \nu^t(\{I, \mu(J)\}) \), and hence \( \nu^{2t}(\{I, \mu(J)\}) = \{I, \mu(J)\} \). By the “repetition-free” condition
(8.10.2)(ii), the permutation $\nu^{2t}$ must act as the identity on the sequence \{I, $\mu(J)$\}, from which it follows that $2d - 1$ (the length of this sequence) divides $2t$. But then $2d - 1$ divides $t$, so that $\nu'$ is the identity permutation in (8.18.5). This completes the proof that $q$ satisfies the conditions for Type (v).

Type (iv): Bottom-bottom reduced. Here I and J have the same length $d$, and therefore the concatenated sequence \{I, $\mu(J)$\} has length $2d$. The same argument as for Type (v) above still applies, with one twist: When we reach the relation $\nu^{2t}(\{I, \mu(J)\}) = \{I, \mu(J)\}$, the “repetition-free” condition implies that $2d$ (rather than $2d - 1$) divides $2t$, and hence $d$ divides $t$, say $t = xd$. This yields the two possibilities for Type (iv) as follows. If $x$ is even, then $2d$ divides $t$, in which case $\nu'$ is the identity permutation in (8.18.5), and therefore $I = I'$ and $J = J'$.

Otherwise $x$ is odd, say $x = 2y + 1$ and hence $\nu' = \nu^{(2y+1)d} = \nu^d$. Then the second equation of (8.18.5) yields $\mu(I) = J'$ and $J = \mu(I')$. The second of these is equivalent to $I' = \mu(J)$. This completes the proof that $q$ satisfies the conditions for Type (iv).

Type (vi): Top-top reduced. This is similar to Type (iv). I and J have the same length $d$, and therefore the two concatenated sequences in the second row of (8.18.4) have length $2d$. Substituting these concatenated sequences into (8.18.3), and then canceling $\nu$ from the second of equation, again yields the pair of equations (8.18.5). As in Type (v) this implies $\nu^{2t}(\{I, \mu(J)\}) = \{I, \mu(J)\}$. Since the concatenated sequences have length $2d$ we finish exactly as in Type (iv), reaching the same conclusion as in Type (iv), as desired.

This completes the proof that $q$ satisfies the conditions for Type (vi), and hence for all types.

The converse parts of (iv)–(vi) are easily verified by using the fact that isomorphism of block cycle 4-tuples is implied by the equations in (8.18.3), for some $t$, together with similarity of blocking matrices. \[\square\]

We conclude this section by noting the structure of those matrix pairs that fail the “nonsymmetry conditions” in (8.7.2)(ii) or (8.10.5)(i).

**Proposition 8.19.** Let $p$ be a matrix pair. Suppose that either:

(i) $p$ would be a nonreduced pair except that its pair of label sequences has the form I and $\mu(I)$; or

(ii) $p$ would be a cycle pair except that its pair of label sequences has the form I and $\nu'\mu(I)$ for some cycle $\nu'$, and its blocking matrix $L$ is similar to $L^{-1}$.

Then $p \cong p' \oplus p'$ for some indecomposable matrix pair $p'$.

**Proof.** (ii) By (8.14.5) we have $f(p) \cong \bar{g} \oplus \bar{g}$ where $g$ is the block cycle pair whose blocking matrix is $L$ and whose label sequences are the same.
as the pair used to define \( p \), that is, \( I \) and \( \nu' \mu(I) \). By Lemma 8.12, \( g \) is self-conjugate, and therefore \( f(p) \cong g \oplus g \). Since \( g \cong \overline{g} \), the 4-tuple \( g \) is a package [Theorem 8.15]; say \( g \cong f(p') \). Then \( f(p) \cong f(p' \oplus p') \) and therefore \( p \cong p' \oplus p' \) [Theorem 8.3]. Moreover, \( g \cong f(p') \) is indecomposable since it is a block cycle 4-tuple. Therefore \( p' \) is also indecomposable.

(i) This is essentially the same as the proof of (ii), except that one uses the deleted cycle analogues, (8.9.6) and Lemma 8.8, of the corresponding block cycle results in that proof. □


In this section \((\Lambda, m, k)\) denotes an unsplit Dedekind-like ring with normalization \((\Gamma, m, F)\). We prove the structure theorems about indecomposable \(\Lambda\)-modules stated in §2.

9.1. Connection with matrices (Brief review). Recall that a “matrix setup” \( X \) is a finite external direct-sum decomposition \( X = \bigoplus_{\nu} \Gamma/m^t \nu \) (1 \( \leq \nu \leq \infty \), where \( m^\infty = 0 \), the “\( \Gamma \)-module associated with \( X \)” [Notation 5.6]. Corresponding to certain pairs \((A, B)\) of matrices over \( F \), with \( A \) invertible, we define three finitely generated \(\Lambda\)-modules,

\[
M = M(A, B) = S(A)/K(B)
\]

where

\[
S(A) = \{ x \in X \mid \rho(x) \in k^{(n)} \cdot A \}
\]

and

\[
K(B) = k^{(q)} \cdot B.
\]

The matrix \( A \) has one column for each uniserial summand \( \Gamma/m^t \) of \( X \), and \( B \) has one column for each of these uniserial summands that has finite length. Thus each column of \( A \) has an associated uniserial summand of \( X \) and an associated column of \( B \). We attach a “length label” to each column of \( A \) and of \( B \), namely the length of its associated uniserial summand of \( X \) [Definition 5.8]. We always view \( B \) as being placed under \( A \), with each column of \( B \) written under the corresponding column of \( A \) and no column of \( B \) under each column of \( A \) whose corresponding summand of \( X \) has infinite length.

Given matrix setups \( X \) and \( X' \) and associated matrix pairs \((A, B)\) and \((A', B')\), we defined \((A, B) \cong (A', B')\) to mean that either matrix pair can be obtained from the other by means of two types of matrix operations that we call “display operations” [Definition 5.9] and “\( k-F \) sweeping-similarity operations” [Theorem 5.11]. Then we proved that \((A, B) \cong (A', B')\) if and only if \( M(A, B) \cong M(A', B') \) [Theorem 5.15].

Since the \(\Lambda\)-module corresponding to the direct sum of two matrix pairs \((A, B)\) (computed with respect to the direct sum of their associated matrix setups) is obviously isomorphic to the direct sum of the associated modules \( M(A, B) \), and since every finitely generated \(\Lambda\)-module is isomorphic to some
M(A, B) [Theorem 5.10], we conclude that \( M(A, B) \) is an indecomposable \( \Lambda \)-module if and only if \((A, B)\) is an indecomposable matrix pair.

Fix an element \( \varepsilon \in F - k \), and note that we then have \( F = k[\varepsilon] \).

9.2. Proof of Theorem 2.7 (Indecomposable modules come from standard diagrams and conversely). In Definitions 8.7 and 8.10 we defined seven canonical forms of matrix pairs \((A, B)\). Each of these canonical forms has the same name as one of our seven types of standard diagrams [Definitions 2.4 and 2.6].

The supplementary statements at the end of Theorems 8.9 and 8.14 state that \( M(A, B) \) is indecomposable whenever \((A, B)\) is one of these pairs. Conversely, Corollary 8.17 states that every indecomposable finitely generated \( \Lambda \)-module is isomorphic to one of this form.

Thus is now suffices to prove:

(9.2.1) Let \((A, B)\) be a matrix pair of one of the seven canonical types. Then there is an associated standard diagram \( \mathcal{D} \) (described below), whose type has the same name as the type of \((A, B)\) and such that \( M(A, B) \cong M(\mathcal{D}) \). Moreover, the set of diagrams that arise in this way is precisely the set of standard diagrams.

Construct the diagram \( \mathcal{D} \) associated with a canonical pair \((A, B)\) as follows. This pair is associated with a matrix setup \( \mathcal{X} \) whose associated \( \Gamma \)-modules is decomposed as displayed in (2.2.1). We call the integer \( m \) in this decomposition the block size of \( \mathcal{X} \) and of \((A, B)\). Thus the pairs arising from deleted cycle 4-tuples [see Definition 8.7] all have block size 1. And the block size of any pair arising from a block cycle 4-tuple [Definition 8.10] equals the number of rows and columns in the blocking matrix of the pair.

First construct a diagram \( \mathcal{D}_0 \) (not a standard diagram!), which consists of one vertical bar for each block of \( m \) uniserial summands of \( \mathcal{X} \), with the top of each bar labeled by the common length of the uniserial summands in that block. When we refer to column blocks of \( A \) or \( B \), we always mean the \( m \)-column blocks associated with blocks of uniserial summands of \( \mathcal{X} \). Thus every basic tile that occurs in \( A \) or \( B \) has two associated vertical bars in \( \mathcal{D}_0 \), corresponding to the two column blocks that pass through it. Similarly, every “exceptional tile” (that is, nonzero \( m \times m \) block \( U, U^{-1}, \) or \( I_m \), that is not part of a basic tile) of \( A \) or \( B \) has exactly one associated vertical bar in \( \mathcal{D}_0 \).

Form \( \mathcal{D} \) by attaching edges to \( \mathcal{D}_0 \) according to the following rules.

(9.2.2) (i) For each basic tile that occurs in \( A \), connect its two corresponding vertical bars with a top-gluing edge. If \( U \) occurs in this basic tile (and hence in the first of the two column blocks of the tile), label the left-hand end of the edge with \( U \).
(ii) For each basic tile that occurs in $B$, connect the two corresponding vertical bars with a bottom-gluing edge.

(iii) For each exceptional tile that occurs in $A$ attach a top-reduction edge to the corresponding vertical bar. If the tile consists of the blocking matrix $U$, label the top-reduction edge with $U$.

(iv) For each exceptional tile that occurs in $B$, attach a bottom-reduction edge to the corresponding vertical bar. If the tile consists of the inverse of the blocking matrix, label the bottom-reduction edge with $U^{-1}$.

It is immediate that $\mathcal{D}$ is the standard diagram with the same name, label sequences and (when relevant) blocking matrix as the canonical form $(A, B)$.

Next we show that $S(A) = S(D)$. Choose any basic tile $T$ that occurs in $A$. This tile interacts with precisely two uniserial summand-blocks of $X$, namely those corresponding to the column blocks containing $T$, and $T$ is the only tile contained in $A$ that interacts with either of these summand-blocks. Let the length labels of the first and second column blocks of $T$ be $i$ and $j$ respectively, and denote the corresponding direct sum of uniserial summands of $X$ by $X_i$ and $X_j$ respectively (ignoring the slight abuse of notation that results when $X_i$ and $X_j$ have the same length; that is, $i = j$ and $X_i \neq X_j$). Thus $X_i = (\Gamma / \mathfrak{m}^i)^{(m)}$ and $X_j = (\Gamma / \mathfrak{m}^j)^{(m)}$. If $x \in X_i$ and $y \in X_j$ and $(x, y) \in S(A)$, then Formula (9.1.1) implies that $\rho(x, y) \in (k^{(m)} \oplus k^{(m)})T$.

Writing this out, using the form of a basic tile given in (8.4.1), yields

$$\begin{align*}
(\rho(x), \rho(y)) &= ((a + b\varepsilon)U, a + b\varepsilon) \quad (a, b \in k^{(m)})
\end{align*}$$

where, to save space, we use $U$ to denote either the blocking matrix or $I$, whichever is appropriate.

Writing $\alpha = a + b\varepsilon$ we see that the ordered pair in (9.2.3) equals $((\alpha U, \alpha))$ where $\alpha$ is an arbitrary element of $F^{(m)}$. Therefore $\rho(x) = \overline{\rho(y)}U$; in other words, $X_i$ and $X_j$ are top-glued as in (2.2.4).

On the other hand, consider any exceptional tile that occurs in $A$. This block equals $U$ or $I$; we write $U$ in either case. This tile interacts with the uniserial summand $X_i$ of $X$ that corresponds to the column block containing the tile, and is the only tile in $A$ that interacts with $X_i$. Take $x \in X_i$. If $x \in S(A)$ then (9.1.1) implies that $\rho(x) \in k^{(m)}U$. Thus $X_i$ is top-reduced in the sense of (2.2.7). This completes the proof that $S(A) = S(D)$.

$K(B) = K(D)$. The analysis here is very similar to the analysis of $S(A)$ except that the basic and exceptional tiles (whichever occur) result in bottom gluing and bottom reduction, respectively.

Thus we now have $M(A, B) = M(D)$. It is clear that all standard diagrams occur in this way. \qed

9.3. Proof of Theorem 2.8 (Uniqueness of standard diagrams). See Theorem 8.18. \qed
9.4. Proof of Proposition 2.9 (Moving $U$). Statement (i) was proved in (2.2.12). Let $M = M(\mathcal{D})$, where $\mathcal{D}$ is a standard diagram, except that the blocking matrix $U$ labels an arbitrary gluing or reduction edge (to be specified).

Statement (ii). We want to show that moving $U$ to the opposite end of its attached vertical bar and then changing it to $U^{-1}$ does not change the isomorphism class of $M$. If this move can be done, then it can also be done in reverse, so there is no loss of generality in assuming that we start at the top of a vertical bar. We need to consider seven situations separately (see (9.4.1)). In each situation we wish to move $U$ from position 1 to position 2, replacing it by $U^{-1}$. In most situations we place a matrix $U$ at position 3 and then compare the effects of moving it to position 1 and position 2, and determine how $U$ must then be changed to preserve the isomorphism class of $M$. We consider the seven situations in the order in which they are displayed, doing the first in more detail than the others.

![Diagram](9.4.1)

Let $\mathcal{D}$ be the diagram before the blocking matrix is moved. Thus $\mathcal{D}$ is a standard diagram except that $U$ is in a position that we shall specify. In the notation of Subsection 9.1, we have $M = S(A)/K(B)$ for the matrix pair $(A, B)$ over $F$, where $(A, B)$ is in one of our standard canonical forms — the form with the same name as $\mathcal{D}$ — except that the blocking matrix is located in a basic tile to be specified, instead of the standard place.

Our proof makes use of the matrix 4-tuple $(A, \overline{A}, B, \overline{B})$ corresponding to the given matrix pair. Recall that $(A, B) \cong (A', B')$ if and only if $(A, \overline{A}, B, \overline{B}) \cong (A', \overline{A}', B', \overline{B}')$ [Theorem 8.3]. We always assume that $M(A, B) \cong M(\mathcal{D})$ before the blocking matrix is moved, and then prove that the isomorphism continues to hold afterwards. Our proofs make use of the fact that sweeping-similarity operations do not change the isomorphism class of $(A, \overline{A}, B, \overline{B})$ [Theorem 6.7]. But recall that we use $F$-$F$ sweeping-similarity instead of $k$-$k$-similarity since our matrices have entries in $F$.

**Top-gluing edge at left (of vertical bar) to bottom-gluing edge.** This is the first situation displayed in (9.4.1). We assume that $U$ is placed in position 3. By statement (i) we can move $U$ to position 1, changing it to $U^{-1}$. Therefore it suffices to prove that we can move $U$ from position 3 to position 2, changing it to $U$.

The matrix 4-tuple in (9.4.2) shows the portion of $(A, \overline{A}, B, \overline{B})$ that we wish to alter: We wish to move $U$ from the two top tiles to the two bottom tiles and replace it with $\overline{U}$. 

![Diagram](9.4.2)
The first step is to reduce the basic tiles, using Lemma 8.4 twice. This changes them to the form shown in the 4-tuple in (9.4.3).

\[
\begin{bmatrix}
  U & I \\
  U \varepsilon & I \varepsilon \\
  I & I \\
  I \varepsilon & I \varepsilon \\
\end{bmatrix}
\] \rightarrow
\begin{bmatrix}
  \overline{U} & I \\
  \overline{U} \varepsilon & I \varepsilon \\
  I & I \\
  I \varepsilon & I \varepsilon \\
\end{bmatrix}.
\]

Next, left multiplication of the first row-block of the upper two matrices by \(U^{-1}\) and the second row-block by \(U^{-1}\) yields the form shown in (9.4.4).

\[
\begin{bmatrix}
  I & 0 \\
  0 & I \\
  I & 0 \\
  0 & I \\
\end{bmatrix}
\begin{bmatrix}
  0 & I \\
  \overline{U}^{-1} & 0 \\
  0 & I \\
  I & 0 \\
\end{bmatrix}
\]

Then, right multiplication of the second column-block of the first upper-lower pair by \(U\) and the second column-block of the second upper-lower pair by \(U\) yields the form shown in (9.4.5).

\[
\begin{bmatrix}
  I & 0 \\
  0 & I \\
  \overline{U} & 0 \\
  0 & I \\
\end{bmatrix}
\begin{bmatrix}
  0 & I \\
  I & 0 \\
  0 & I \\
  U & 0 \\
\end{bmatrix}.
\]

Finally, note that this matrix 4-tuple in (9.4.5) can also be obtained by starting with the 4-tuple in (9.4.6) and applying Lemma 8.4 to each of the two pairs of basic tiles:

\[
\begin{bmatrix}
  I & I \\
  I \varepsilon & I \varepsilon \\
  \overline{U} & I \\
  \overline{U} \varepsilon & I \varepsilon \\
\end{bmatrix}
\] \rightarrow
\begin{bmatrix}
  I & I \\
  I \varepsilon & I \varepsilon \\
  U & I \\
  U \varepsilon & I \varepsilon \\
\end{bmatrix}.
\]

This completes the proof that we can move \(U\) from position 3 to \(\overline{U}\) in position 2.

*Top-gluing edge at right (of vertical bar) to bottom-gluing edge.* This is done similarly.
**Top-gluing edge at left (of vertical bar) to bottom-reduction edge.** We want to change the first 4-tuple displayed in (9.4.7) to the second. The procedure is the same as the first situation we considered.

\[
\begin{bmatrix}
U & I \\
U\varepsilon & I\varepsilon
\end{bmatrix}
\begin{bmatrix}
\tilde{U} & I \\
\tilde{U}\varepsilon & I\varepsilon
\end{bmatrix}
\begin{bmatrix}
I & I \\
I\varepsilon & I\varepsilon
\end{bmatrix}
\begin{bmatrix}
I & I \\
I\varepsilon & I\varepsilon
\end{bmatrix}.
\] (9.4.7)

**Top-reduction edge to bottom-gluing edge at left (of vertical bar).** Interchange the roles of \( A \) and \( B \) in the previous case.

**Top-gluing edge at right (of vertical bar) to bottom-reduction edge.** Here we ignore position 3, and move \( U \) directly from position 1 to position 2. Moreover, it is easier to ignore the 4-tuple and work directly with \((A,B)\), transforming the first matrix pair in (9.4.8) to the second by means of the \( k-F \) sweeping-similarity operations of Theorem 5.11. In fact, we need only right-multiply the first column block of the first matrix pair by \( U^{-1} \) in order to obtain the second matrix pair.

\[
\begin{bmatrix}
U & I \\
U\varepsilon & I\varepsilon
\end{bmatrix}
\begin{bmatrix}
I & I \\
I\varepsilon & I\varepsilon
\end{bmatrix}.
\] (9.4.8)

**Top-reduction edge to bottom-gluing edge at right (of vertical bar).** Reverse the roles of \( A \) and \( B \) in the previous case.

**Top-reduction edge to bottom-reduction edge.** This again is the same, but replace the two basic tiles (A-matrices) in (9.4.8) by the exceptional tiles \( U \) and \( I \), respectively.

**Supplementary statement.** Now suppose that \( \mathcal{D} \) is one of the standard diagrams with a pair of reduction edges. Then the similarity invariant of \( \mathcal{D} \) is the similarity class of \( U\tilde{U}^{-1} \) rather than of \( U \) itself [Theorem 2.8]. Suppose that \( U \) is attached to some top gluing edge. Then replacing \( U \) by \( \tilde{U}^{-1} \), and not moving it, leaves the similarity invariant of \( \mathcal{D} \) unchanged. Now, by (i), moving \( \tilde{U}^{-1} \) to the other end of the gluing edge replaces it by \( U \) again. Thus, moving \( U \) from the top of one vertical bar to the top of an adjacent vertical bar does not change the isomorphism class of \( M(\mathcal{D}) \). After doing this, we can move \( U \) to the bottom of that bar, replacing it by \( U^{-1} \). If that bar has an attached gluing edge, the same reasoning as before allows us to move \( U^{-1} \) to the other end of that edge without changing the isomorphism class of \( M(\mathcal{D}) \). Since all standard diagrams are connected diagrams, a combination of these moves enable us to move \( U \) from top to top (or, similarly, bottom to bottom) of any pair of vertical bars.

**Proposition 9.5.** Suppose that one of the following holds.

(i) \( \mathcal{D} = \mathcal{D}_{\text{Nd}} \), except that Condition (2.4.2)(iii) fails (that is, \( J = \mu(I) \) holds); or
(ii) \( D = D_{Cy} \), except that Condition (2.6.4)(ii) fails (that is, we have both \( U \sim U^{-1} \) and \( J = \nu' \mu(I) \) for some cycle \( \nu' \)).

Then \( M(D) \) is the direct sum of two isomorphic indecomposable \( \Lambda \)-modules.

Proof. As at the end of Subsection 9.2, we have \( M(A, B) = M(D) \) for an appropriate \( (A, B) \). Therefore this proposition is a module-theoretic translation of Proposition 8.19. □

9.6. Proof of Theorem 2.11 (\( \Lambda \)-homomorphisms versus \( \Gamma \)-homomorphisms). Recall that a “separated \( \Lambda \)-module” is a \( \Lambda \)-submodule of some \( \Gamma \)-module, and that a “separated cover”, the fundamental structure on which this paper is built, is defined in Definition 4.6.

Now let \( \phi: S(D) \to M(D) \) be the natural homomorphism. We claim that \( \phi \) is a separated cover. We already observed that \( M(D) \cong M(A, B) = S(A)/K(B) \) for a suitable matrix pair \( (A, B) \) [(9.1.1)], and, at the end of Subsection 9.2, we showed that \( S(A) = S(D) \) and \( K(B) = K(D) \). Therefore the claim follows from Theorem 5.10(ii). Similarly, \( \phi': S(D') \to M(D') \) is a separated cover.

Our main “almost functorial property” of separated covers [Theorem 4.12] states that any homomorphism \( f \) of finitely generated \( \Lambda \)-modules can be lifted to a \( \Lambda \)-homomorphism \( f^\ast \) of their separated covers, and this lifting preserves monomorphisms and surjections. Thus it now suffices to further extend \( f^\ast S(D') \to S(D) \) to a \( \Gamma \)-homomorphism \( f^{\ast*}: X(D') \to X(D) \).

First note that \( X(D) = \Gamma \cdot S(D) \) and \( X(D') = \Gamma \cdot S(D') \), by the second equation in (9.1.1), because the matrices \( A \) and \( A' \) are invertible. Now \( \Gamma \cdot S(D) \cong \Gamma \otimes_\Lambda S(D) \) and \( \Gamma \cdot S(D') \cong \Gamma \otimes_\Lambda S(D') \), by Lemma 5.2, so that \( f^{\ast*} = 1 \otimes f^\ast \) is the desired extension of \( f^\ast \). □

Remark 9.7 (on repetition-freeness). When the condition on repetition-freeness (2.6.4)(iv) fails for cycle diagrams, it is easy to see that a larger block size can be used, with a correspondingly different blocking matrix. The new, larger blocking matrix can fail to be indecomposable under similarity, in which case \( M(D) \) decomposes.

When the repetition-freeness condition (2.6.3)(iii) fails for one of the other three diagrams that occur with arbitrary block size, a similar (but slightly more complicated) thing happens. Let \( p \) be the matrix pair for which the condition fails. Then, in the associated 4-tuple \( f(p) \), a larger block size can be used. Let its (larger) blocking matrix be \( L \). Our desire for a canonical form then allows us to disregard the smaller block size. Changing to the larger blocking matrix \( L \) can also change the type of diagram that we are dealing with. Moreover, the new, larger blocking matrix might be decomposable under similarity, which easily shows that \( f(p) \) decomposes. If this decomposition of \( f(p) \) contains a pair of mutually conjugate summands or a single self-conjugate summand, and contains additional terms, then \( f(p) \) is
a direct sum of two packages, and hence $p$ decomposes. In this case, $M(p)$ decomposes as well.


In this section $(\Lambda, m, k)$ denotes a strictly split Dedekind-like ring with normalization $\Gamma$. The proofs in this section are minor modifications of the corresponding similar — but not identical — results in the unsplit case. The purpose of the truncated proofs that follow is to give a directory to the needed earlier results in this paper.

10.1. Connection with matrices (Brief review). Let $M$ be any finitely generated $\Lambda$-module. Then $M \cong M(A_1, A_2, B_1, B_2)$ where $(A_1, A_2, B_1, B_2)$ is a matrix 4-tuple (over $k$) associated with some matrix setup $\mathcal{X}$ [Notation 6.3 – Theorem 6.6]. Moreover, two such modules are isomorphic if and only if their corresponding matrix 4-tuples can be obtained from each other by means of $k$-$k$ sweeping-similarity operations and display operations [Theorem 6.10].

Now suppose that $M$ is indecomposable. Then, after suitable $k$-$k$ sweeping-similarity operations and display operations, $(A_1, A_2, B_1, B_2)$ becomes either a deleted cycle 4-tuple or a block cycle 4-tuple [Theorem 7.4], and each 4-tuple of either of these types yields an indecomposable $M$ [Theorem 7.3].

10.2. Proof of Theorem 3.5 (Indecomposable modules versus standard diagrams). The first step is to show that each deleted cycle or block cycle 4-tuple yields, respectively, a module $M(D_{DCY})$ or $M(D_{BCY})$, and this is an easy consequence of the definitions of these two types of 4-tuples [Definitions 7.2] and diagrams [Definitions 3.3]. Thus it now suffices to prove the uniqueness assertions of Theorem 3.5, and these follow from Theorem 7.3.

10.3. Proof of Proposition 3.6 (Moving $U$). This proof refers to the block cycle 4-tuple displayed in diagram (7.2.4), where the blocking matrix is called $L$ instead of $U$. The matrix operations we use are among the $k$-$k$ sweeping-similarity operations listed in Theorem 6.7.

We can simultaneously left multiply the first $m$ rows of $A_1$ and $B_1$ by $L^{-1}$, thus replacing the $L$ in the upper left corner of $A_1$ by $I$ and replacing the $I$ in the upper left corner of $A_2$ by $L^{-1}$. Therefore, we have moved $L$ from the left side of a top-gluing edge in $D$ to the right side of that edge, replacing $L$ by $L^{-1}$, as desired.

Once $L^{-1}$ is in this new position, we simultaneously right multiply the $j_1$-labeled columns of $A_2$ and $B_2$ by $L$, thus replacing $L^{-1}$ by $I$ and replacing the $I$ below it (in $B_2$) by $L$. Therefore, we have moved $L^{-1}$ from the top
of the $j_1$-labeled bar in $\mathcal{D}$ to the bottom of that bar, replacing it by $L$, as desired. Note that $L$ is now attached to the bottom-gluing edge attached to the $j_1$-labeled vertical bar.

Analogous simultaneous row operations in $B_1$ and $B_2$ now move $L$ from one end of its attached bottom-gluing edge to the other, replacing it by $L^{-1}$. Continuing in the fashion around the entire cycle completes the proof of the proposition.

10.4. Proof of Theorem 3.8 ($\Lambda$-homomorphisms versus $\Gamma$-homomorphisms). We have $M = S/K$ where $S = S(\mathcal{D})$ and $K = K(\mathcal{D})$, as in (3.2.9). If we can show that the natural homomorphism $S \to S/K$ is a separated cover of $S/K \cong M$, then we can just repeat the proof, in Subsection 9.6, of the corresponding theorem in the unsplit case. For this, see Theorem 6.6. □


In this section we complete our commutative noetherian tame-wild theorem (complete local case) by describing all indecomposable finitely generated modules over Klein rings (defined below). We also show how to use our description of indecomposable modules over Dedekind-like rings [§2, §3] to describe modules over homomorphic images of these rings. The point here is that our construction of indecomposable $\Lambda$-modules given in §2 and §3 involves both $\Lambda$ and its normalization $\Gamma$, and so we need to describe how to deal with $\Gamma$ when passing to homomorphic images of $\Lambda$.

We use the commutative case of the following well-known result. (For the finitely generated case, which is all we need, see for example [CR1, Theorem 6.30].)

Lemma 11.1. Let $A$ be a quasi-Frobenius (i.e., artinian self-injective) ring with left socle $H$. Then every left $A$-module is the direct sum of a projective $A$-module and and an $A/H$-module.

Definition 11.2 (Klein rings). For any module $M$ over a ring $\Omega$, we let $\mu_{\Omega}(M)$ denote the minimal number of generators required by $M$.

We call the artinian local ring $(\Omega, n, k)$ a Klein ring if $\mu_{\Omega}(n) = 2$, $\mu_{\Omega}(n^2) = 1$, $n^3 = 0$, and every element of $n$ has square 0. (See [KL1, Introduction, Theorem 2.10, and §5] for more about these rings.)

The next result reduces the description of modules over Klein rings to modules over homomorphic images of Dedekind-like rings.

Theorem 11.3. Suppose that $(\Omega, n, k)$ is a Klein ring; then the following hold.

(i) $\Omega$ is a quasi-Frobenius ring with simple socle $n^2$, and $k$ has characteristic 2.
\( \Omega/n^2 \) is a homomorphic image of a strictly split Dedekind-like ring \((\Lambda, m, k)\) which is \(\mathfrak{m}\)-adically complete; in fact, \(\Omega/n^2 \cong \Lambda/m^2\).

(iii) Every \(\Omega\)-module is the direct sum of a free module and an \(\Omega/n^2 \cong \Lambda/m^2\)-module.

Proof. (i) Since \(n^3 = 0\) we have \(n^2 \subseteq \mathrm{soc} \Omega\). We can write \(n = (x, y)\). From the definition, we conclude that \(x^2 = y^2 = 0\), but \(xy \neq 0\) since \(\mu_\Omega(n^2) = 1\), so in fact \(n^2 = (xy)\). Thus \(n^2\) is a simple submodule of \(\mathrm{soc} \Omega\). To see that \(n^2\) is the entire socle of \(\Omega\) suppose that \((ax + by)\in n^2\). Then \(n = (x, y)\) shows that \(axy = 0 = bxy\), and hence both \(a\) and \(b\) are nonunits, that is, elements of \(n\). Therefore, \(ax + by \in n^2\), as desired.

Also from the definition of Klein ring, it follows that \(0 = (x + y)^2 = x^2 + 2xy + y^2 = 2xy\), from which we conclude that \(2 \in n\), and hence the residue field \(k\) has characteristic 2. Since \(\Omega\) is artinian and local with simple socle, \(\Omega\) is quasi-Frobenius [F, Theorem 3.1].

(ii) \(\Omega/n^2\) is a homomorphic image of a Dedekind-like ring \((\Lambda, m, k)\) of the claimed form, by [KL1, Proposition 3.4]. Say \(\Omega/n^2 \cong \Lambda/I\). Since the maximal ideal of \(\Omega/n^2\) has square zero, we have \(I \subseteq m^2\). To see that equality holds it suffices to show that \(\Omega/n^2\) and \(\Lambda/m^2\) both have composition length 3. This holds because \(\mu_\Omega(n)\) and \(\mu_\Lambda(m)\) both equal 2, and hence each maximal ideal modulo its square has \(k\)-dimension 2.

(iii) This follows from statements (i) and (ii), together with Lemma 11.1 and the fact that projective modules over local rings are free. \(\Box\)

The rest of this section deals with modules over homomorphic images of unsplit or strictly split Dedekind-like rings.

Lemma 11.4. Let \(I \neq 0\) be an ideal of an unsplit or strictly split Dedekind-like ring \((\Lambda, m, k)\) with normalization \(\Gamma\). Then exactly one of the following holds.

(i) \(\mu_\Lambda(I) = 1\), \(I\) contains regular elements of \(\Lambda\), and \(\Lambda(\Gamma I)/I \cong k\).

(ii) \(\mu_\Lambda(I) = 2\), \(I\) contains regular elements, and \(\Gamma I = I\).

(iii) \(\mu_\Lambda(I) = 1\), \(I\) contains no regular elements of \(\Lambda\), and \(\Gamma I = I\).

The module \(\Lambda(\Lambda/I)\) has finite length in situations (i) and (ii), but not (iii).

In situation (i), \((\Gamma I)/I = \mathrm{soc} \Lambda(\Lambda/I)\).

Proof. Recall that \(\Gamma/m\) is a 2-dimensional \(k\)-vector space in both the unsplit and strictly split cases [Notation 1.1]. It follows from Nakayama’s lemma that \(\mu_\Lambda(\Lambda \Gamma) = 2\). Again by Nakayama’s lemma, we have

\[
0 \neq I/(mI) \subsetneq (\Gamma I)/(mI).
\]

Note that the \(\Lambda\)-modules on both sides of the inclusion in (11.4.1) are annihilated by the ideal \(m\) and are therefore \(k\)-vector spaces.

Consider the case that \(I\) contains regular elements of \(\Lambda\). These are also regular elements of the principal ideal ring \(\Gamma\), and therefore \((\Gamma I)/(mI) \cong \]
Γ/\mathfrak{m} as \Lambda-modules, hence has dimension 2 as a \( k \)-vector space. Suppose \( \Gamma I \neq I \). Then the inclusion in (11.4.1) is strict. Since the right-hand side has \( k \)-dimension 2, \( I/(\mathfrak{m}I) \) therefore has \( k \)-dimension 1, and therefore, by Nakayama's lemma, \( \mu_\Lambda(I) = 1 \). Moreover, these dimensions show that \( (\Gamma I)/I \) has \( k \)-dimension 1 as claimed. Thus all of the assertions in situation (i) hold in this case. On the other hand, suppose that \( \Gamma = \mathfrak{m} \). Then the inclusion in (11.4.1) becomes equality, showing that \( I/(\mathfrak{m}I) \) has \( k \)-dimension 2, and hence \( \mu_\Lambda(I) = 2 \). Thus situation (ii) holds.

This leaves the case that \( I \) consists of zero divisors, and hence the Dedekind-like ring \( \Lambda \) is strictly split. It follows that \( I \) is contained in one of the two coordinate rings of \( \Gamma \). In this situation situation (iii) is easily seen to hold, and \( \Lambda/I \) does not have finite length.

Now we claim that, in situation (i),

\[
\text{(11.4.2)} \quad \text{soc}_\Lambda(\Lambda/I) = (\Gamma I)/I.
\]

We already have the inclusion \( \supseteq \) by (i). Let \( \text{soc}(\Lambda/I) = Y/I \). Since \( \mathfrak{m} \) annihilates the simple \( \Lambda \)-module, we have \( \mathfrak{m}Y \subseteq I \). Note that the left-hand side is a \( \Gamma \)-module while the right-hand side is not (since \( \Gamma I \neq I \) in situation (i)). Therefore strict inequality holds. Moreover \( \mathfrak{m}I \) is the unique maximal \( \Lambda \)-submodule of \( I \), since \( I \) is principal and generated by a regular element. It follows from these two observations that \( \mathfrak{m}Y \subseteq \mathfrak{m}I \). Substituting \( \mathfrak{m} = \pi \Gamma \), and canceling the regular element \( \pi \) then shows that \( Y \subseteq \Gamma Y \subseteq \Gamma I \), completing the proof of the claim.

The finite length assertion about situations (i) and (ii) follows from the facts that \( \Lambda \) has Krull dimension 1 and \( I \) contains regular elements. \( \square \)

**Definition 11.5.** An AVR — artinian valuation ring — is an artinian local principal ideal ring. These are precisely the artinian rings whose ideals are totally ordered by inclusion. Note that a field is an AVR but (by our convention in Notation 1.1) not a DVR.

**Lemma 11.6.** Let \((\Lambda, \mathfrak{m}, k)\) be an unsplit or strictly split Dedekind-like ring with normalization \( \Gamma \), and let \( I \) be an ideal of \( \Gamma \) such that \( 0 \neq I \subseteq \mathfrak{m} \). Then \( \Lambda/I \) is the pullback of the following commutative square analogous to (1.1.1).

\[
\begin{array}{ccc}
\Lambda/I & \subset & \Gamma/I \\
\downarrow \rho & & \downarrow \rho \\
k & \subset & \Gamma \\
\end{array}
\]

Moreover, \( \Gamma/I \) is an AVR or the direct sum of two AVR\s or the direct sum of one DVR and one AVR.

**Proof.** The fact that \( \Lambda/I \) is the pullback of (11.6.1) is an immediate consequence of the fact that \( \Lambda \) is the pullback of (1.1.1). The statements about
Γ/I are immediate consequences of the fact that Γ is either a DVR or the direct sum of two DVRs. □

The next theorem reduces the description of modules over homomorphic images of the Dedekind-like rings in this paper to the situation in diagram (11.6.1).

**Theorem 11.7.** Let I be an ideal of an unsplit or strictly split Dedekind-like ring (Λ, m, k) with normalization Γ, and suppose that 0 ≠ I ⊆ m. Then exactly one of the following holds.

(i) I ≠ ΓI. Then every Λ/I-module is the direct sum of a free Λ/I-module and a Λ/(ΓI)-module, and Λ/(ΓI) is a pullback as in (11.6.1).

(ii) I = ΓI. Then Λ/I is the pullback of commutative square (11.6.1).

**Proof.** (i) By Lemma 11.4, the ring Λ/I is an artinian ring with simple socle (ΓI)/I and (obviously) has no nontrivial idempotents. Therefore Λ/I is quasi-Frobenius [F, Theorem 3.1]. The desired decomposition of Λ-modules therefore follows from Lemma 11.1.

(ii) By hypothesis I is an ideal of Γ, and so Lemma 11.6 applies. □

It now suffices to describe all indecomposable finitely generated Λ/I-modules in the situation of diagram (11.6.1), where 0 ≠ I ⊆ m and I is an ideal of Γ. Our structure theorems in §2 and §3 all construct Λ-modules from Γ-modules, using standard diagrams. The key to moving between Λ and Γ is given in the next lemma, and the final answer is given in Theorem 11.9.

**Lemma 11.8.** Let (Λ, m, k) be an unsplit or strictly split Dedekind-like ring with normalization Γ, and let I be an ideal of Γ such that 0 ≠ I ⊆ m. Let M = M(D) be any finitely generated indecomposable Λ-module constructed from a diagram D and a Γ-module X, as in §2 and §3. Then IM = 0 (i.e., M is a Λ/I-module) if and only if IX = 0 (i.e., X is a Γ/I-module).

**Proof.** We have M = S(D)/K(D) by (2.2.11) or (3.2.9) respectively, according to whether Λ is unsplit or strictly split. Moreover, as shown in Subsection 9.6 and Theorem 6.6, the natural surjection φ: S = S(D) → M is a separated cover of M in both cases; and X = ΓS (in the unsplit case, by (9.1.1) and the fact that A is invertible, in the strictly split case, by (6.4.3) and the fact that A1 and A2 are invertible). It therefore suffices to prove the following more general result:

(11.8.1) Let φ: S → M be a separated cover of a Λ-module, and let X = ΓS be any Γ-module generated by S. Then IM = 0 if and only if IX = 0.

To prove the nontrivial “only if” statement, suppose that IM = 0. Then IS ⊆ ker φ. Since S is a Λ-submodule of some Γ-module, it follows that IS
is a $\Gamma$-submodule of $\ker \phi$. Since $\phi$ is a separated cover, it follows that $IS = 0$ [Lemma 4.9]. Since $I$ is an ideal of $\Gamma$ we therefore have $IX = IGS = IS = 0$ as claimed.

**Theorem 11.9.** Let $(\Lambda, m, k)$ be an unsplit or strictly split Dedekind-like ring with normalization $\Gamma$, and $I$ an ideal of $\Gamma$ such that $0 \neq I \subseteq m$. Let $\mathcal{D}$ be a standard diagram, as in §2 or §3, constructed from some $\Gamma$-module $X$.

Then $M(\mathcal{D})$ is a $\Lambda/I$-module if and only if:

(i) If $\Lambda$ is unsplit, then all length-labels in $\mathcal{D}$ are less than or equal to the length of the AVR $\Gamma/I$ (as a $\Gamma/I$-module).

(ii) If $\Lambda$ is strictly split (so that $\Gamma = \Gamma_1 \oplus \Gamma_2$, and $I = I_1 \oplus I_2$, with each $I_\nu \subseteq \Gamma_\nu$), then all $\Gamma_\nu$-length labels ($\nu = 1, 2$) in $\mathcal{D}$ are less than or equal to the length of the DVR or AVR $\Gamma_\nu/I_\nu$.

**Proof.** As shown in Lemma 11.8, we have $IM = 0$ if and only if $IX = 0$. Thus the question reduces to finding when $IX = 0$.

Suppose first that $\Lambda$ is unsplit, hence $\Gamma$ is a DVR. Then $\Gamma/I$ is an AVR, and hence $X$ is a direct sum of uniserial $\Gamma$-modules whose length is at most the length of $\Gamma/I$. In other words, all length labels in $\mathcal{D}$ are at most the length of $\Gamma/I$, as claimed.

Suppose next that $\Lambda$ is strictly split. Then $\Gamma = \Gamma_1 \oplus \Gamma_2$ where each $\Gamma_\nu$ is a DVR, and $I = I_1 \oplus I_2$ where each $I_\nu$ is an ideal of $\Gamma_\nu$ contained in the maximal ideal of $\Gamma_\nu$. In this situation $\Gamma_\nu/I_\nu$ is either an AVR (if $I_\nu \neq 0$) or a DVR (if $I_\nu = 0$). The rest of the proof is exactly as in the unsplit case.

**Remark 11.10.** It is possible to know that a commutative complete local ring $(\Upsilon, p, k)$ is a homomorphic image of a complete Dedekind-like ring $(\Lambda, m, k)$ without knowing $\Lambda$ and its normalization $\Gamma$ in advance. For example, $\Upsilon$ might be equal $\Omega/n^2$ where $\Omega$ is a Klein ring with radical $n$ [Theorem 11.3].

Assume that $\Upsilon$ is known to be a homomorphic image of a complete Dedekind-like (local) ring. We may assume that the socle of $\Upsilon$ has $k$-dimension 2 (for if it equals 1, we apply the reduction in the supplementary statement of Theorem 11.7). Thus we are now in the situation shown in (11.6.1): $I$ is an ideal of $\Gamma$ contained in $m$. In this situation, one must know something about $\Gamma$ in order to obtain the structure of indecomposable $\Lambda$-modules from Theorem 11.9. A careful reading of the results in Sections 2 and 3 to which Theorem 11.9 refers shows that it suffices to know $\Gamma/I$, rather than $\Gamma$ itself.

We remark that the proofs of [KL1, Propositions 3.4 and 3.6] give constructions of $\Lambda$ and $\Gamma$ in the strictly split and unsplit cases, respectively, and this can presumably be used to find the structure of $\Gamma/I$. 

12. Examples.

This section gives several naturally occurring examples of unsplit and strictly split Dedekind-like rings, both complete and not complete. We repeatedly use the fact that tensoring a pullback diagram with a flat module yields a pullback diagram (use the idea in e.g. [K, Proposition 2.10] or [L3, Lemma 6.1]). We also use Lemma 2.21 of [KL1], which reads: If \((Λ, m, k)\) is an unsplit or strictly split Dedekind-like ring with normalization \(Γ\), then the \(m\)-adic completion \((\hat{Λ}, \hat{m}, k)\) is, respectively, an unsplit or strictly split Dedekind-like ring with normalization \(\hat{Γ}\), and \(\text{rad} \hat{Γ} = \hat{m}\).

Examples 12.1 (Simplest examples). The ring \(\mathbb{R} + X \mathbb{C}[[X]]\) of formal power series over the complex numbers, with real constant term is a complete unsplit Dedekind-like ring [KL1, Examples 2.18].

The ring \(k[[X,Y]]/(X \cdot Y)\), with \(k\) any field, is complete strictly split Dedekind-like, and so is \(\mathbb{Z}_p[[X]]/(p \cdot X)\), where \(p\) is any prime number and \(\mathbb{Z}_p\) is the \(p\)-adic completion of the integers [KL1, Examples 2.17].

Example 12.2. Let \(Λ = ZG\), the integral group ring of any finite cyclic group of squarefree order. Then, for every maximal ideal \(m\) of \(Λ\), the \(m\)-localization \(Λ_m\) is either strictly split Dedekind-like or a DVR. (See [L2], where slightly different terminology is used.) Hence by [KL1, Lemma 2.21], this is also true of the \(m\)-adic completion \(\hat{Λ}_m\).

Example 12.3. (In the language of algebraic geometry:) The coordinate ring of the union of the two lines \(X = 0\) and \(Y = 0\) over the field \(k\) is the prototypical strictly split Dedekind-like ring \(k[[X,Y]]/(XY)\). More generally, let \(Λ\) be the coordinate ring of the union of finitely many distinct lines, where each pair of lines intersect in a point, but no three of the lines intersect in a point. Then each of the localizations of \(Λ\) at one of the maximal ideals corresponding to one of these points of intersection is a strictly split Dedekind-like ring. By [KL1, Lemma 2.21], the same is true of the completions.

Example 12.4 (Quadratic orders). Consider the quadratic order \(Λ = \mathbb{Z}[\sqrt{n}]\) for some square-free integer \(n\). We show that the \(p\)-adic completion \(\hat{Λ}_p\) of \(Λ\) is strictly split Dedekind-like, unsplit Dedekind-like, or a DVR for every rational prime \(p\).

It is well-known ([We, Theorem 6-1-1], for example) that \(Λ\) is a Dedekind domain if and only if \(n\) is congruent to 2 or 3 modulo 4. Thus, we can suppose that \(n\) is congruent to 1 modulo 4. In this situation it is well-known that \(Γ = \mathbb{Z}[(1 + \sqrt{n})/2]\) is the normalization of \(Λ\). (See, for example, [We, Theorem 6-1-1].) If \(p\) is an odd prime, the \(p\)-adic completions of \(Λ\) and \(Γ\) are obviously the same, and hence \(\hat{Λ}_p\) is a DVR, so we need only consider the 2-adic completion \(\hat{Λ}_2\).
Let $C = 2\Gamma$; clearly $C \subseteq \Lambda$, so that $C$ is contained in the conductor of $\Gamma$ into $\Lambda$. By [We, Theorem 6-2-1], the ideal $C$ is maximal in $\Gamma$ if $n$ is congruent to 5 modulo 8, while $C$ is the product of two distinct maximal ideals in $\Gamma$ if $n$ is congruent to 1 modulo 8. We consider these two cases individually.

Consider the case $n \equiv 5 \pmod{8}$. Here, the residue ring $\Gamma/C$ is a field with exactly four elements, and $\Lambda/C$ is a proper subring of $\Gamma/C$, so that $k = \Lambda/C$ is a field with exactly two elements. Therefore, $C$ is also a maximal ideal of $\Lambda$ and so is the conductor from $\Gamma$ into $\Lambda$. Moreover, the rings $\Lambda$ and $\Gamma$ fit into a conductor square as in (1.1.1) (except for the kernels of the vertical maps being the radical), in which $F = \Gamma/C$ is a (separable) quadratic field extension of $k$. Since tensoring a pullback diagram with a flat module again yields a pullback diagram, tensoring this diagram with $\hat{\mathbb{Z}}_2$ shows that $\Lambda_2$ is unsplit Dedekind-like. In fact, the localization $\Lambda_2$ is also unsplit Dedekind-like.

Consider the case $n \equiv 1 \pmod{8}$. Here the residue ring $\Gamma/C$ is the product of two fields each with exactly two elements, and $\Lambda/C$ is a proper subring of $\Gamma/C$. Again it follows that $k = \Lambda/C$ is a field with exactly two elements, so that $C$ is a maximal ideal of $\Lambda$ and is the conductor from $\Gamma$ into $\Lambda$. Moreover, the rings $\Lambda$ and $\Gamma$ fit into a conductor square as in (1.1.1) (except for the kernels of the vertical maps being the radical), in which the inclusion in the bottom row can be taken to be the diagonal inclusion of $k$ into $\Gamma/C \cong k \times k$. As before, tensoring with $\hat{\mathbb{Z}}_2$ shows that $\Lambda_2$ is strictly split Dedekind-like. Here the localization $\Lambda_2$ is never strictly split Dedekind-like, because it is an integral domain. (However it is nonstrictly split Dedekind-like, as defined in [KL1, Definitions 2.5].)

**Example 12.5.** Let $\Lambda$ be a subring of square-free index (say, $n$) in $\Gamma = \mathbb{Z}^s$ (direct product of $s$ copies of the integers). Then we claim that, for every maximal ideal $m$ of $\Lambda$, the localization $\Lambda_m$ is a strictly split Dedekind-like ring or a DVR. (Hence, by [KL1, Lemma 2.21], the same is true of the $m$-adic completion $\Lambda_m$.) In fact, $\Lambda_m$ is strictly split Dedekind-like precisely for the finite number of $m$ that contain a prime factor of $n$.

We may assume that $n > 1$. Since $|\Gamma/\Lambda| = n$ we have

\[(12.5.1) \quad \Gamma n \subset \Lambda \subset \Gamma.\]

Choose a maximal ideal $m$ of $\Lambda$. We want to show that $\Lambda_m$ is strictly split Dedekind-like with normalization $\Gamma_m$, or else is a DVR.

If $m \not\supseteq \Gamma n$, then some element of $\Gamma n$ becomes a unit in $\Lambda_m$, and we have $\Lambda_m = \Gamma_m$, a local principal ideal domain, equivalently, a DVR. Thus we suppose, from now on, that $m \supseteq \Gamma n$. There are only finitely many such $m$ since $\Gamma/\Gamma n$ is a finite ring.

Since $n \in m$, the prime ideal $m$ contains some prime number $p$ that divides $n$. Since localizing at $m$ can be done by first localizing at $p$, we may replace
the three terms in (12.5.1) by their \( p \)-localizations. The effect of this is that, after a change of notation, we may assume that \( n = p \), a prime number. Moreover the original additive group \( \Gamma/\Lambda \) of square-free order \( n \) has now been replaced by its \( p \)-localization, hence is a group of order \( p \). We now have (after our change of notation) that \( \Lambda \) is the pullback of the following diagram.

\[
\begin{array}{ccc}
\Lambda & \subset & \Gamma \\
\downarrow & & \downarrow \\
\overline{\Lambda} = \Lambda / (\Gamma p) & \subset & \overline{\Gamma} = \Gamma / (\Gamma p).
\end{array}
\]

Note that \( \overline{\Lambda} \) and \( \overline{\Gamma} \) are finite rings. Now localize the whole diagram at \( m \). Then \( \Lambda_m \) remains the pullback of the localized diagram. Thus, after a further change of notation, we have that \( \Lambda \) is a local ring. We now prove that \( \Lambda \) is strictly split Dedekind-like with normalization \( \Gamma \).

Think of the original rings \( \Lambda \) and \( \Gamma \) as module-finite \( \mathbb{Z} \)-algebras. Then our first localization — at \( p \) — converted \( \Lambda \) and \( \Gamma \) to module-finite \( \mathbb{Z}_p \)-algebras. In particular, it converted \( \mathbb{Z}^s \) (the original \( \Gamma \)) to \( \mathbb{Z}_p^s \). The effect of the further localization at \( m \) on each coordinate ring \( \mathbb{Z}_p \) is to either leave it unchanged or replace it with \( 0 \). Thus we now have \( \Gamma = \mathbb{Z}_p^t \) for some \( t \leq s \). Therefore \( \overline{\Gamma} = \mathbb{Z}_p^t \), the direct product of \( t \) copies of the field of \( p \) elements.

Now consider the finite, therefore artinian ring \( \overline{\Lambda} \). Since \( \overline{\Lambda} \subset \overline{\Gamma} \), \( \overline{\Lambda} \) has radical zero and is therefore a direct product of fields. On the other hand \( \overline{\Lambda} \) is a local ring, because the local ring \( \Lambda \) maps onto it. Therefore \( \overline{\Lambda} \) is a field (of characteristic \( p \)). Since each of the coordinate rings \( \mathbb{Z}_p \) of \( \overline{\Gamma} \) is a \( \overline{\Lambda} \)-vector space, we see that \( \overline{\Lambda} = \mathbb{Z}_p \).

The abelian group \( \overline{\Gamma}/\overline{\Lambda} \cong \Gamma/\Lambda \) has order \( p \). On the other hand its order is \( |\overline{\Gamma}|/|\overline{\Lambda}| = p^t/p = p^{t-1} \). Therefore \( t = 2 \). We now have that \( \Gamma = \mathbb{Z}_p^2 \), \( \overline{\Gamma} = \mathbb{Z}_p^2 \), and \( \overline{\Lambda} = \mathbb{Z}_p^2 \). Therefore the local ring \( (\Lambda, \Gamma p, \mathbb{Z}_p) \) is strictly split Dedekind-like, as desired.

**Example 12.6.** It is easy to see, by continuing the reasoning in the previous proof, that the rings \( \Lambda \) in Example 12.5 can be explicitly constructed as follows.

Choose any square-free positive integer \( n = p_1 p_2 \ldots p_u > 1 \) and any positive integer \( s > 1 \). For each prime \( p_\nu \) choose a pair of coordinate rings, say coordinates \( i_\nu \neq j_\nu \) of \( \mathbb{Z}^s \). Let \( f_\nu, g_\nu \) be the natural map of coordinate rings \( i_\nu, j_\nu \) respectively, onto \( \mathbb{Z}_{p_\nu} \). Then let

\[
\Lambda = \left\{ (x_1, \ldots, x_s) \in \mathbb{Z}^s \mid (\forall \nu) \ f_\nu(x_{i_\nu}) = g_\nu(x_{j_\nu}) \right\}.
\]

For some amusing module-theoretic parlor tricks over these rings — that is, many examples of nonuniqueness of direct-sum decompositions of finitely generated \( \Lambda \)-modules — see [L4].
### 13. Terminological index.

<table>
<thead>
<tr>
<th>Item</th>
<th>Section</th>
<th>Item</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>associated</td>
<td>481</td>
<td>diagram (split case)</td>
<td>3.2</td>
</tr>
<tr>
<td>$4$-tuple $f(p)$ of matrix pair $p$</td>
<td>8.1</td>
<td>diagram (unsplit case)</td>
<td>2.2</td>
</tr>
<tr>
<td>$\Gamma$-module (split case)</td>
<td>6.3</td>
<td>matrix (split case)</td>
<td>6.4</td>
</tr>
<tr>
<td>$\Gamma$-module (unsplit case)</td>
<td>5.6</td>
<td>matrix (unsplit case)</td>
<td>5.7</td>
</tr>
<tr>
<td>$\Lambda$-module (split case)</td>
<td>6.4</td>
<td>local ring</td>
<td>1.1</td>
</tr>
<tr>
<td>$\Lambda$-module (unsplit case)</td>
<td>5.8</td>
<td>matrix</td>
<td></td>
</tr>
<tr>
<td>AVR (artinian valuation ring)</td>
<td>11.5</td>
<td>$4$-tuple</td>
<td>6.4</td>
</tr>
<tr>
<td>block size, in diagram (split case)</td>
<td>3.2</td>
<td>canonical forms (split case)</td>
<td>7.2</td>
</tr>
<tr>
<td>diagram (unsplit case)</td>
<td>2.2</td>
<td>canonical forms (unsplit case)</td>
<td>8.7, 8.10</td>
</tr>
<tr>
<td>matrix $4$-tuple</td>
<td>(7.2.3)</td>
<td>operations theorem (split case)</td>
<td>6.7</td>
</tr>
<tr>
<td>matrix pair</td>
<td>8.10, 9.2</td>
<td>operations theorem (unsplite case)</td>
<td>5.11</td>
</tr>
<tr>
<td>blocking matrix</td>
<td></td>
<td>pair</td>
<td>5.7</td>
</tr>
<tr>
<td>split case (block cycle $4$-tuple)</td>
<td>(7.2.3)</td>
<td>setup (split case)</td>
<td>6.3</td>
</tr>
<tr>
<td>split case (module)</td>
<td>3.3</td>
<td>setup (unsplite case)</td>
<td>5.6</td>
</tr>
<tr>
<td>unsplit case (matrix pair)</td>
<td>8.10</td>
<td>$A^r$ triangular form</td>
<td>5.12</td>
</tr>
<tr>
<td>unsplit case (module)</td>
<td>2.6</td>
<td>mirror image (permutation)</td>
<td>2.3, 8.6</td>
</tr>
<tr>
<td>concatenation (of sequences)</td>
<td>8.6</td>
<td>moving $U$</td>
<td></td>
</tr>
<tr>
<td>conjugate</td>
<td></td>
<td>split case</td>
<td>3.6</td>
</tr>
<tr>
<td>element, matrix, or $4$-tuple</td>
<td>8.1</td>
<td>unsplit case</td>
<td>2.9</td>
</tr>
<tr>
<td>self-conjugate $4$-tuple</td>
<td>8.1</td>
<td>package (matrix $4$-tuple)</td>
<td>8.1</td>
</tr>
<tr>
<td>corresponding</td>
<td></td>
<td>reduce (top or bottom)</td>
<td>2.2</td>
</tr>
<tr>
<td>column (split case)</td>
<td>6.4</td>
<td>repetition-free sequence (2.6.1)</td>
<td></td>
</tr>
<tr>
<td>column (unsplite case)</td>
<td>5.7</td>
<td>separated cover</td>
<td></td>
</tr>
<tr>
<td>summand (split case)</td>
<td>6.3</td>
<td>almost functorial</td>
<td>4.12</td>
</tr>
<tr>
<td>summand (unsplit case)</td>
<td>5.6</td>
<td>definition</td>
<td>4.6</td>
</tr>
<tr>
<td>cycle</td>
<td>8.6</td>
<td>earlier work</td>
<td>4.8(i)</td>
</tr>
<tr>
<td>Dedekind-like ring</td>
<td>1.1</td>
<td>uniqueness</td>
<td>4.13</td>
</tr>
<tr>
<td>direct sum of matrix $4$-tuples</td>
<td>7.1</td>
<td>connotation of words</td>
<td>4.17</td>
</tr>
<tr>
<td>display operation</td>
<td></td>
<td>definition</td>
<td>4.3</td>
</tr>
<tr>
<td>(split case)</td>
<td>6.5</td>
<td>simultaneous cyclic permutations</td>
<td>4.16</td>
</tr>
<tr>
<td>(unsplite case)</td>
<td>5.9</td>
<td>standard</td>
<td></td>
</tr>
<tr>
<td>DVR</td>
<td>1.1</td>
<td>copy of $F$</td>
<td>2.1, 5.6</td>
</tr>
<tr>
<td>efficient diagram</td>
<td></td>
<td>copy of $k$</td>
<td>3.1, 6.3</td>
</tr>
<tr>
<td>split case</td>
<td>3.7</td>
<td>diagram (split case)</td>
<td>3.3</td>
</tr>
<tr>
<td>unsplit case</td>
<td>2.10</td>
<td>diagram (unsplite case)</td>
<td>2.4, 2.6</td>
</tr>
<tr>
<td>glue</td>
<td></td>
<td>generator</td>
<td>1.1</td>
</tr>
<tr>
<td>split case</td>
<td>3.2</td>
<td>image (of element of $F$)</td>
<td>5.6</td>
</tr>
<tr>
<td>unsplit case</td>
<td>2.2</td>
<td>image (of element of $k$)</td>
<td>6.3</td>
</tr>
<tr>
<td>Klein ring</td>
<td>11.2</td>
<td>sweeping-similarity operations</td>
<td></td>
</tr>
<tr>
<td>Krull-Schmidt theorem</td>
<td>(Dedekind-like rings)</td>
<td>1.3</td>
<td>$k$-$k$ (split case)</td>
</tr>
<tr>
<td>(Dedekind-like rings)</td>
<td></td>
<td>$k$-$F$ (unsplite case)</td>
<td>5.11</td>
</tr>
<tr>
<td>label sequence, in</td>
<td></td>
<td>tame</td>
<td>1 (¶1)</td>
</tr>
<tr>
<td>diagram (split case)</td>
<td>(3.2.10)</td>
<td>Theorem 90 for matrices</td>
<td>8.11</td>
</tr>
<tr>
<td>diagram (unsplite case)</td>
<td>(2.3.1)</td>
<td>tile</td>
<td></td>
</tr>
<tr>
<td>matrix $4$-tuple (deleted cycle)</td>
<td>(7.2.1)</td>
<td>basic</td>
<td>(8.4.1)</td>
</tr>
<tr>
<td>matrix $4$-tuple (block cycle)</td>
<td>(7.2.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>matrix pair (no blocking matrix)</td>
<td>8.7</td>
<td>exceptional</td>
<td>8.10</td>
</tr>
<tr>
<td>matrix pair (with blocking matrix)</td>
<td>8.10</td>
<td>unit forward rotation</td>
<td>2.1, 8.6</td>
</tr>
<tr>
<td>length label, in</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Received September 9, 1999 and revised November 29, 2000. Levy’s research was partially supported by an NSA grant.
Mathematics Department
Florida Atlantic University
Boca Raton, Florida 33431-0991
E-mail address: klingler@fau.edu

Mathematics Department
University of Wisconsin
Madison, Wisconsin 53706-1388
E-mail address: levy@math.wisc.edu
APPLICATION OF REPRESENTATION FORMULAE TO COMPARISON AND NONEXISTENCE THEOREMS FOR ELLIPTIC BOUNDARY VALUE PROBLEMS

A.A. Kosmode’m’anskii Jr.

Using the representation formulas obtained earlier, new comparison theorems for elliptic boundary value problems are developed. Properties of support function of convex domain are applied for proofs and for obtaining nonexistence theorems for solutions of capillary problems in the absence of gravity.

Let $D_0$ and $D_1$ ($D_0 \subset D_1$) be plane convex figures. Denote by $A_i$ and $p_i$ the area and the perimeter of figure $D_i$ ($i = 0, 1$). Let’s suppose that the inequality

$$\frac{A_1}{p_1} > \frac{A_0}{p_0}$$

holds. In the present paper, using (1), we shall obtain the comparison theorems for some elliptic boundary value problems (Sections 1-3). Proofs of these theorems were based on representation formulas, obtained earlier. The general steps of proving one of these are given in the Appendix. Further (Section 4) we will formulate the sufficient condition for (1) in terms of mixed area and will use these results to prove some nonexistence theorems for solutions of capillary problems in the absence of gravity (Section 5).

1. The comparison theorem for solutions of second boundary value problem for Helmholtz equation.

Let $D$ be a convex planar domain with $C^{2,\alpha}$ boundary $\Gamma$. Hereinafter we denote by $n$ the outward normal to $\Gamma$. Let $u(x, y)$ be a solution to the following problem

$$\Delta u = ku \quad \text{in } D, \quad u_n|_{\Gamma} = R > 0.$$ 

In [7] we have proved the following:

**Theorem 1.** Let $u$ be the solution to problem (2), and $z$ be a solution of second boundary value problem for the Poisson equation

$$\Delta z = \frac{R p}{A} \quad \text{in } D, \quad z_n|_{\Gamma} = R,$$
such that
\[ \int \int_D z \, dx \, dy = 0. \]

Then the solution of problem (2) can be represented as
\[ u = \frac{R_p}{kA} + z + \omega, \]
where \( \omega \) satisfies the inequality \( \max |\omega| < C|k| \) in \( D \).

Let \( u_i \) be the solution to (2) in the domain \( D_i \). From (4) we immediately obtain:

**Theorem 2.** Let domains \( D_0 \) and \( D_1 \) be such that the conditions of Theorem 1 and inequality (1) hold. Then there exists a number \( k_0 > 0 \) such that for any positive number \( k < k_0 \) the inequality \( u_0 > u_1 \) holds in the domain \( D_0 \).

The result of Theorem 2 we had announced in [6].

2. The comparison theorem for solutions of the third boundary value problem for Poisson equation.

In the third boundary value problem it is required to find a solution of the equation
\[ \Delta u = -1 \]
in domain \( D \) with the boundary conditions
\[ u + \beta \frac{\partial u}{\partial n} = 0 \quad (\beta > 0). \]

The solution of this problem satisfies the theorem of representation ([5], [8]).

**Theorem 3.** Let the boundary \( \Gamma \) of plane convex domain \( D \) belongs to the class \( C^{2,\alpha} \) and its curvature is separated from zero. Then
\[ u = \frac{\beta A}{p} + u_\infty + \omega, \]
where \( u_\infty \) is the solution of Equation (5) such that
\[ \frac{\partial u_\infty}{\partial n} = -\frac{A}{p}, \quad \int_\Gamma u_\infty \, ds = 0, \]
and function \( \omega \) satisfies the inequality \( \max |\omega| < C\beta^{-1} \) in \( D \).

Let \( u_i \) be the solutions of the third boundary value problems in domains \( D_i \). From representation (7) we immediately obtain:
Theorem 4. Let the domains $D_0$ and $D_1$ be such that the conditions of Theorem 3 and inequality (1) hold. Then there exists $\beta_0 > 0$ such that for any $\beta > \beta_0$ the inequality $u_1 > u_0$ holds in $D_0$.

3. The comparison theorem of capillary surfaces heights in case of small gravity.

It is well-known (see [3]) that the searching of the form of liquid free surface in cylindrical tube under capillary forces and force of gravity is equivalent to the following boundary value problem. It is required to find the solution of the equation

$$(8) \quad \text{div}Tu = ku$$

in domain $D$ with boundary condition

$$(9) \quad (Tu, n) = \cos \gamma, \quad (Tu = \nabla u / \sqrt{1 + |\nabla u|^2}).$$

In the absence of gravity, the equation of liquid free surface takes the form

$$(10) \quad \text{div}Tu = \frac{p \cos \gamma}{A},$$

but boundary condition remains in form (9).

Below we consider M. Miranda question ([3], Sec. 5.3, 5.4): Does a liquid in a “wide” capillary tube rise lower than in a “narrow” one. This question is equivalent to the following problem: Let $u_0$ and $u_1$ be solutions of Equation (8) in domains $D_0$ and $D_1$ ($D_0 \subset D_1$) with boundary conditions (9) on boundaries $\Gamma_0$ and $\Gamma_1$. Is it right that $u_0 > u_1$ in $D_0$?

In [3] some conditions for an affirmative answer are given, and also an example for which the answer is negative.

D. Siegel has proved in [13] for plane domain with $C^{2,\alpha}$ boundary the following:

Theorem 5. Let there exists a solution $z$ to the problem (10)-(9). Then solution $u$ of the problem (8)-(9) can be represented as

$$(11) \quad u = \frac{p \cos \gamma}{kA} + z + \omega$$

while the function $\omega$ satisfies the inequality $\max |\omega| < C|k|$ in $D$.

$L_2$-estimate of $\omega$ was received in [7].

Now the comparison theorem is immediately following from representation (11).

Theorem 6. Let $0 < \gamma < \pi/2$, the inequality (1) holds and there exist the solutions of the problem (10)-(9) in domains $D_0$ and $D_1$, then exists $k_0 > 0$ such that for any $0 < k < k_0$ the inequality $u_0 > u_1$ holds in $D_0$. 
We note that for special cases of domains $D_i$ ($D_1$ is a disk or $D_0$ is a disk of sufficiently small radius), the comparison Theorems 2, 4 and 6 have been obtained by other methods for arbitrary positive $k$ and $\beta$ in [3], [12], [5].

On the other hand, it is evident that if the inequality reverse (1) holds, then $u_0 < u_1$ in $D_0$.


Let us obtain now a sufficient condition under which the inequality (1) holds. We have proved the same implication in [6], where we assume sufficient smoothness of a boundaries. In present paper this result is reduced in Example 3.

Let $A_{01}$ be the mixed area of figures $D_0$ and $D_1$.

**Theorem 7.** Let figures $D_0$ and $D_1$ be such that

\[(p_0 + p_1)A_1 \geq 2A_{01}p_1.\]  

Then the inequality (1) holds.

**Proof.** We shall use the formulas from standard manuals ([1], [11]) on the geometry of convex figures.

Let $D_\theta = (1 - \theta)D_0 + \theta D_1$ be the linear family of convex figures. It is well-known that the area $A_\theta$ of the figure $D_\theta$ is given by formula

\[A_\theta = (1 - \theta)^2A_0 + 2\theta(1 - \theta)A_{01} + \theta^2A_1,\]

and its perimeter $p_\theta$ is given by formula

\[p_\theta = (1 - \theta)p_0 + \theta p_1.\]

We note that inequality (1) immediately follows from (12) and the Frobenius inequality

\[2A_{01} \geq \frac{A_0p_1}{p_0} + \frac{A_1p_0}{p_1}.\]

We shall give another proof whose details give additional information. Let us consider the function

\[f(\theta) = \frac{A_\theta}{p_\theta}.\]

We shall prove that this function is concave in the segment $[0, 1]$ and its left derivative $f'(1)$ is positive because of (12). Hence we shall prove that the function $f(\theta)$ monotonically increases. Using formulas (13) and (14), we obtain

\[f(\theta) = \frac{(1 - \theta)^2A_0 + 2\theta(1 - \theta)A_{01} + \theta^2A_1}{(1 - \theta)p_0 + \theta p_1}.\]

On the other hand

\[(1 - \theta)f(0) + \theta f(1) = (1 - \theta)\frac{A_0}{p_0} + \theta \frac{A_1}{p_1}.\]
After elementary algebraic transformations we see that the concavity condition for \( f(\theta) \)

\[
f(\theta) \geq (1 - \theta)f(0) + \theta f(1)
\]
is equivalent to Frobenius inequality. If we calculate the left derivative

\[
f'(1) = \lim_{\epsilon \to 0} \frac{f(1) - f(1 - \epsilon)}{\epsilon},
\]
using formulas (13) and (14), we obtain

\[
(15) \quad f(1) - f(1 - \epsilon) = \frac{F_1}{p_1} - \frac{\epsilon^2 F_0 + 2\epsilon(1 - \epsilon)F_0 + (1 - \epsilon)^2 F_1}{\epsilon p_0 + (1 - \epsilon)p_1}
\]

\[
= \epsilon \left( \frac{p_0 + p_1}{p_0} \right) A_1 - \frac{2A_0 p_1}{p_1^2} + O(\epsilon^2).
\]

It is evident that the derivative \( f'(1) \) (the coefficient of \( \epsilon \) in (15)) is nonnegative because of (12).

Theorem 7 has been proved.

If coefficient of \( \epsilon \) in (15) is negative then inequality opposite (12) holds. This means that between figures \( D_\theta \) there exists the figure such that \( f(\theta) > f(1) \).

On the other hand we shall obtain the condition for inequality opposite (1) if we calculate the right derivative of the function \( f(\theta) \) in zero.

**Theorem 8.** Let

\[
2A_{01} \leq \frac{p_0 + p_1}{p_0} A_0.
\]

Then

\[
\frac{A_1}{p_1} < \frac{A_0}{p_0}.
\]

Let us consider three important special cases.

**Example 1.** Let \( D_1 \) be a disk with radius \( R_1 \). In this case we have \( 2A_{01} = R_1 p_0 \). It is evident that \( p_0 < 2\pi R_1 \), hence

\[
R_1 p_0 2\pi R_1 < (p_0 + 2\pi R_1) \pi R_1^2,
\]
and inequality (12) holds.

**Example 2.** Let \( D_0 \) be a disk with radius

\[
r < \frac{A_1}{p_1} \frac{1}{1 - \frac{2\pi A_1}{p_1^2}}.
\]

We know that \( 2A_{01} = rp_1 \). After algebraic transformations we obtain

\[
r(p_1^2 - 2\pi A_1) < A_1 p_1,
\]
or

\[
r p_1 p_1 < (2\pi r + p_1) A_1,
\]
hence inequality (12) holds.

We note that from isoperimetric inequality

$$\frac{A_1}{p_1} \frac{1}{1 - \frac{2\pi A_1}{p_1^2}} \leq \frac{2A_1}{p_1}$$

(equality holds only if $D_1$ is a disk), hence, using Example 1, we can to improve the previous result: Let $D_0 \subset D_1$ and $D_0$ is contained in the disc of radius $R_0 < 2A_1/p_1$ then inequality (1) holds.

**Example 3.** Let domain $D_1$ has smooth boundary, whose curvature $K_1$ satisfies the inequality

$$0 < K_1 \leq \frac{p_1}{A_1}.$$  \hspace{1cm} (16)

Then inequality (12) holds.

Indeed, let $h_i(\phi)$ be the support function of the domain $D_i$. Then the following formulas are valid ([1])

$$2A_{01} = \int_0^{2\pi} (h_0h_1 - h_0'h_1')d\phi,$$

$$\frac{1}{K_1} = h_1'' + h_1,$$

$$p = \int_0^{2\pi} h d\phi, \hspace{1cm} 2A = \int_0^{2\pi} (h^2 - h''^2) d\phi.$$

Using the inequality (16) we obtain

$$2A_{01} = \int_0^{2\pi} h_0(h_1'' + h_1)d\phi$$

$$= \int_0^{2\pi} (h_0 - h_1)(h_1'' + h_1)d\phi + \int_0^{2\pi} (h_1^2 - h_1'^2)d\phi$$

$$\leq 2A_1 - \frac{A_1}{p_1}(p_1 - p_0)$$

$$= \frac{A_1(p_1 + p_0)}{p_1}.$$

Hence the inequality (12) holds.

Examples stated above show that condition (12) can be used for checking inequality (1).
5. The nonexistence theorems for solutions of capillary problem in the absence of gravity.

Let us return to the problem (10)-(9). If $\gamma = 0$ the important condition of the existence of solution for this problem in the domain $D_1$ is the following: Let $D_0$ be an arbitrary subdomain of $D_1$ then if the solution of the problem (9)-(10) exists then inequality (1) holds. ([2]).

Giusti has proved ([4]) that for convex domains the sufficient condition of the existence is the inequality (16). Moreover ([3]), if the solution of the problem (10)-(9) exists for $\gamma = 0$ then it exists for any $0 < \gamma \leq \pi/2$. Using this statements and our previous speculations we can reformulate the result [2] as sufficient condition of nonexistence for problem (10)-(9).

**Theorem 9.** Let exists such convex subdomain $D_0$ of domain $D_1$ that the inequality opposite (12) holds. Then if $\gamma = 0$ then a solution of (10)-(9) does not exist.

**Proof.** Really, if inequality opposite (12) holds then there exists a domain $D \subset D_1$ in linear family $D_\theta = (1 - \theta)D_0 + \theta D_1$ such that

$$\frac{A}{p} > \frac{A_1}{p_1}.$$  

In particular, (10)-(9) has no solutions, if $D_1$ is a regular polygon. Indeed, we can put as $D_0$ the disk inscribed into $D_1$.

Using results of Section 4, we can add the following simple condition of nonexistence of solutions for (10)-(9) in case of $\gamma = 0$.

**Theorem 10.** Let we can inscribe into $D_1$ the disk of radius

$$r > \frac{p_1 A_1}{p_1^2 - 2\pi A_1}.$$  

Then in case of $\gamma = 0$ the solution of the problem (10)-(9) does not exist.

**Proof.** We immediately obtain from (17)

$$rp_1 > \frac{2\pi r + p_1}{p_1} A_1.$$  

We can take a disk of radius $r$ as the domain $D_0$. It is evident that the inequality (18) is the inequality opposite (12), hence we can apply Theorem 9.

Let us obtain now the generalization of Theorem 10.

**Theorem 11.** Let we can inscribe in domain $D_1$ the disk of radius $r$ such that inequality (17) holds. Then problem (10)-(9) has no solutions for any contact angle $\gamma$ satisfying the inequality

$$\cos \gamma > \frac{A_1}{rp_1} \left( 1 + \sqrt{\frac{4\pi(p_1 r - \pi r^2 - A_1)}{p_1^2 - 4\pi A_1}} \right).$$
Proof. We remind the general idea of nonexistence proofs: If we can find subdomain \( \mathcal{D} \subset D_1 \) such that

\[
\frac{A_1}{p_1} < \frac{\overline{A} \cos \gamma}{\overline{p}},
\]

then problem (10)-(9) has no solution ([3]).

We shall find the subdomain \( \mathcal{D} \) in a certain linear family \( D_\theta \). We can reformulate the nonexistence condition in the following form: Let there exist a subdomain \( D_0 \subset D_1 \) and number \( \theta \in (0, 1) \) such that

\[
f(1) < f(\theta),
\]

then the problem (10)-(9) has no solutions.

Let us construct the corresponding linear family. Let the convex domain \( D_0 \subset D_1 \) be such that

\[
2A_0 p_0 > (p_0 + p_1)A_0, \quad 2A_0 p_1 > (p_0 + p_1)A_1.
\]

It follows from Theorems 7 and 8, that function \( f(\theta) \) reaches its maximum value in the interval \((0, 1)\). Let us find this value. We represent the function \( f(\theta) \) in the form

\[
f(\theta) = -\frac{E\theta}{\Delta p} + \frac{S}{(\Delta p)^2} - \frac{G}{(\Delta p)^2(\theta \Delta p + p_0)},
\]

where

\[
E = 2A_0 - A_0 - A_1, \quad S = 2p_1(A_0 - A_0) + p_0(A_0 - A_1),
\]

\[
G = 2p_0 p_1 A_0 - p_1^2 A_0 - p_0^2 A_1, \quad \Delta p = p_1 - p_0.
\]

It follows from the Frobenius inequality that \( G > 0 \), and inequalities (22) shows that \( E > 0 \).

After calculations we see that function \( f(\theta) \) reaches its maximum value in the point

\[
\theta = \frac{1}{\Delta p} \left( \frac{\sqrt{G}}{E} - p_0 \right),
\]

and this value is equal to

\[
f(\theta) = \frac{2}{(\Delta p)^2} \left( (p_0 + p_1)A_0 - p_1 A_0 - p_0 A_1 - \sqrt{G E} \right).
\]

Using inequality (21), we can reformulate the sufficient condition of nonexistence of solution to the problem (10)-(9): Let there exist a convex subdomain \( D_0 \) of the domain \( D_1 \) such that inequalities (22) holds, then for any contact angle \( \gamma \), satisfying inequality

\[
\cos \gamma > \frac{A_1(\Delta p)^2}{2p_1((p_0 + p_1)A_0 - p_1 A_0 - p_0 A_1 - \sqrt{G E})},
\]

problem (10)-(9) has no solutions.
Let $D_0$ be the disk, which radius $r$ satisfies the inequality (17). Then

$$A_0 = \pi r^2, \quad p_0 = 2\pi r, \quad 2A_{01} = rp_1,$$

$$G = \pi r^2(p_1^2 - 4\pi A_1), \quad E = rp_1 - \pi r^2 - A_1.$$  

Note that the second of inequalities (22) holds because of (17) and the first one holds automatically. Substituting last formulas in inequality (24) we obtain Theorem 11 after algebraic transformations.

**Example.** Let domain $D_1$ be the regular $n$-polygon circumscribed around a circle of radius $r$. Then the inequality (19) takes the form

$$\cos \gamma > \frac{1}{2} \left(1 + \sqrt{\frac{\pi}{n} \tan \frac{\pi}{n}}\right).$$

For large $n$ we can write more simple formula

$$\gamma < \frac{\pi}{n \sqrt{6}}.$$

We see that this estimate is weaker than the exact one ([3], Th. 6.2): $\gamma < \pi/n$, but it holds the same form in case of smoothed angles.

**Appendix.**

Let us consider now the general steps for proof Theorem 1. Hereafter we denote by $C$ (with subscripts or without them) the constants depending on geometrical characteristics of domain $D$.

It is evident that in convex domain $D$ Poincaré inequality holds

$$\int_D \int_D u^2 \, dx \, dy \leq \frac{1}{A} \left( \int_D \int_D u \, dx \, dy \right)^2 + \mu \int_D \int_D |\nabla u|^2 \, dx \, dy.$$

Let

$$v = u - \frac{Rp}{kA}.$$

Function $v$ satisfies the equation

$$\Delta v = kv + \frac{Rp}{A}$$

and $v_n = R$ on $\Gamma$. Let us integrate (2) over $D$. It is easy to see that

$$\int_D \int_D v \, dx \, dy = 0.$$

We subtract Equation (3) from Equation (25). We obtain

$$\Delta (v - z) = k(v - z) + kz.$$

Denote $\omega = v - z$. Then

$$\Delta \omega = k\omega + kz.$$
We multiply (26) on $\omega$ and integrate over $D$. We obtain

$$
\int_D \Delta \omega dx \, dy = \int_D k\omega^2 dx \, dy + \int_D z \omega dx \, dy.
$$

We transform the left side of (27) by well-known formulas, taking into account that $\omega_n = 0$. We obtain

$$
- \int_D |\nabla \omega|^2 dx \, dy = \int_D k\omega^2 dx \, dy + \int_D k z \omega dx \, dy.
$$

Taking into account that

$$
\int_D \omega dx \, dy = 0
$$

we use the Poincaré inequality

$$
\left( k + \frac{1}{\mu} \right) \|\omega\|_{L^2}^2 \leq k \|\omega\|_{L^2}^2 + \int_D |\nabla \omega|^2 dx \, dy
$$

$$
\leq k \int_D \omega^2 dx \, dy + \int_D |\nabla \omega|^2 dx \, dy,
$$

and the Cauchy-Schwarz-Bunyakovskii inequality

$$
-k \int_D \omega z dx \, dy \leq |k| \left( \int_D z^2 dx \, dy \right)^{1/2} \left( \int_D \omega^2 dx \, dy \right)^{1/2}.
$$

We obtain after algebraic transformations

$$
\|\omega\|_{L^2} \leq \frac{\mu |k|}{1 + \mu k} \|z\|_{L^2}.
$$

By S.L. Sobolev embedding theorem:

$$
\left( \max_D \omega \right)^2 \leq C_1 \|\nabla_2 \omega\|_{L^2}^2 + C_2 \|\omega\|_{L^2}^2.
$$

$L_2$ - norm of second derivitives of $\omega$ in plane convex domain is estimated from $L_2$ - norm of operator $\Delta \omega$. The detailed proof of this estimate for solution of the first boundary value problem is given in [10]. The same proof yields the same estimate for solutions of second boundary value problem as well. Indeed, let $g = \omega_{xx} \omega_{yy} - \omega_{xy}^2$. It is evident that

$$
\omega_{xx}^2 + 2\omega_{xy}^2 + \omega_{yy}^2 = (\Delta \omega)^2 - 2g.
$$
Using the identity
\[
2 \int_D \int \, g \, dx \, dy = \int_\Gamma (\omega_x \omega_{xy} - \omega_y \omega_{xx}) \, dx - (\omega_x \omega_{yy} - \omega_y \omega_{xy}) \, dy,
\]
we obtain because of boundary condition of problem (2) and convexity of domain $D$
\[
2 \int_D \int g \, dx \, dy = \int_\Gamma K |\nabla \omega|^2 \, ds > 0
\]
and the estimate is proved.

Furthermore, we obtain from (26)
\[
\| \nabla^2 \omega \|^2_{L^2} \leq \| \Delta \omega \|^2_{L^2} \leq 2k^2(\| \omega \|^2_{L^2} + \| \zeta \|^2_{L^2}).
\]

The statement of Theorem 1 follows from the substitution of the latter estimate in (30) using (29).

**Remark.** Analyzing the proof of Theorem 1, we can see that requirement to convexity of domain $D$ is excessive. Indeed, we can require only the realizability of the Poincaré inequality, S.L. Sobolev embedding theorem and the possibility to estimate $\| \nabla^2 u \|_{L^2}$ by means of $\| \Delta u \|_{L^2}$. These conditions are contained in [10].

**Acknowledgments.** We would like to thank Professor R. Finn for his comments and suggestions. Professor R. Finn has played a critical role in the design and the completion of this paper. We are grateful to Doctor G.N. Krichevets for helpful discussion.

**References**


Received March 19, 1999 and revised June 19, 2000.

Moscow State University of Railways (MIIT)
Dept. of Computational Mathematics
ul. Obraztsova 15
Moscow
Russia
E-mail address: mitja@farber.mccme.ru
HIGHER-DIMENSIONAL SUBSHIFTS OF FINITE TYPE,
FACTOR MAPS AND MEASURES OF MAXIMAL
ENTROPY

RONALD MEESTER AND JEFFREY E. STEIF

We investigate factor maps of higher-dimensional subshifts
of finite type. In particular, we are interested in how the
number of ergodic measures of maximal entropy behaves un-
der such factor maps. We show that this number is preserved
under almost invertible maps, but not in general under fi-
nite to one factor maps. One of our tools, which is of inde-
pendent interest, is a higher-dimensional characterization of
entropy-preserving factor maps that extends the well-known
one-dimensional characterization result.

1. Introduction.

In this paper we discuss some aspects of higher-dimensional subshifts of fi-
nite type. The book of Lind and Marcus ([4]) is an excellent introduction to
the theory of one-dimensional symbolic dynamics. It turns out however, as
is well-known, that the higher-dimensional theory is different from the one-
dimensional theory. In higher dimensions, new concepts arise that do not
have one-dimensional analogues. In addition, many one-dimensional results
are simply not true in higher dimensions. In this paper we address issues
of both types. An example of a new question without a one-dimensional
analogue is the behaviour of the number of ergodic measures of maximal
entropy under various types of factor maps. For an example of the second
type, we show that the one-dimensional characterization of entropy preserv-
ing factor maps (which says that under an obvious irreducibility condition,
entropy preservation is equivalent to the map being finite to one) must be
replaced by something different in higher dimensions. We also mention here
the recent paper [6] which contains a number of interesting results concern-
ing higher-dimensional subshifts of finite type.

Let us start with a description of the setup. Throughout, $X$ denotes a $d$-
dimensional shift space with finite alphabet $\mathcal{A}(X)$, that is, $X$ is a closed (in
the product topology) translation invariant subset of $\{\mathcal{A}(X)\}^{\mathbb{Z}^d}$. A pattern
in $X$ is the restriction of a configuration in $X$ to a finite subset of $\mathbb{Z}^d$. The
restriction of a configuration $x$ to a finite set $A \subseteq \mathbb{Z}^d$ is denoted by $\pi_A(x)$.
We write $B_n$ for the box $\{-n, \ldots, n\}^d$ and $\pi_n(x)$ for $\pi_{B_n}(x)$. The collection of all patterns $\{\pi_n(x) : x \in X\}$ is denoted by $B_n(X)$.

For a countable set of forbidden patterns $\mathcal{F} = \{F_1, F_2, \ldots\}$, we write $S_\mathcal{F}$ for the shift space consisting of those configurations that do not contain any of the patterns in $\mathcal{F}$. We can and often will assume that for all $i$, $F_i$ is a pattern on a box. The size of a pattern on $B_n$ is defined to be $n$. A shift space $X$ is called a subshift of finite type (SOFT) if $X$ can be written as $X = S_\mathcal{F}$, where $\mathcal{F}$ consists of only finitely many patterns.

Consider two shift spaces $X$ and $Y$. Let, for some number $\alpha$, $\Phi$ be a map $\Phi : B_\alpha(X) \to A(Y)$. We call $\Phi$ a block map. We can use this block map to define a map $\phi : X \to Y$ by putting

$$\phi(x)_z = \Phi(\pi_\alpha(T_z(x)))$$

where $T_z$ denotes translation by the vector $z$. Clearly $\phi$ commutes with shifts. $\phi$ is called a factor map from $X$ to $Y$. (Clearly $\phi$ is continuous in the product topology and it is well-known that all shift commuting continuous maps from $X$ to $Y$ are of this form.) The smallest $\alpha$ for which a given factor map $\phi$ arises in this way is called the range of $\phi$. All factor maps in this paper are assumed to be onto, i.e., $Y = \phi(X)$. An invertible factor map is called a conjugacy, and if there exists a conjugacy between $X$ and $Y$, then $X$ and $Y$ are called conjugate. We will sometimes abuse notation and write $\phi(x)$ where $x \in B_n(X)$ for some $n$, that is, we view $\phi$ as a mapping $\phi : B_\alpha(X) \to B_{n-\alpha}(Y)$.

If $X$ is a shift space, let $h(X)$ denote the topological entropy of $X$ and if $\mu$ is a translation invariant measure on $X$, let $h(\mu)$ denote the measure-theoretical entropy of $\mu$. See [9] for precise definitions. For a finite partition $\mathcal{P}$ of a probability space $(X, \mu)$, we let $h_\mu(\mathcal{P})$ be its entropy defined to be $-\sum_{i} \mu(P_i) \log \mu(P_i)$ where the $P_i$’s denote the atoms of $\mathcal{P}$. The set of translation invariant probability measures on $X$ is denoted by $\mathcal{M}(X)$. The variational principle (see [5]) states that

$$h(X) = \sup_{\mu \in \mathcal{M}(X)} h(\mu).$$

Moreover, the supremum is achieved at some measure: Such a measure is called a measure of maximal entropy. Let $\mathcal{M}_{\text{max}}$ denote the set of measures of maximal entropy. In one dimension, an irreducible SOFT (precise definitions follow later) has a unique measure of maximal entropy, the so-called Parry measure. In higher dimensions this is no longer true. Burton and Steif ([1]) gave examples of strongly irreducible SOFT’s with multiple ergodic measures of maximal entropy. It is well-known that the set of extremal elements in $\mathcal{M}_{\text{max}}$ is exactly the set of ergodic measures of maximal entropy. This implies that $\mathcal{M}_{\text{max}}$ is a face in the simplex $\mathcal{M}$ and hence is also a simplex. Cardinality is denoted by $|\cdot|$ throughout.
This paper is organized as follows: Section 2 contains all the non-standard technical definitions, Section 3 contains our main results, and the last two sections are devoted to the proofs of the results.

2. Further definitions and preliminaries.

In this section we give some technical definitions.

**Definition 2.1.** A shift space $X$ is called **strongly irreducible** (s.i.) if there is an $s \geq 0$ such that whenever we have two (finite) patterns $\eta_1$ on $A_1 \subset \mathbb{Z}^d$ and $\eta_2$ on $A_2 \subset \mathbb{Z}^d$ of $X$ and the distance between $A_1$ and $A_2$ is greater than $s$, then there is an $\eta \in X$ that is an extension of both $\eta_1$ and $\eta_2$.

We call the smallest such $s$ with the above property the **separation distance** of $X$.

**Definition 2.2.** $X$ is called **weakly irreducible** if for every $\eta_1$ and $\eta_2$ as above, and every $z \in \mathbb{Z}^d$, there is a $u \in \mathbb{Z}^d$ on the halfline through $z$ and the origin, and an $\eta$, such that $\eta$ extends both $\eta_1$ and $T_u(\eta_2)$.

**Remark 1.** The definition of weak irreducibility in [1] is not satisfactory: The one-dimensional SOFT with only one disallowed patterns $(1,0)$ is weakly irreducible according to their definition. But clearly this is not desirable, since all elements of this SOFT are of the form $\ldots 0000011111 \ldots$ In our definition this SOFT is not weakly irreducible.

**Remark 2.** If $X$ is s.i. and $\phi : X \to Y$ is a (onto) factor map, then it is easy to see that $Y$ is also s.i.

**Definition 2.3.** A factor map $\phi : X \to Y$ is said to be **finite to one** if every $y \in Y$ has only finitely many pre-images.

For the definition of almost invertibility we need the concept of transitive points. In one dimension, a point $x \in X$ is called transitive if in $x$ we see all patterns of $X$ in both directions. There are several ways to extend this idea to higher dimensions. We choose for maybe the strongest possible definition, in order to make the forthcoming Theorem 3.4 as strong as possible.

**Definition 2.4.** A point $x$ in a shift space $X$ is called **transitive** if for all half-lines $\ell$ starting at the origin with rational direction, and all $n \geq 0$, the set

$$\left\{ \pi_n(T_z(x)) : z \in \mathbb{Z}^d \cap \ell \right\}$$

contains all patterns of $X$ of size $n$.

Observe that this definition is stronger than requiring that the $\mathbb{Z}^d$-orbit of $x$ is dense in $X$, which is one of the usual definitions.
Definition 2.5. A factor map $\phi : X \to Y$ is called almost invertible if every transitive point in $Y$ has only one pre-image.

In one dimension, an almost invertible factor map from an (weakly) irreducible SOFT is necessarily finite to one (Proposition 9.2.2 in [4]). This is no longer true in higher dimensions as the following example shows.

Example. Consider, in one dimension, a factor map $\phi : X \to Y$ that is almost invertible but not invertible and where $X$ is s.i. (Such an example can be given by sending the golden mean shift to the even shift (Examples 1.2.3 and 1.2.4 in [4]) induced by the block map $00 \to 1$, $01 \to 0$, $10 \to 0$.) Next, define a two-dimensional shift space $X'$ by the requirement that $x \in X'$ if for all $k$, $(x(z,k))$, $z \in \mathbb{Z}$ is an element of $X$. $Y'$ is defined similarly. Define a factor map $\phi' : X' \to Y'$ by just applying $\phi$ row by row. It is clear that $\phi'$ is not finite to one. On the other hand, if $y' \in Y'$ is transitive, the restriction of $y'$ to a horizontal line is transitive in $Y$, and therefore has only one pre-image in $X$. It follows that $y'$ has only one pre-image and we conclude that $\phi'$ is almost invertible.

3. Results.

We first discuss two known basic facts concerning conjugacies between SOFT’s. They are not related to our results but we give them for the sake of the reader. Theorem 2.1.10 in [4] shows that conjugacies preserve SOFTness in one dimension. Their proof does not extend to higher dimensions. The following extension of this result is due to Klaus Schmidt and reported in [7].

Theorem 3.1. Let $X$ and $Y$ be two $d$-dimensional shift spaces, and suppose that $X$ is a SOFT and that $X$ and $Y$ are conjugate. Then $Y$ is also a SOFT.

A nearest neighbor SOFT is a SOFT all whose forbidden patterns are between nearest neighbors, that is, all forbidden patterns are on sets of the form $\{x, x + e_i\}$ for $x \in \mathbb{Z}^d$ and where the $e_i$’s denote the unit vectors. The following result is proved in [10].

Theorem 3.2. Every SOFT is conjugate to a nearest neighbor SOFT.

We now consider measures of maximal entropy. In particular, we are interested in what happens with the number of ergodic measures of maximal entropy under factor maps. Without any assumption, there is no hope for an interesting result, as indicated by the following examples.

Example. Let, for positive integers $N$ and $M$, $X_{N,M}$ be the following nearest neighbor SOFT. The alphabet $\mathcal{A}(X_{N,M})$ is the set $\{-N, \ldots, -1, 1, \ldots, M\}$. The forbidden patterns are all neighboring pairs of the form $(a, b)$ with $ab \leq -2$, i.e., a positive number cannot sit next to a negative
number unless they both have absolute value 1. These SOFT’s were studied in [1] and [2]. In [1], it was shown that when \( N = M \) and \( M \) is sufficiently large, then \( X \) has exactly two ergodic measures of maximal entropy which we call \( \mu^+ \) and \( \mu^- \). If we have a configuration chosen according to \( \mu^+ \), flipping all coordinates (i.e., multiplying them by minus one) leads to a choice distributed according to \( \mu^- \). In [2], it was shown that if \( N = M - 1 \) and \( M \) is sufficiently large, then \( X_{N,M} \) has only one measure of maximal entropy. We can now define a factor map \( \phi : X_{M,M+1} \to X_{M,M} \) simply by changing all the \( M+1 \)'s to \( M \). When \( M \) is large enough, our previous comments imply that this is a factor map that maps a space with a unique measure of maximal entropy onto a space with two ergodic measures of maximal entropy. On the other hand, mapping \( X_{M,M} \) onto the full \( M \)-shift by taking absolute values coordinatewise leads, for large \( M \), to a factor map that maps a space with two ergodic measures of maximal entropy onto a space with a unique measure of maximal entropy.

In both of the last two mappings, the image shift space has strictly lower entropy than the domain space and so let us see what happens when entropy is preserved. We shall write \( \phi \mu \) for the push-forward of the measure \( \mu \) under \( \phi \), that is, \( \phi \mu(A) = \mu(\phi^{-1}A) \), for all Borel sets \( A \).

**Theorem 3.3.** Let \( X \) and \( Y \) be shift spaces with \( h(X) = h(Y) \) and let \( \phi : X \to Y \) be a factor map. Then:

(i) Every \( \mu \in \mathcal{M}_{\text{max}}(X) \) is the push-forward of a measure \( \nu \) in \( \mathcal{M}_{\text{max}}(X) \), i.e., \( \mu = \phi \nu \) for some \( \nu \in \mathcal{M}_{\text{max}}(X) \);

(ii) If \( X \) is a strongly irreducible SOFT and \( \mu \in \mathcal{M}(X) \), then \( h(\phi \mu) = h(\mu) \). In particular, if \( \mu \in \mathcal{M}_{\text{max}}(X) \), then \( \phi \mu \in \mathcal{M}_{\text{max}}(Y) \) and so \( \phi \) takes \( \mathcal{M}_{\text{max}}(X) \) to \( \mathcal{M}_{\text{max}}(Y) \) and is surjective by (i).

The question of whether the induced mapping from \( \mathcal{M}_{\text{max}}(X) \) to \( \mathcal{M}_{\text{max}}(Y) \) when \( h(X) = h(Y) \) is injective will arise in Theorems 3.4 and 3.5. We note here the obvious fact that if \( \phi \) is a conjugacy from \( X \) to \( Y \), then the induced mapping \( \phi \) from \( \mathcal{M}_{\text{max}}(X) \) to \( \mathcal{M}_{\text{max}}(Y) \) is an isomorphism of simplices; in particular, the number of ergodic measures of maximal entropy are the same.

**Remark 3.** Note that all statements in Theorem 3.3 are false without the entropy condition. The first example preceding Theorem 3.3 shows that (i) cannot generally be true, since we can map from a space with a unique measure of maximal entropy onto a space with two ergodic measures of maximal entropy. The second example preceding Theorem 3.3 shows that (ii) is false without the entropy condition since it is not hard to see that the image of \( \mu^+ \) is not i.i.d. uniform, the unique measure of maximal entropy on the full \( M \)-shift. (In fact, it is not even i.i.d.)
Next we investigate what happens when the factor map is not quite invertible. We start with almost invertible.

**Theorem 3.4.** Let \( X \) be a strongly irreducible SOFT and let \( \phi : X \rightarrow Y \) be almost invertible. Then \( h(X) = h(Y) \) and the induced mapping \( \phi : \mathcal{M}_{\text{max}}(X) \rightarrow \mathcal{M}_{\text{max}}(Y) \) is bijective. In particular, since this mapping is convex and continuous, it is an isomorphism of these simplices and hence the number of ergodic measures of maximal entropy for the two systems are the same.

Note that we do not assume that \( Y \) is a SOFT.

Another relaxation of invertibility is to require the factor map to be *finite to one*. The following theorem tells us that as far as measures of maximal entropy are concerned, finite to one is very different from almost invertibility.

**Theorem 3.5.** There exists a strongly irreducible SOFT \( X \) and a shift space \( Y \) with the same entropy, together with a two to one factor map \( \phi \) from \( X \) onto \( Y \) such that \( X \) has two ergodic measures of maximal entropy and \( Y \) has a unique measure of maximal entropy.

In our proofs of the above results we shall need a higher-dimensional characterization of entropy preserving factor maps. For one-dimensional irreducible SOFT’s, preservation of entropy under factor maps is characterized by the requirement that the factor map is finite to one (Theorem 8.1.16 in [4]). This is no longer true in higher dimensions. For instance, one can map the full 2-dimensional 2-shift onto itself via the block map \( \Phi \) of range 1 given by

\[
\Phi((x_{ij})_{i,j \in \{-1,0,1\}}) = x_{00} + x_{10} \pmod{2}.
\]

(Note that this map operates row by row.) Clearly this map preserves entropy but it is not finite to one.

**Remark 4.** It follows from Theorem 3.3 that a mapping between spaces of equal entropy can never be from a space with a unique measure of maximal entropy to a space with multiple ergodic measures of maximal entropy. The reader should compare this with Theorem 3.5.

For a factor map \( \phi \) from \( X \) onto \( Y \), a **diamond** of \( \phi \) is a set of two elements \( x \neq x' \) in \( X \) which differ in only finitely many coordinates and for which \( \phi(x) = \phi(x') \). (The name “diamond” comes from the one-dimensional graphical representation of these objects.) The proper analogue of Theorem 8.1.16 in [4] in higher dimensions is given by the following theorem. We remark that the equivalence of (a) and (b) follows from Theorem 3.2 in [6] together with the fact that a strongly irreducible SOFT is automatically entropy minimal; see the forthcoming Lemma 4.1. We will need (in particular) characterization (c) below, and since the proof of the equivalence of (c) with
the other statements depends on the proof of (a) \(\iff\) (b), we include a full proof of the theorem.

**Theorem 3.6.** Let \(X\) be a strongly irreducible SOFT and let \(\phi : X \rightarrow Y\) be a factor map based on a block map \(\Phi : B_\alpha(X) \rightarrow A(Y)\). Then the following six statements are equivalent:

(a) \(h(X) > h(Y)\).

(b) \(\phi\) has a diamond.

(c) There exist \(y \in Y\), a positive constant \(k_1\) and a constant \(c_1 > 1\) such that for infinitely many \(n\) we have

\[
|\{x \in B_n(X) : \phi(x) = \pi_{n-\alpha}(y)\}| \geq k_1 c_1^{2^{n+1}d}.
\]

(d) There exist \(y \in Y\), a positive constant \(k_2\) and a constant \(c_2 > 1\) such that for infinitely many \(n\) we have

\[
|\pi_n(\phi^{-1}(y))| \geq k_2 c_2^{2^{n+1}d}.
\]

(e) There exist \(y \in Y\), a positive constant \(k_3\) and a constant \(c_3 > 1\) such that for all \(n \geq 0\) we have

\[
|\{x \in B_n(X) : \phi(x) = \pi_{n-\alpha}(y)\}| \geq k_3 c_3^{2^{n+1}d}.
\]

(f) There exist \(y \in Y\), a positive constant \(k_4\) and a constant \(c_4 > 1\) such that for all \(n \geq 0\) we have

\[
|\pi_n(\phi^{-1}(y))| \geq k_4 c_4^{2^{n+1}d}.
\]

**Remark 5.** The proof readily shows that the implication from (a) to (b) above is true for any SOFT \(X\).

### 4. Factor maps and entropy: Proofs.

In this section, we prove Theorem 3.6. We shall need the following lemma.

**Lemma 4.1.** Let \(X\) be a strongly irreducible SOFT, and consider a pattern \(p\) that occurs in \(X\). If we add this pattern to the list of forbidden patterns, obtaining a new SOFT \(X'\), then \(h(X') < h(X)\).

**Proof.** Our proof uses measures, which might appear a little strange since the statement is purely topological. Let \(\mu\) be a measure of maximal entropy for \(X'\). If \(h(X) = h(X')\) then it would follow that \(\mu\) is also a measure of maximal entropy for \(X\). Now every measure of maximal entropy has so called uniform conditional probabilities (see [2], Proposition 1.20), that is, the conditional distribution of patterns on a finite set \(A\) given any configuration \(\eta\) on the outside of \(A\) is uniform over all compatible configurations on \(A\) that extend \(\eta\).
Let the size of \( p \) be \( M \). Next let \( N > M + s \) (where \( s \) is the separation distance of \( X \)) and consider an allowed configuration on the external boundary \( \partial B_N \) of \( B_N \). Since \( X \) is s.i., the extra forbidden pattern \( p \), when placed in \( B_M \), is compatible with this boundary condition, and by the property of uniform conditional probabilities, there is positive \( \mu \) probability to see \( p \) in \( B_M \). But this is a contradiction since \( \mu \) concentrates on \( X' \) in which \( p \) does not occur.

**Proof of Theorem 3.6.** (a) \( \Rightarrow \) (b): The factor map \( \phi \) is based on a block map \( \Phi : \mathcal{B}_\alpha(X) \to \mathcal{A}(Y) \) for some \( \alpha \). We choose \( m \) so large that \( m > 2\alpha, 2\ell \) where \( \ell \) is the size of the largest forbidden pattern in \( X \). Observe that \( \phi \) maps patterns of size \( n + m \) onto patterns of size \( n + m - \alpha \). From the definition of entropy we have that

\[
|\mathcal{B}_{n+m}(X)| = e^{(2n+1)d(h(X)+o(1))},
\]

and

\[
|\mathcal{B}_{n+m-\alpha}(Y)| = e^{(2n+1)d(h(Y)+o(1))},
\]

as \( n \to \infty \). Denote the set of patterns of \( X \) in the annulus \( B_{n+m} \setminus B_n \) by \( \mathcal{E}(n,m) \). It is clear that the cardinality of \( \mathcal{E}(n,m) \) is bounded above by \( e^{C(2n+1)^{d-1}} \), for some suitable positive constant \( C \) (which depends on \( m \)). For \( \eta \in \mathcal{E}(n,m) \), we denote by \( \mathcal{P}(\eta) \) the set of all extensions of \( \eta \) in \( B_{n+m} \) that occur in \( X \). Let \( \epsilon > 0 \) such that \( h(X) > h(Y) + 2\epsilon \). From (1) it follows that for \( n \) large enough there exists a \( \eta_0 \in \mathcal{E}(n,m) \) such that

\[
|\mathcal{P}(\eta_0)| \geq e^{(2n+1)^d(h(X) - \epsilon) - C(2n+1)^{d-1}}.
\]

From (2) we see that \( \phi(\mathcal{P}(\eta_0)) \) contains for large \( n \) at most \( \exp \left( (2n+1)^d(h(Y) + \epsilon) \right) \) elements. It follows (since \( h(X) > h(Y) + 2\epsilon \) that for \( n \) large enough, there exists at least one element \( \eta_0 \) in \( \mathcal{B}_{n+m-\alpha}(Y) \) which has two pre-images in \( \mathcal{P}(\eta_0) \). We denote these two pre-images by \( z(1) \) and \( z(2) \). Since \( z(1) \) and \( z(2) \) agree on \( B_{n+m} \setminus B_n \) and \( m > 2\ell \), we can extend \( z(1) \) and \( z(2) \) so that they are also equal outside of \( B_{n+m} \). Since also \( m > 2\alpha \), it follows that the two extensions have the same image under \( \phi \) and therefore form a diamond.

(b) \( \Rightarrow \) (a): Assume there is a diamond \( \{ x(1), x(2) \} \) and assume that \( x(1) \) and \( x(2) \) differ inside the box \( B_n \) only. Let \( m \) be large (we shall see later how large), and construct a new shift space \( X^{(m)} \) as follows. The alphabet of \( X^{(m)} \) consists of all elements of \( \mathcal{B}_m(X) \). An element \( \eta \in \{ \mathcal{B}_m(X) \}^{\mathbb{Z}^d} \) is in \( X^{(m)} \) if the configuration obtained by centering each \( \eta_n \) at location \( (2m+1)z \) is an element of \( X \). There is a natural one-to-one correspondence (which is not a conjugacy!) between \( X \) and \( X^{(m)} \): Cutting a configuration of \( X \) into appropriate disjoint translates of \( B_m \) leads to an element in \( X^{(m)} \) and this process is reversible. It is easy to see that \( X^{(m)} \) is a s.i. SOFT and that its
entropy is equal to \((2m + 1)^d h(X)\). A similar construction leads to the shift space \(Y^{(m)}\), whose entropy is equal to \((2m + 1)^d h(Y)\).

Next we define a second SOFT \(X^{(m)*}\) which is obtained from \(X^{(m)}\) by disallowing those elements of the alphabet \(B_m(X)\) that have a copy of \(\pi_n(x(1))\) in the middle. It follows from Lemma 4.1 that \(h(X^{(m)*}) < h(X^{(m)})\). The map \(\phi\) induces a factor map \(\phi^{(m)}\) from \(X^{(m)}\) to \(Y^{(m)}\) and a map \(\phi^{(m)*}\) from \(X^{(m)*}\) to \(Y^{(m)}\) which satisfies all the conditions of a factor map except perhaps being surjective. We claim that this latter map is onto. To see this, just note that if \(y^{(m)} \in Y^{(m)}\) has a pre-image \(x^{(m)} \in X^{(m)} \setminus X^{(m)*}\), then \(y^{(m)}\) also has a pre-image in \(X^{(m)*}\) which is obtained by replacing all “middle copies” of \(\pi_n(x(1))\) in \(x^{(m)}\) by a copy of \(\pi_n(x(2))\), provided \(m\) is sufficiently large. This leads to the inequality \(h(X^{(m)*}) \geq h(Y^{(m)})\). Putting everything together we obtain

\[
h(X) = \frac{1}{(2m + 1)^d} h(X^{(m)}) > \frac{1}{(2m + 1)^d} h(X^{(m)*}) \geq \frac{1}{(2M + 1)^d} h(Y^{(m)}) = h(Y),
\]

which is what we wanted to prove.

(b) \(\Rightarrow\) (f): Assume there is a diamond \(\{x(1), x(2)\}\) and assume that they differ inside the box \(B_n\) only. Let \(m\) as above be such that \(m > 2\alpha, 2\ell\), where \(\ell\) denotes the size of the largest forbidden pattern in \(X\). Consider a regular rectangular grid of translates of \(B_{n+m}\), where any two such boxes are separated by a distance \(s\), the separation distance of the SOFT. We can ‘fill’ each of these boxes by the appropriate translate of \(\pi_{n+m}(x(1))\). By strong irreducibility and compactness, we can extend this configuration to a configuration \(x_0 \in X\). Define \(y_0 = \phi(x_0)\). Next, we want to replace any of the patterns \(\pi_{n+m}(x(1))\) in one of the boxes of the grid by \(\pi_{n+m}(x(2))\). We claim that this can be done, in that the new configuration is still in \(X\). To see this, note that \(x(2) \in X\) and therefore contains no forbidden pattern. Hence, the only possibility for a forbidden pattern to be created by the replacement of \(\pi_{n+m}(x(1))\) by \(\pi_{n+m}(x(2))\) is that this forbidden pattern intersects both the translate of \(\pi_{n+m}(x(2))\) and the complement of that translate. But since \(m > 2\ell\) and \(x(1)\) and \(x(2)\) agree on \(B_{n+m} \setminus B_n\), this is impossible. It follows by construction, using that \(m > 2\alpha\), that the image under \(\phi\) has not changed by this replacement. Since we can do this in any of the boxes in the grid, it follows straightforwardly that \(y_0\) satisfies the requirement in (f).

(f) \(\Rightarrow\) (d): This is obvious.

(d) \(\Rightarrow\) (c): This is obvious.

(f) \(\Rightarrow\) (e): This is obvious.

(e) \(\Rightarrow\) (c): This is obvious.
(c) ⇒ (b): (This is similar to the proof that (a) ⇒ (b).) Choose \( m \) so large that \( m > 2\alpha, 2\ell \). Since the number of configurations in \( B_{n+m} \setminus B_n \) is of order \( c^{(2n+1)^d-1} \) (as \( m \) is fixed), there exists \( \eta_0 \) defined on \( B_{n+m} \setminus B_n \) such that

\[
\{ x \in B_{n+m}(X) : x = \eta_0 \text{ on } B_{n+m} \setminus B_n, \phi(\pi_n(x)) = \pi_{n-\alpha}(y) \} \geq k_1 c_1 (2n+1)^d.
\]

In particular, there exist \( x_1 \neq x_2 \) in the above set. Since \( m > 2\ell \), we can extend \( x_1 \) and \( x_2 \) to \( \tilde{x}_1 \) and \( \tilde{x}_2 \) in \( X \) such that they agree outside of \( B_{n+m} \) (and hence outside of \( B_n \)). Since \( m > 2\alpha \), \( \phi(\tilde{x}_1) = \phi(\tilde{x}_2) \) yielding a diamond. \( \square \)

5. Factor maps and measures of maximal entropy: Proofs.

For the proof of Theorem 3.3 we need the following lemma which comes from [3]. In fact, there is a small detail missing in their proof. When applying the Hahn-Banach Theorem, one should require that the extension also has norm one so that one can then use the fact (see [8], p. 116) that an operator of norm 1 on \( C(X) \) which sends 1 to 1 is a positive operator. It is then necessarily given by integration against some probability measure.

**Lemma 5.1.** Let \( X \) and \( Y \) be shift spaces and let \( \phi : X \to Y \) be a factor map. Let \( \mu \) be a probability measure on \( Y \). Then there exists at least one measure \( \nu \) on \( X \) such that \( \mu \) is the push-forward of \( \nu \) under \( \phi \), i.e., \( \mu = \phi \nu \).

If in addition \( \mu \) is stationary, then \( \nu \) can also be taken to be stationary.

**Proof of Theorem 3.3.** (i) This follows straightforwardly from Lemma 5.1, the fact that factor maps cannot increase measure-theoretic entropy and the variational principle.

(ii) This is more involved and we use an argument based on conditional entropy. We write \( \mu' = \phi \mu \). Since \( h(X) = h(Y) \) it will be enough (using the variational principle) to show that \( h(\mu) = h(\mu') \). The first thing to do is to write \( h(\mu') \) in terms of \( \mu \) and a partition on \( X \). Let \( P_n(Y) \) be the partition of \( Y \) that specifies all coordinates in \( B_n \), and denote by \( P_n(X) \) the partition of \( X \) that specifies all coordinates in \( B_n \). Finally, \( Q_{n+\alpha}(X) \) is the partition of \( X \) that specifies the projection on \( B_n \) of the image under \( \phi \). Note that \( P_{n+\alpha}(X) \) refines \( Q_{n+\alpha}(X) \). We now write

\[
h(\mu') = \lim_{n \to \infty} \frac{h_{\mu'}(P_n(Y))}{(2n+1)^d}
\]

\[
= \lim_{n \to \infty} -\sum_{y \in B_n(Y)} \mu(\phi^{-1}(y)) \log \mu(\phi^{-1}(y)) (2n+1)^d
\]

\[
= \lim_{n \to \infty} \frac{h_\mu(Q_{n+\alpha}(X))}{(2n+2\alpha+1)^d}.
\]

(Here \( \phi^{-1}(y) \) are the elements in \( B_{n+\alpha}(X) \) which map to \( y \).)
We can write the entropy $h(\mu)$ as follows:

\[
\begin{align*}
    h(\mu) &= \lim_{n \to \infty} \frac{h_\mu(P_{n+\alpha}(X))}{(2n + 2\alpha + 1)^d} \\
    &= \lim_{n \to \infty} \left\{ \frac{h_\mu(Q_{n+\alpha}(X))}{(2n + 2\alpha + 1)^d} + \frac{h_\mu(P_{n+\alpha}(X) \mid Q_{n+\alpha}(X))}{(2n + 2\alpha + 1)^d} \right\} \\
    &= h(\mu') + \lim_{n \to \infty} \frac{h_\mu(P_{n+\alpha}(X) \mid Q_{n+\alpha}(X))}{(2n + 2\alpha + 1)^d}.
\end{align*}
\]

Therefore, we need to show that the limit in the last expression is equal to zero. The numerator inside the limit is by definition equal to

\[
\sum_{q \in Q_{n+\alpha}(X)} \mu(q) \left\{ - \sum_{p \in P_{n+\alpha}(X)} \frac{\mu(p \cap q)}{\mu(q)} \log \frac{\mu(p \cap q)}{\mu(q)} \right\}.
\]

Clearly, the term between the curly brackets is itself the entropy of a partition with respect to a probability measure and therefore bounded above by the logarithm of the number of elements in this partition which have positive probability. By Theorem 3.6, we know that for all $y \in Y, \epsilon > 0$, for $n$ large enough we have that

\[
|\{x \in B_{n+\alpha}(X) : \phi(x) = \pi_n(y)\}| \leq e^{\epsilon(2(n+\alpha)+1)^d}.
\]

Let $N(y) = N_\epsilon(y)$ be the smallest number $k$ so that (4) holds for all $n \geq k$. There exists a number $N$ such that

\[
\mu'(\{y : N(y) \leq N\}) = \mu(\{\phi^{-1}\{y : N(y) \leq N\}\}) \geq 1 - \epsilon.
\]

This implies that up to at most an $\epsilon$-portion, all atoms of the partition $Q_{N+\alpha}$ contain less than $\exp(\epsilon(2(N + \alpha) + 1)^d)$ elements. The remaining atoms contain (by the previous remark) at most $|A(X)|^{(2(N+\alpha)+1)^d}$ elements. Hence the expression in (3) is bounded above by

\[
e^{\epsilon(2(N + \alpha) + 1)^d} + \epsilon(2(N + \alpha) + 1)^d \log |A(X)|,
\]

and the proof is complete. \qed

For the proof of Theorem 3.4 we need the following lemma.

**Lemma 5.2.** Let $X$ be a strongly irreducible SOFT, and let $T(X)$ be the set of transitive points of $X$. For every $\mu \in M_{\max}(X)$, we have $\mu(T(X)) = 1$.

**Proof.** Fix a half-line $\ell$ as in the definition of transitivity, and a pattern $p$ of size $n$. Consider a collection of translates $B(1), B(2), \ldots$ of $B_n$ centered at vertices on $\ell$ and such that any two boxes in this collection are separated by at least $s$ (the separation distance of the shift space). By strong irreducibility, the pattern $p$ has positive $\mu$-probability. Hence there is positive $\mu$-probability to see $p$ in $B(1)$. By the property of uniform conditional probabilities (explained in the proof of Lemma 4.1) and the fact that the
distances between the different translates is at least \( s \), it is easy to see that there exists \( \delta = \delta(n) \) such that for all \( k \) and for any conditioning of the configuration on \( \bigcup_{i=1}^{k} B(i) \), the conditional probability to see \( p \) in \( B(k+1) \) is at least \( \delta \). This easily implies that \( \mu \)-a.s. we see \( p \) in some box \( B(k) \).

Finally, we note that there are only countably many halflines \( \ell \) and countably many patterns to check. This shows that the \( \mu \)-probability to see all patterns on every half-line with rational direction is equal to one. \( \square \)

**Proof of Theorem 3.4.** The proof of Proposition 9.2.2 in [4] does not depend on the dimension and implies that \( \phi \) has no diamonds, which implies by Theorem 3.6 that \( h(X) = h(Y) \). In view of Theorem 3.3, it suffices to show that if \( \mu, \nu \in M_{\text{max}}(X) \) with \( \mu \neq \nu \), then \( \phi \mu \neq \phi \nu \).

According to Lemma 5.2, both \( \mu \) and \( \nu \) live on \( T(X) \). Since it is immediate to check that \( \phi \) maps \( T(X) \) into \( T(Y) \), it follows that \( \phi \mu \) and \( \phi \nu \) both live on \( T(Y) \). If \( \mu \neq \nu \), then there exists a cylinder set \( A \) such that \( \mu(A) \neq \nu(A) \). Note that \( \phi(A) \) is measurable, since it is the continuous image of a compact set. We claim that

\[
\phi \mu(\phi(A)) \neq \phi \nu(\phi(A)),
\]

which implies that \( \phi \mu \neq \phi \nu \). To do this, we first observe that

\[
A \subseteq \phi^{-1}\phi(A) \subseteq A \cup (X \setminus T(X)),
\]

The first inclusion is obvious. For the second, suppose that \( x \in \phi^{-1}\phi(A) \) and \( x \in T(X) \). Since the image of a transitive point is automatically transitive (as is easily verified), \( \phi(x) \in T(Y) \) and by the almost invertibility of \( \phi \), \( \phi(x) \) has only one pre-image. This unique pre-image must then be \( x \), and it follows that \( x \in A \). Since by definition \( \mu(\phi^{-1}\phi(A)) = \phi \mu(\phi(A)) \), we take the \( \mu \)-measure in (6), giving

\[
\mu(A) \leq \phi \mu(\phi(A)) \leq \mu(A) + \mu(X \setminus T(X)) = \mu(A),
\]

where the last equation follows from Lemma 5.2. Hence all inequalities in the last equation are in fact equalities. A similar statement is true when we replace \( \mu \) by \( \nu \) and we finally obtain (5). \( \square \)

**Proof of Theorem 3.5.** Let \( X \) be the SOFT \( X_{M,M} \) defined in Section 3, and take \( M \) so large that \( X \) has exactly two ergodic measures of maximal entropy. Denote a block of size 1 by \( \{b_{i,j} : -1 \leq i, j \leq 1\} \). Define the following block map \( \Phi : B_{1}(X) \rightarrow \{A(X)\}^{8} \):

\[
\Phi(\{b_{i,j} : -1 \leq i, j \leq 1\}) = \text{sgn}(b_{0,0}) (b_{-1,1}, b_{0,1}, b_{1,1}, b_{-1,0}, b_{1,0}, b_{-1,-1}, b_{0,-1}, b_{1,-1}),
\]

and denote by \( \phi \) the factor map on \( X \) based on \( \Phi \). Finally, we write \( Y = \phi(X) \).
We first claim that $\phi$ is two to one. To see this fix $y \in Y$. Suppose $x$ is a pre-image of $y$. Given the value of $y$ in the origin, we know the absolute value of the values of $x$ in the set $B_1 \setminus (0, 0)$; in addition, we know which of these values have the same sign. This leaves us with exactly two possibilities. Choose one of these. If we shift the block one unit to the right, say, then we know from the value of $y$ at $(1, 0)$ what the absolute value of $x$ is at the origin. But given our previous choice, we have no freedom in the sign anymore. Continue in this way; it is then clear that our initial choice exhausts all potential freedom, and $y$ has at most two pre-images. As it is clear that $\phi(x) = \phi(y)$, every point has exactly two pre-images.

Now, one observes that $h(X) = h(Y)$ as easily follows from the argument in the previous paragraph. We next claim that $Y$ has a unique measure of maximal entropy. It follows from Theorem 3.3 that every measure of maximal entropy for $Y$ must be a push forward of a measure of maximal entropy for $X$. However, since the two ergodic measures of maximal entropy $\mu^+$ and $\mu^-$ are obtained from each other by flipping coordinates, we see that $\phi \mu^+ = \phi \mu^-$. This last measure is therefore the only measure of maximal entropy for $Y$. □

Acknowledgement. We thank one of the referees for very useful comments which very much improved the paper.

References

[10] RONALD MEESTER AND JEFFREY E. STEIF


Received December 15, 1999 and revised December 15, 2000. The first author research was supported by a grant from the Swedish National Science Foundation.

UNIVERSITY OF UTRECHT
MATHEMATICAL INSTITUTE
P.O. Box 80.010
3508 TA UTRECHT
THE NETHERLANDS
E-mail address: meester@math.uu.nl

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA 30332-0160
E-mail address: steif@math.gatech.edu
<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeffrey Adams</td>
<td>Extensions of tori in $SL(2)$</td>
<td>257</td>
</tr>
<tr>
<td>Frank Betten and Gestur Ólafsson</td>
<td>Causal compactification and Hardy Spaces for spaces of Hermitian type</td>
<td>273</td>
</tr>
<tr>
<td>David P. Blecher and Vern I. Paulsen</td>
<td>Multipliers of operator spaces, and the injective envelope</td>
<td>1</td>
</tr>
<tr>
<td>Yanghyun Byun and Seunghun Yi</td>
<td>Product formula for self-intersection numbers</td>
<td>313</td>
</tr>
<tr>
<td>Daomin Cao, E.S. Noussair and Shusen Yan</td>
<td>Existence and nonexistence of interior-peaked solution for a nonlinear Neumann problem</td>
<td>19</td>
</tr>
<tr>
<td>Nathan M. Dunfield</td>
<td>Alexander and Thurston norms of fibered 3-manifolds</td>
<td>43</td>
</tr>
<tr>
<td>Abdelkhalek Faouzi</td>
<td>Sur les opérateurs nilpotents à images des itérés fermées dans un espace de Banach</td>
<td>59</td>
</tr>
<tr>
<td>Paul M.N. Feehan</td>
<td>Critical-exponent Sobolev norms and the Slice Theorem for the quotient space of connections</td>
<td>71</td>
</tr>
<tr>
<td>Lorenz Halbeisen and Benedikt Löwe</td>
<td>Techniques for approaching the dual Ramsey property in the projective hierarchy</td>
<td>119</td>
</tr>
<tr>
<td>J.A. Jeong, G.H. Park and D.Y. Shin</td>
<td>Stable rank and real rank of graph $C^*$-algebras</td>
<td>331</td>
</tr>
<tr>
<td>Lee Klingler and Lawrence S. Levy</td>
<td>Representation type of commutative Noetherian rings II: Local tameness</td>
<td>387</td>
</tr>
<tr>
<td>Lee Klingler and Lawrence S. Levy</td>
<td>Representation type of commutative Noetherian rings II: Local tameness</td>
<td>387</td>
</tr>
<tr>
<td>A.A. Kosmodem’yanskii Jr.</td>
<td>Application of representation formulae to comparison and nonexistence theorems for elliptic boundary value problems</td>
<td>485</td>
</tr>
<tr>
<td>Antonio Lanteri and Hidetoshi Maeda</td>
<td>Elliptic surfaces and ample vector bundles</td>
<td>147</td>
</tr>
<tr>
<td>Lawrence S. Levy with Lee Klingler</td>
<td></td>
<td>387</td>
</tr>
<tr>
<td>Lawrence S. Levy with Lee Klingler</td>
<td></td>
<td>387</td>
</tr>
<tr>
<td>Benedikt Löwe with Lorenz Halbeisen</td>
<td></td>
<td>119</td>
</tr>
<tr>
<td>Abdallah Lyzzaik</td>
<td>Quasiholomorphic mappings and univalent harmonic mappings onto punctured bounded convex domains</td>
<td>159</td>
</tr>
</tbody>
</table>
Hidetoshi Maeda with Antonio Lanteri 147
Ronald Meester and Jeffrey E. Steif: *Higher-dimensional subshifts of finite type, factor maps and measures of maximal entropy* 497
E.S. Noussair with Daomin Cao and Shusen Yan 19
Yoshiyuki Ohyama and Kouki Taniyama: *Vassiliev invariants of knots in a spatial graph* 191
Gestur Ólafsson with Frank Betten 273
G.H. Park with J.A. Jeong and D.Y. Shin 331
Vern I. Paulsen with David P. Blecher 1
Zinovy Reichstein and Boris Youssin: *Splitting fields of G-varieties* 207
P. Sawyer: *The asymptotic expansion of spherical functions on symmetric cones* 251
D.Y. Shin with J.A. Jeong and G.H. Park 331
Jeffrey E. Steif with Ronald Meester 497
Kouki Taniyama with Yoshiyuki Ohyama 191
Shusen Yan with Daomin Cao and E.S. Noussair 19
Seunghun Yi with Yanghyun Byun 313
Boris Youssin with Zinovy Reichstein 207
Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use \LaTeX, but papers in other varieties of \TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as \LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of Bib\TeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal’s preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.
Extensions of tori in SL(2)  
JEFFREY ADAMS  

257

Causal compactification and Hardy Spaces for spaces of Hermitian type  
FRANK BETTEN AND GESTUR ÓLAFSSON  

273

Product formula for self-intersection numbers  
YANGHYUN BYUN AND SEUNGHUN YI  

313

Stable rank and real rank of graph C*-algebras  
J.A. JEONG, G.H. PARK AND D.Y. SHIN  

331

Representation type of commutative Noetherian rings I: Local wildness  
LEE KLINGLER AND LAWRENCE S. LEVY  

345

Representation type of commutative Noetherian rings II: Local tameness  
LEE KLINGLER AND LAWRENCE S. LEVY  

387

Application of representation formulae to comparison and nonexistence theorems for elliptic boundary value problems  
A.A. KOSMODEM'YANSKII JR.  

485

Higher-dimensional subshifts of finite type, factor maps and measures of maximal entropy  
RONALD MEESTER AND JEFFREY E. STEIF  

497