Documenta Mathematica

Journal der
Deutschen Mathematiker-Vereinigung
Gegründet 1996

Band 18 · 2013

Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, veröffentlicht Forschungsarbeiten aus allen mathematischen Gebieten und wird in traditioneller Weise referiert. Es wird indiziert durch Mathematical Reviews, Science Citation Index Expanded, Zentralblatt für Mathematik.

Artikel können als TEX-Dateien per E-Mail bei einem der Herausgeber eingereicht werden. Hinweise für die Vorbereitung der Artikel können unter der unten angegebenen WWW-Adresse gefunden werden.

Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, publishes research manuscripts out of all mathematical fields and is refereed in the traditional manner. It is indexed in Mathematical Reviews, Science Citation Index Expanded, Zentralblatt für Mathematik.

Manuscripts should be submitted as TEX -files by e-mail to one of the editors. Hints for manuscript preparation can be found under the following web address.

http://www.math.uni-bielefeld.de/documenta

ISSN 1431-0635 (Print), ISSN 1431-0643 (Internet)

SPARC
Leading Edge

Documenta Mathematica is a Leading Edge Partner of SPARC, the Scholarly Publishing and Academic Resource Coalition of the Association of Research Libraries (ARL), Washington DC, USA.

Address of Technical Managing Editor: Ulf Rehmann, Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Typesetting in TEX.

Documenta Mathematica
Band 18, 2013

The Λ-Adic Shimura-Shintani-Waldspurger Correspondence

Matteo Longo, Marc-Hubert Nicole

Received: October 25, 2012

Communicated by Takeshi Saito

Abstract. We generalize the Λ-adic Shintani lifting for $\mathrm{GL}_2(\mathbb{Q})$ to indefinite quaternion algebras over $\mathbb{Q}$.

2010 Mathematics Subject Classification: Primary 11F37, 11F30, 11F85

1. Introduction

Langlands's principle of functoriality predicts the existence of a staggering wealth of transfers (or lifts) between automorphic forms for different reductive groups. In recent years, attempts at the formulation of p-adic variants of Langlands's functoriality have been articulated in various special cases. We prove the existence of the Shimura-Shintani-Waldspurger lift for p-adic families. More precisely, Stevens, building on the work of Hida and Greenberg-Stevens, showed in [21] the existence of a Λ-adic variant of the classical Shintani lifting of [20] for $\mathrm{GL}_2(\mathbb{Q})$. This Λ-adic lifting can be seen as a formal power series with coefficients in a finite extension of the Iwasawa algebra $\Lambda := \mathbb{Z}_p[\![X]\!]$ equipped with specialization maps interpolating classical Shintani lifts of classical modular forms appearing in a given Hida family.
Shimura in [19], resp. Waldspurger in [22] generalized the classical Shimura-Shintani correspondence to quaternion algebras over $\mathbb{Q}$, resp. over any number field. In the p-adic realm, Hida ([7]) constructed a Λ-adic Shimura lifting, while Ramsey ([17]) (resp. Park [12]) extended the Shimura (resp. Shintani) lifting to the overconvergent setting.
In this paper, motivated by ulterior applications to Shimura curves over $\mathbb{Q}$, we generalize Stevens's result to any non-split rational indefinite quaternion algebra B, building on work of Shimura [19] and combining this with a result of Longo-Vigni [9]. Our main result, for which the reader is referred to Theorem 3.8 below, states the existence of a formal power series and specialization maps interpolating Shimura-Shintani-Waldspurger lifts of classical forms in a given

p-adic family of automorphic forms on the quaternion algebra B. The Λ-adic variant of Waldspurger's result appears computationally challenging (see remark in [15, Intro.]), but it seems within reach for real quadratic fields (cf. [13]).

As an example of our main result, we consider the case of families with trivial character. Fix a prime number p and a positive integer N such that $p \nmid N$. Embed the set $\mathbb{Z}^{\geq 2}$ of integers greater or equal to 2 in $\mathrm{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times)$ by sending $k \in \mathbb{Z}^{\geq 2}$ to the character $x \mapsto x^{k-2}$. Let f_∞ be an Hida family of tame level N passing through a form f_0 of level $\Gamma_0(Np)$ and weight k_0. There is a neighborhood U of k_0 in $\mathrm{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times)$ such that, for any $k \in \mathbb{Z}^{\geq 2} \cap U$, the weight k specialization of f_∞ gives rise to an element $f_k \in S_k(\Gamma_0(Np))$. Fix a factorization $N = MD$ with $D > 1$ a square-free product of an even number of primes and $(M, D) = 1$ (we assume that such a factorization exists). Applying the Jacquet-Langlands correspondence we get for any $k \in \mathbb{Z}^{\geq 2} \cap U$ a modular form f_k^{JL} on Γ, which is the group of norm-one elements in an Eichler order R of level Mp contained in the indefinite rational quaternion algebra B of discriminant D. One can show that these modular forms can be p-adically interpolated, up to scaling, in a neighborhood of k_0. More precisely, let $\mathcal{O}$ be the ring of integers of a finite extension F of $\mathbb{Q}_p$ and let $\mathbb{D}$ denote the $\mathcal{O}$-module of $\mathcal{O}$-valued measures on $\mathbb{Z}_p^2$ which are supported on the set of primitive elements in $\mathbb{Z}_p^2$. Let Γ_0 be the group of norm-one elements in an Eichler order $R_0 \subseteq B$ containing R. There is a canonical action of Γ_0 on $\mathbb{D}$ (see [9, §2.4] for its description). Denote by F_k the extension of F generated by the Fourier coefficients of f_k. Then there is an element $\Phi \in H^1(\Gamma_0, \mathbb{D})$ and maps $\rho_k : H^1(\Gamma_0, \mathbb{D}) \longrightarrow H^1(\Gamma, F_k)$ such that $\rho(k)(\Phi) = \phi_k$, the cohomology class associated to f_k^{JL}, with k in a neighborhood of k_0 (for this we need a suitable normalization of the cohomology class associated to f_k^{JL}, which we do not touch for simplicity in this introduction). We view Φ as a quaternionic family of modular forms. To each ϕ_k we may apply the Shimura-Shintani-Waldspurger lifting ([19]) and obtain a modular form h_k of weigh $k + 1/2$, level $4Np$ and trivial character. We show that this collection of forms can be p-adically interpolated. For clarity's sake, we present the liftings and their Λ-adic variants in a diagram, in which the horizontal maps are specialization maps of the p-adic family to weight k; JL stands for the Jacquet-Langlands correspondence; SSW stands for the Shimura-Shintani-Waldspurger lift; and the dotted arrows are constructed in this paper:

$$\begin{tikzcd}
f_\infty \arrow[r, mapsto] \arrow[d, mapsto, "\Lambda-\text{adic JL}"'] & f_k \arrow[d, mapsto, "\text{JL}"] \\
\Phi \arrow[r, mapsto, "\rho_k"] \arrow[d, mapsto, dotted, "\Lambda-\text{adic SSW}"'] & \phi_k \arrow[d, mapsto, "\text{SSW}"] \\
\Theta \arrow[r, mapsto, dotted] & h_k
\end{tikzcd}$$

More precisely, as a particular case of our main result, Theorem 3.8, we get the following

Theorem 1.1. *There exists a p-adic neighborhood U_0 of k_0 in* $\mathrm{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times)$, *$p$-adic periods Ω_k for $k \in U_0 \cap \mathbb{Z}^{\geq 2}$ and a formal expansion*

$$\Theta = \sum_{\xi \geq 1} a_\xi q^\xi$$

with coefficients a_ξ in the ring of $\mathbb{C}_p$-valued functions on U_0, such that for all $k \in U_0 \cap \mathbb{Z}^{\geq 2}$ we have

$$\Theta(k) = \Omega_k \cdot h_k.$$

Further, $\Omega_{k_0} \neq 0$.

2. Shintani integrals and Fourier coefficients of half-integral weight modular forms

We express the Fourier coefficients of half-integral weight modular forms in terms of period integrals, thus allowing a cohomological interpretation which is key to the production of the Λ-adic version of the Shimura-Shintani-Waldspurger correspondence. For the quaternionic Shimura-Shintani-Waldspurger correspondence of interest to us (see [15], [22]), the period integrals expressing the values of the Fourier coefficients have been computed generally by Prasanna in [16].

2.1. The Shimura-Shintani-Waldspurger lifting. Let $4M$ be a positive integer, $2k$ an even non-negative integer and χ a Dirichlet character modulo $4M$ such that $\chi(-1) = 1$. Recall that the space of half-integral weight modular forms $S_{k+1/2}(4M, \chi)$ consists of holomorphic cuspidal functions h on the upper-half place $\mathfrak{H}$ such that

$$h(\gamma(z)) = j^{1/2}(\gamma, z)^{2k+1}\chi(d)h(z),$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4M)$, where $j^{1/2}(\gamma, z)$ is the standard square root of the usual automorphy factor $j(\gamma, z)$ (cf. [15, 2.3]).
To any quaternionic integral weight modular form we may associate a half-integral weight modular form following Shimura's work [19], as we will describe below.
Fix an odd square free integer N and a factorization $N = M \cdot D$ into coprime integers such that $D > 1$ is a product of an even number of distinct primes. Fix a Dirichlet character ψ modulo M and a positive even integer $2k$. Suppose that

$$\psi(-1) = (-1)^k.$$

Define the Dirichlet character χ modulo $4N$ by

$$\chi(x) := \psi(x) \left(\frac{-1}{x}\right)^k.$$

Let B be an indefinite quaternion algebra over $\mathbb{Q}$ of discriminant D. Fix a maximal order $\mathcal{O}_B$ of B. For every prime $\ell|M$, choose an isomorphism

$$i_\ell : B \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \simeq \mathbb{M}_2(\mathbb{Q}_\ell)$$

such that $i_\ell(\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_\ell) = \mathbb{M}_2(\mathbb{Z}_\ell)$. Let $R \subseteq \mathcal{O}_B$ be the Eichler order of B of level M defined by requiring that $i_\ell(R \otimes_{\mathbb{Z}} \mathbb{Z}_\ell)$ is the suborder of $\mathbb{M}_2(\mathbb{Z}_\ell)$ of upper triangular matrices modulo ℓ for all $\ell|M$. Let Γ denote the subgroup of the group $R_1^\times$ of norm 1 elements in $R^\times$ consisting of those γ such that $i_\ell(\gamma) \equiv \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ mod ℓ for all $\ell|M$. We denote by $S_{2k}(\Gamma)$ the $\mathbb{C}$-vector space of weight $2k$ modular forms on Γ, and by $S_{2k}(\Gamma, \psi^2)$ the subspace of $S_{2k}(\Gamma)$ consisting of forms having character ψ^2 under the action of $R_1^\times$. Fix a Hecke eigenform

$$f \in S_{2k}(\Gamma, \psi^2)$$

as in [19, Section 3].

Let V denote the $\mathbb{Q}$-subspace of B consisting of elements with trace equal to zero. For any $v \in V$, which we view as a trace zero matrix in $\mathbb{M}_2(\mathbb{R})$ (after fixing an isomorphism $i_\infty : B \otimes \mathbb{R} \simeq \mathbb{M}_2(\mathbb{R})$), set

$$G_v := \{\gamma \in \mathrm{SL}_2(\mathbb{R}) |\, \gamma^{-1} v \gamma = v\}$$

and put $\Gamma_v := G_v \cap \Gamma$. One can show that there exists an isomorphism

$$\omega : \mathbb{R}^\times \xrightarrow{\sim} G_v$$

defined by $\omega(s) = \beta^{-1}\left(\begin{smallmatrix} s & 0 \\ 0 & s^{-1} \end{smallmatrix}\right)\beta$, for some $\beta \in \mathrm{SL}_2(\mathbb{R})$. Let $\mathfrak{t}_v$ be the order of $\Gamma_v \cap \{\pm 1\}$ and let γ_v be an element of Γ_v which generates $\Gamma_v \{\pm 1\} / \{\pm 1\}$. Changing γ_v to γ_v^{-1} if necessary, we may assume $\gamma_v = \omega(t)$ with $t > 0$. Define V^* to be the $\mathbb{Q}$-subspace of V consisting of elements with strictly negative norm. For any $\alpha = \left(\begin{smallmatrix} a & b \\ c & -a \end{smallmatrix}\right) \in V^*$ and $z \in \mathcal{H}$, define the quadratic form

$$Q_\alpha(z) := cz^2 - 2az - b.$$

Fix $\tau \in \mathcal{H}$ and set

$$P(f, \alpha, \Gamma) := -\big(2(-\mathrm{nr}(\alpha))^{1/2}/\mathfrak{t}_\alpha\big) \int_\tau^{\gamma_\alpha(\tau)} Q_\alpha(z)^{k-1} f(z) dz$$

where $\mathrm{nr} : B \to \mathbb{Q}$ is the norm map. By [19, Lemma 2.1], the integral is independent on the choice τ, which justifies the notation.

Remark 2.1. The definition of $P(f, \alpha, \Gamma)$ given in [19, (2.5)] looks different: the above expression can be derived as in [19, page 629] by means of [19, (2.20) and (2.22)].

Let $R(\Gamma)$ denote the set of equivalence classes of V^* under the action of Γ by conjugation. By [19, (2.6)], $P(f, \alpha, \Gamma)$ only depends on the conjugacy class of α, and thus, for $\mathcal{C} \in R(\Gamma)$, we may define $P(f, \mathcal{C}, \Gamma) := P(f, \alpha, \Gamma)$ for any choice of $\alpha \in \mathcal{C}$. Also, $q(\mathcal{C}) := -\mathrm{nr}(\alpha)$ for any $\alpha \in \mathcal{C}$.

Define $\mathcal{O}'_B$ to be the maximal order in B such that $\mathcal{O}'_B \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \simeq \mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_\ell$ for all $\ell \nmid M$ and $\mathcal{O}'_B \otimes_{\mathbb{Z}} \mathbb{Z}_\ell$ is equal to the local order of $B \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ consisting of elements γ such that $i_\ell(\gamma) = \left(\begin{smallmatrix} a & b/M \\ cM & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathbb{Z}_\ell$, for all $\ell | M$. Given $\alpha \in \mathcal{O}'_B$, we can find an integer b_α such that

$$(1) \qquad i_\ell(\alpha) \equiv \begin{pmatrix} * & b_\alpha/M \\ * & * \end{pmatrix} \quad \mathrm{mod}\ i_\ell(R \otimes_{\mathbb{Z}} \mathbb{Z}_\ell), \quad \forall \ell | M.$$

Define a locally constant function η_ψ on V by $\eta_\psi(\alpha) = \psi(b_\alpha)$ if $\alpha \in \mathcal{O}'_B \cap V$ and $\eta(\alpha) = 0$ otherwise, with $\psi(a) = 0$ if $(a, M) \neq 1$ (for the definition of locally constant functions on V in this context, we refer to [19, p. 611]).
For any $\mathcal{C} \in R(\Gamma)$, fix $\alpha_{\mathcal{C}} \in \mathcal{C}$. For any integer $\xi \geq 1$, define

$$a_\xi(\tilde{h}) := \left(2\mu(\Gamma \backslash \mathfrak{H})\right)^{-1} \cdot \sum_{\mathcal{C} \in R(\Gamma), q(\mathcal{C}) = \xi} \eta_\psi(\alpha_{\mathcal{C}}) \xi^{-1/2} P(f, \mathcal{C}, \Gamma).$$

Then, by [19, Theorem 3.1],

$$\tilde{h} := \sum_{\xi \geq 1} a_\xi(\tilde{h}) q^\xi \in S_{k+1/2}(4N, \chi)$$

is called the Shimura-Shintani-Waldspurger lifting of f.

2.2. Cohomological interpretation. We introduce necessary notation to define the action of the Hecke action on cohomology groups; for details, see [9, §2.1]. If G is a subgroup of $B^\times$ and S a subsemigroup of $B^\times$ such that (G, S) is an Hecke pair, we let $\mathcal{H}(G, S)$ denote the Hecke algebra corresponding to (G, S), whose elements are written as $T(s) = GsG = \coprod_i Gs_i$ for $s, s_i \in S$ (finite disjoint union). For any $s \in S$, let $s^* := \mathrm{norm}(s)s^{-1}$ and denote by S^* the set of elements of the form s^* for $s \in S$. For any $\mathbb{Z}[S^*]$-module M we let $T(s)$ act on $H^1(G, M)$ at the level of cochains $c \in Z^1(G, M)$ by the formula $(c|T(s))(\gamma) = \sum_i s_i^* c(t_i(\gamma))$, where $t_i(\gamma)$ are defined by the equations $Gs_i\gamma = Gs_j$ and $s_i\gamma = t_i(\gamma)s_j$. In the following, we will consider the case of $G = \Gamma$ and

$$S = \{s \in B^\times | i_\ell(s) \text{ is congruent to } \left(\begin{smallmatrix} 1 & * \\ 0 & * \end{smallmatrix}\right) \text{ mod } \ell \text{ for all } \ell | M\}.$$

For any field L and any integer $n \geq 0$, let $V_n(L)$ denote the L-dual of the L-vector space $\mathcal{P}_n(L)$ of homogeneous polynomials in 2 variables of degree n. We let $\mathbb{M}_2(L)$ act from the right on $P(x, y)$ as $P|\gamma(x, y) := P(\gamma(x, y))$, where for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ we put

$$\gamma(x, y) := (ax + yb, cx + dy).$$

This also equips $V_n(L)$ with a left action by $\gamma \cdot \varphi(P) := \varphi(P|\gamma)$. To simplify the notation, we will write $P(z)$ for $P(z, 1)$.
Let F denote the finite extension of $\mathbb{Q}$ generated by the eigenvalues of the Hecke action on f. For any field K containing F, set

$$\mathbb{W}_f(K) := H^1\left(\Gamma, V_{k-2}(K)\right)^f$$

where the superscript f denotes the subspace on which the Hecke algebra acts via the character associated with f. Also, for any sign $\pm$, let $\mathbb{W}_f^\pm(K)$ denote the $\pm$-eigenspace for the action of the archimedean involution ι. Remember that ι is defined by choosing an element ω_∞ of norm -1 in $R^\times$ such that such

that $i_\ell(\omega_\infty) \equiv \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ mod M for all primes $\ell | M$ and then setting $\iota := T(w_\infty)$ (see [9, §2.1]). Then $\mathbb{W}_f^\pm(K)$ is one dimensional (see, e.g., [9, Proposition 2.2]); fix a generator $\phi_f^\pm$ of $\mathbb{W}_f^\pm(F)$.
To explicitly describe $\phi_f^\pm$, let us introduce some more notation. Define

$$f|\omega_\infty(z) := (Cz+D)^{-k/2}\overline{f(\omega_\infty(\bar{z}))}$$

where $i_\infty(\omega_\infty) = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$. Then $f|\omega_\infty \in S_{2k}(\Gamma)$ as well. If the eigenvalues of the Hecke action on f are real, then we may assume, after multiplying f by a scalar, that $f|\omega_\infty = f$ (see [19, p. 627] or [10, Lemma 4.15]). In general, let $I(f)$ denote the class in $H^1(\Gamma, V_{k-2}(\mathbb{C}))$ represented by the cocycle

$$\gamma \longmapsto \left[P \mapsto I_\gamma(f)(P) := \int_\tau^{\gamma(\tau)} f(z)P(z)dz \right]$$

for any $\tau \in \mathcal{H}$ (the corresponding class is independent on the choice of τ). With this notation,

$$P(f,\alpha,\Gamma) = -\big(2(-\mathrm{nr}(\alpha))^{1/2}/\mathfrak{t}_\alpha\big) \cdot I_{\gamma_{\alpha_\mathcal{C}}}(f)\big(Q_{\alpha_\mathcal{C}}(z)^{k-1}\big).$$

Denote by $I^\pm(f) := (1/2)\cdot I(f) \pm (1/2)\cdot I(f)|\omega_\infty$, the projection of $I(f)$ to the eigenspaces for the action of ω_∞. Then $I(f) = I^+(f) + I^-(f)$ and $I_f^\pm = \Omega_f^\pm \cdot \phi_f^\pm$, for some $\Omega_f^\pm \in \mathbb{C}^\times$.
Given $\alpha \in V^*$ of norm $-\xi$, put $\alpha' := \omega_\infty^{-1}\alpha\omega_\infty$. By [19, 4.19], we have

$$\eta(\alpha)\xi^{-1/2}P(f,\alpha,\Gamma) + \eta(\alpha')\xi^{-1/2}P(f,\alpha',\Gamma) = -\eta(\alpha)\cdot \mathfrak{t}_\alpha{}^{-1} \cdot I^+_{\gamma_\alpha}\big(Q_{\alpha_\mathcal{C}}(z)^{k-1}\big).$$

We then have

$$a_\xi(\tilde{h}) = \sum_{\mathcal{C}\in R_2(\Gamma), q(\mathcal{C})=\xi} \frac{-\eta_\psi(\alpha_\mathcal{C})}{2\mu(\Gamma\backslash\mathcal{H})\cdot \mathfrak{t}_{\alpha_\mathcal{C}}} \cdot I^+_{\gamma_{\alpha_\mathcal{C}}}\big(Q_{\alpha_\mathcal{C}}(z)^{k-1}\big).$$

We close this section by choosing a suitable multiple of h which will be the object of the next section. Given $Q_\alpha(z) = cz^2 - 2az - b$ as above, with α in V^*, define $\tilde{Q}_\alpha(z) := M \cdot Q_\alpha(z)$. Then, clearly, $I^\pm(f)(\tilde{Q}_{\alpha_\mathcal{C}}(z)^{k-1})$ is equal to $M^{k-1}I^\pm(f)(Q_{\alpha_\mathcal{C}}(z)^{k-1})$. We thus normalize the Fourier coefficients by setting
(2)

$$a_\xi(h) := -\frac{a_\xi(\tilde{h})\cdot M^{k-1}\cdot 2\mu(\Gamma\backslash\mathcal{H})}{\Omega_f^+} = \sum_{\mathcal{C}\in R(\Gamma), q(\mathcal{C})=\xi} \frac{\eta_\psi(\alpha_\mathcal{C})}{\mathfrak{t}_{\alpha_\mathcal{C}}}\cdot \phi_f^+\big(\tilde{Q}_{\alpha_\mathcal{C}}(z)^{k-1}\big).$$

So

$$h := \sum_{\xi\geq 1} a_\xi(h)q^\xi \tag{3}$$

belongs to $S_{k+1/2}(4N,\chi)$ and is a non-zero multiple of $\tilde{h}$.

3. The Λ-adic Shimura-Shintani-Waldspurger correspondence

At the heart of Stevens's proof lies the control theorem of Greenberg-Stevens, which has been worked out in the quaternionic setting by Longo–Vigni [9]. Recall that $N \geq 1$ is a square free integer and fix a decomposition $N = M \cdot D$ where D is a square free product of an even number of primes and M is coprime to D. Let $p \nmid N$ be a prime number and fix an embedding $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_p$.

3.1. The Hida Hecke algebra. Fix an ordinary p-stabilized newform

$$f_0 \in S_{k_0}\big(\Gamma_1(Mp^{r_0}) \cap \Gamma_0(D), \epsilon_0\big) \tag{4}$$

of level $\Gamma_1(Mp^{r_0}) \cap \Gamma_0(D)$, Dirichlet character ϵ_0 and weight k_0, and write $\mathcal{O}$ for the ring of integers of the field generated over $\mathbb{Q}_p$ by the Fourier coefficients of f_0.

Let Λ (respectively, $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$) denote the Iwasawa algebra of $W := 1 + p\mathbb{Z}_p$ (respectively, $\mathbb{Z}_p^\times$) with coefficients in $\mathcal{O}$. We denote group-like elements in Λ and $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$ as $[t]$. Let $\mathfrak{h}_\infty^{\mathrm{ord}}$ denote the p-ordinary Hida Hecke algebra with coefficients in $\mathcal{O}$ of tame level $\Gamma_1(N)$. Denote by $\mathcal{L} := \mathrm{Frac}(\Lambda)$ the fraction field of Λ. Let $\mathcal{R}$ denote the integral closure of Λ in the primitive component $\mathcal{K}$ of $\mathfrak{h}_\infty^{\mathrm{ord}} \otimes_\Lambda \mathcal{L}$ corresponding to f_0. It is well known that the Λ-algebra $\mathcal{R}$ is finitely generated as Λ-module.

Denote by $\mathcal{X}$ the $\mathcal{O}$-module $\mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}\text{-alg}}(\mathcal{R}, \bar{\mathbb{Q}}_p)$ of continuous homomorphisms of $\mathcal{O}$-algebras. Let $\mathcal{X}^{\mathrm{arith}}$ the set of arithmetic homomorphisms in $\mathcal{X}$, defined in [9, §2.2] by requiring that the composition

$$W \hookrightarrow \Lambda \xrightarrow{\kappa} \bar{\mathbb{Q}}_p$$

has the form $\gamma \mapsto \psi_\kappa(\gamma)\gamma^{n_\kappa}$ with $n_\kappa = k_\kappa - 2$ for an integer $k_\kappa \geq 2$ (called the weight of κ) and a finite order character $\psi_\kappa : W \to \bar{\mathbb{Q}}_p$ (called the wild character of κ). Denote by r_κ the smallest among the positive integers t such that $1 + p^t\mathbb{Z}_p \subseteq \ker(\psi_\kappa)$. For any $\kappa \in \mathcal{X}^{\mathrm{arith}}$, let P_κ denote the kernel of κ and $\mathcal{R}_{P_\kappa}$ the localization of $\mathcal{R}$ at κ. The field $F_\kappa := \mathcal{R}_{P_\kappa}/P_\kappa\mathcal{R}_{P_\kappa}$ is a finite extension of $\mathrm{Frac}(\mathcal{O})$. Further, by duality, κ corresponds to a normalized eigenform

$$f_\kappa \in S_{k_\kappa}\big(\Gamma_0(Np^{r_\kappa}), \epsilon_\kappa\big)$$

for a Dirichlet character $\epsilon_\kappa : (\mathbb{Z}/Np^{r_\kappa}\mathbb{Z})^\times \to \bar{\mathbb{Q}}_p$. More precisely, if we write $\psi_{\mathcal{R}}$ for the character of $\mathcal{R}$, defined as in [6, Terminology p. 555], and we let ω denote the Teichmüller character, we have $\epsilon_\kappa := \psi_\kappa \cdot \psi_{\mathcal{R}} \cdot \omega^{-n_\kappa}$ (see [6, Cor. 1.6]). We call $(\epsilon_\kappa, k_\kappa)$ the signature of κ. We let κ_0 denote the arithmetic character associated with f_0, so $f_0 = f_{\kappa_0}$, $k_0 = k_{\kappa_0}$, $\epsilon_0 = \epsilon_{\kappa_0}$, and $r_0 = r_{\kappa_0}$. The eigenvalues of f_κ under the action of the Hecke operators T_n ($n \geq 1$ an integer) belong to F_κ. Actually, one can show that f_κ is a p-stabilized newform on $\Gamma_1(Mp^{r_\kappa}) \cap \Gamma_0(D)$.

Let Λ_N denote the Iwasawa algebra of $\mathbb{Z}_p^\times \times (\mathbb{Z}/N\mathbb{Z})^\times$ with coefficients in $\mathcal{O}$. To simplify the notation, define $\Delta := (\mathbb{Z}/Np\mathbb{Z})^\times$. We have a canonical isomorphism of rings $\Lambda_N \simeq \Lambda[\Delta]$, which makes Λ_N a Λ-algebra, finitely generated as

Λ-module. Define the tensor product of Λ-algebras

$$\mathcal{R}_N := \mathcal{R} \otimes_\Lambda \Lambda_N,$$

which is again a Λ-algebra (resp. Λ_N-algebra) finitely generated as a Λ-module, (resp. as a Λ_N-module). One easily checks that there is a canonical isomorphism of Λ-algebras

$$\mathcal{R}_N \simeq \mathcal{R}[\Delta]$$

(where Λ acts on $\mathcal{R}$); this is also an isomorphism of Λ_N-algebras, when we let $\Lambda_N \simeq \Lambda[\Delta]$ act on $\mathcal{R}[\Delta]$ in the obvious way.
We can extend any $\kappa \in \mathcal{X}^{\mathrm{arith}}$ to a continuous $\mathcal{O}$-algebra morphism

$$\kappa_N : \mathcal{R}_N \longrightarrow \bar{\mathbb{Q}}_p$$

setting

$$\kappa_N\left(\sum_{i=1}^n r_i \cdot \delta_i\right) := \sum_{i=1}^n \kappa(r_i) \cdot \psi_{\mathcal{R}}(\delta_i)$$

for $r_i \in \mathcal{R}$ and $\delta_i \in \Delta$. Therefore, κ_N restricted to $\mathbb{Z}_p^\times$ is the character $t \mapsto \epsilon_\kappa(t)t^{n_\kappa}$. If we denote by $\mathcal{X}_N$ the $\mathcal{O}$-module of continuous $\mathcal{O}$-algebra homomorphisms from $\mathcal{R}_N$ to $\bar{\mathbb{Q}}_p$, the above correspondence sets up an injective map $\mathcal{X}^{\mathrm{arith}} \hookrightarrow \mathcal{X}_N$. Let $\mathcal{X}_N^{\mathrm{arith}}$ denote the image of $\mathcal{X}^{\mathrm{arith}}$ under this map. For $\kappa_N \in \mathcal{X}_N^{\mathrm{arith}}$, we define the signature of κ_N to be that of the corresponding κ.

3.2. The control theorem in the quaternionic setting. Recall that $B/\mathbb{Q}$ is a quaternion algebra of discriminant D. Fix an auxiliary real quadratic field F such that all primes dividing D are inert in F and all primes dividing Mp are split in F, and an isomorphism $i_F : B \otimes_{\mathbb{Q}} F \simeq \mathbb{M}_2(F)$. Let $\mathcal{O}_B$ denote the maximal order of B obtained by taking the intersection of B with $\mathbb{M}_2(\mathcal{O}_F)$, where $\mathcal{O}_F$ is the ring of integers of F. More precisely, define

$$\mathcal{O}_B := \iota^{-1}\big(i_F^{-1}\big(i_F(B \otimes 1) \cap \mathbb{M}_2(\mathcal{O}_F)\big)\big)$$

where $\iota : B \hookrightarrow B \otimes_{\mathbb{Q}} F$ is the inclusion defined by $b \mapsto b \otimes 1$. This is a maximal order in B because $i_F(B \otimes 1) \cap \mathbb{M}_2(\mathcal{O}_F)$ is a maximal order in $i_F(B \otimes 1)$. In particular, i_F and our fixed embedding of $\bar{\mathbb{Q}}$ into $\bar{\mathbb{Q}}_p$ induce an isomorphism

$$i_p : B \otimes_{\mathbb{Q}} \mathbb{Q}_p \simeq \mathbb{M}_2(\mathbb{Q}_p)$$

such that $i_p(\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_p) = \mathbb{M}_2(\mathbb{Z}_p)$. For any prime $\ell | M$, also choose an embedding $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_\ell$ which, composed with i_F, yields isomorphisms

$$i_\ell : B \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \simeq \mathbb{M}_2(\mathbb{Q}_\ell)$$

such that $i_p(\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_\ell) = \mathbb{M}_2(\mathbb{Z}_\ell)$. Define an Eichler order $R \subseteq \mathcal{O}_B$ of level M by requiring that for all primes $\ell | M$ the image of $R \otimes_{\mathbb{Z}} \mathbb{Z}_\ell$ via i_ℓ consists of upper triangular matrices modulo ℓ. For any $r \geq 0$, let Γ_r denote the subgroup of the group $R_1^\times$ of norm-one elements in R consisting of those γ such that $i_\ell(\gamma) = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $c \equiv 0 \mod Mp^r$ and $a \equiv d \equiv 1 \mod Mp^r$, for all primes $\ell | Mp$. To conclude this list of notation and definitions, fix an embedding $F \hookrightarrow \mathbb{R}$ and let

$$i_\infty : B \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{M}_2(\mathbb{R})$$

be the induced isomorphism.
Let $\mathbb{Y} := \mathbb{Z}_p^2$ and denote by $\mathbb{X}$ the set of primitive vectors in $\mathbb{Y}$. Let $\mathbb{D}$ denote the $\mathcal{O}$-module of $\mathcal{O}$-valued measures on $\mathbb{Y}$ which are supported on $\mathbb{X}$. Note that $\mathbb{M}_2(\mathbb{Z}_p)$ acts on $\mathbb{Y}$ by left multiplication; this induces an action of $\mathbb{M}_2(\mathbb{Z}_p)$ on the $\mathcal{O}$-module of $\mathcal{O}$-valued measures on $\mathbb{Y}$, which induces an action on $\mathbb{D}$. The group $R^\times$ acts on $\mathbb{D}$ via i_p. In particular, we may define the group:

$$\mathbb{W} := H^1(\Gamma_0, \mathbb{D}).$$

Then $\mathbb{D}$ has a canonical structure of $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$-module, as well as $\mathfrak{h}_\infty^{\mathrm{ord}}$-action, as described in [9, §2.4]. In particular, let us recall that, for any $[t] \in \mathcal{O}[\![\mathbb{Z}_p^\times]\!]$, we have

$$\int_{\mathbb{X}} \varphi(x,y) d\big([t]\cdot\nu\big) = \int_{\mathbb{X}} \varphi(tx,ty)d\nu,$$

for any locally constant function φ on $\mathbb{X}$.
For any $\kappa \in \mathcal{X}^{\mathrm{arith}}$ and any sign $\pm \in \{-,+\}$, set

$$\mathbb{W}_\kappa^\pm := \mathbb{W}_{f_\kappa^{\mathrm{JL}}}^\pm(F_\kappa) = H^1\big(\Gamma_{r_\kappa}, V_{n_\kappa}(F_\kappa)\big)^{f_\kappa,\pm}$$

where f_κ^{JL} is any Jacquet-Langlands lift of f_κ to Γ_{r_κ}; recall that the superscript f_κ denotes the subspace on which the Hecke algebra acts via the character associated with f_κ, and the superscript $\pm$ denotes the $\pm$-eigenspace for the action of the archimedean involution ι. Also, recall that $\mathbb{W}_\kappa^\pm$ is one dimensional and fix a generator $\phi_\kappa^\pm$ of it.
We may define specialization maps

$$\rho_\kappa : \mathbb{D} \longrightarrow V_{n_\kappa}(F_\kappa)$$

by the formula

$$\rho_\kappa(\nu)(P) := \int_{\mathbb{Z}_p\times\mathbb{Z}_p^\times} \epsilon_\kappa(y)P(x,y)d\nu \tag{5}$$

which induces (see [9, §2.5]) a map:

$$\rho_\kappa : \mathbb{W}^{\mathrm{ord}} \longrightarrow \mathbb{W}_\kappa^{\mathrm{ord}}.$$

Here $\mathbb{W}^{\mathrm{ord}}$ and $\mathbb{W}_\kappa^{\mathrm{ord}}$ denote the ordinary submodules of $\mathbb{W}$ and $\mathbb{W}_\kappa$, respectively, defined as in [3, Definition 2.2] (see also [9, §3.5]). We also let $\mathbb{W}_{\mathcal{R}} := \mathbb{W} \otimes_\Lambda \mathcal{R}$, and extend the above map ρ_κ to a map

$$\rho_\kappa : \mathbb{W}_{\mathcal{R}}^{\mathrm{ord}} \longrightarrow \mathbb{W}_\kappa^{\mathrm{ord}}$$

by setting $\rho_\kappa(x\otimes r) := \rho_\kappa(x)\cdot\kappa(r)$.

Theorem 3.1. *There exists a p-adic neighborhood $\mathcal{U}_0$ of κ_0 in $\mathcal{X}$, elements $\Phi^\pm$ in $\mathbb{W}_{\mathcal{R}}^{\mathrm{ord}}$ and choices of p-adic periods $\Omega_\kappa^\pm \in F_\kappa$ for $\kappa \in \mathcal{U}_0 \cap \mathcal{X}^{\mathrm{arith}}$ such that, for all $\kappa \in \mathcal{U}_0 \cap \mathcal{X}^{\mathrm{arith}}$, we have*

$$\rho_\kappa(\Phi^\pm) = \Omega_\kappa^\pm \cdot \phi_\kappa^\pm$$

and $\Omega^{\pm}_{\kappa_0} \neq 0$.

Proof. This is an easy consequence of [9, Theorem 2.18] and follows along the lines of the proof of [21, Theorem 5.5], cf. [10, Proposition 3.2]. $\square$

We now normalize our choices as follows. With $\mathcal{U}_0$ as above, define

$$\mathcal{U}_0^{\mathrm{arith}} := \mathcal{U}_0 \cap \mathcal{X}^{\mathrm{arith}}.$$

Fix $\kappa \in \mathcal{U}_0^{\mathrm{arith}}$ and an embedding $\bar{\mathbb{Q}}_p \hookrightarrow \mathbb{C}$. Let f_κ^{JL} denote a modular form on Γ_{r_κ} corresponding to f_κ by the Jacquet-Langlands correspondence, which is well defined up to elements in $\mathbb{C}^\times$. View $\phi_\kappa^{\pm}$ as an element in $H^1(\Gamma_{r_\kappa}, V_n(\mathbb{C}))^{\pm}$. Choose a representative $\Phi^{\pm}_\gamma$ of $\Phi^{\pm}$, by which we mean that if $\Phi^{\pm} = \sum_i \Phi_i^{\pm} \otimes r_i$, then we choose a representative $\Phi^{\pm}_{i,\gamma}$ for all i. Also, we will write $\rho_\kappa(\Phi)(P)$ as

$$\int_{\mathbb{Z}_p\times\mathbb{Z}_p^\times} \epsilon_\kappa(y)P(x,y)d\Phi^{\pm}_\gamma := \sum_i \kappa(r_i)\cdot \int_{\mathbb{Z}_p\times\mathbb{Z}_p^\times} \epsilon_\kappa(y)P(x,y)d\Phi^{\pm}_{i,\gamma}.$$

With this notation, we see that the two cohomology classes

$$\gamma \longmapsto \int_{\mathbb{Z}_p\times\mathbb{Z}_p^\times} \epsilon_\kappa(y)P(x,y)d\Phi^{\pm}_\gamma(x,y)$$

and

$$\gamma \longmapsto \Omega_\kappa^{\pm}\cdot \int_\tau^{\gamma(\tau)} P(z,1)f_\kappa^{\mathrm{JL},\pm}(z)dz$$

are cohomologous in $H^1(\Gamma_{r_\kappa}, V_{n_\kappa}(\mathbb{C}))$, for any choice of $\tau \in \mathcal{H}$.

3.3. Metaplectic Hida Hecke algebras. Let $\sigma : \Lambda_N \to \Lambda_N$ be the ring homomorphism associated to the group homomorphism $t \mapsto t^2$ on $\mathbb{Z}_p^\times \times (\mathbb{Z}/N\mathbb{Z})^\times$, and denote by the same symbol its restriction to Λ and $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$. We let Λ_σ, $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]_\sigma$ and $\Lambda_{N,\sigma}$ denote, respectively, Λ, $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$ and Λ_N viewed as algebras over themselves via σ. The ordinary metaplectic p-adic Hida Hecke algebra we will consider is the Λ-algebra

$$\widetilde{\mathcal{R}} := \mathcal{R} \otimes_\Lambda \Lambda_\sigma.$$

Define as above

$$\widetilde{\mathcal{X}} := \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}\text{-alg}}(\widetilde{\mathcal{R}}, \bar{\mathbb{Q}}_p)$$

and let the set $\widetilde{\mathcal{X}}^{\mathrm{arith}}$ of arithmetic points in $\widetilde{\mathcal{X}}$ to consist of those $\tilde\kappa$ such that the composition

$$W \hookrightarrow \Lambda \xhookrightarrow{\lambda\mapsto 1\otimes\lambda} \widetilde{\mathcal{R}} \xhookrightarrow{\tilde\kappa} \bar{\mathbb{Q}}_p$$

has the form $\gamma \mapsto \psi_{\tilde\kappa}(\gamma)\gamma^{n_{\tilde\kappa}}$ with $n_{\tilde\kappa} := k_{\tilde\kappa} - 2$ for an integer $k_{\tilde\kappa} \geq 2$ (called the weight of $\tilde\kappa$) and a finite order character $\psi_{\tilde\kappa} : W \to \bar{\mathbb{Q}}$ (called the wild character of $\tilde\kappa$). Let $r_{\tilde\kappa}$ the smallest among the positive integers t such that $1 + p^t\mathbb{Z}_p \subseteq \ker(\psi_{\tilde\kappa})$.

We have a map $p : \widetilde{\mathcal{X}} \to \mathcal{X}$ induced by pull-back from the canonical map $\mathcal{R} \to \widetilde{\mathcal{R}}$. The map p restricts to arithmetic points.

As above, define the Λ-algebra (or Λ_N-algebra)

$$\widetilde{\mathcal{R}}_N := \mathcal{R} \otimes_\Lambda \Lambda_{N,\sigma} \tag{6}$$

via $\lambda \mapsto 1 \otimes \lambda$.

We easily see that $\widetilde{\mathcal{R}}_N \simeq \widetilde{\mathcal{R}}[\Delta]$ as Λ_N-algebras, where we enhance $\widetilde{\mathcal{R}}[\Delta]$ with the following structure of $\Lambda_N \simeq \Lambda[\Delta]$-algebra: for $\sum_i \lambda_i \cdot \delta_i \in \Lambda[\Delta]$ (with $\lambda_i \in \Lambda$ and $\delta_i \in \Delta$) and $\sum r_j \cdot \delta'_j \in \widetilde{\mathcal{R}}[\Delta]$ (with $r_j = \sum_h r_{j,h} \otimes \lambda_{j,h} \in \widetilde{\mathcal{R}}$, $r_{j,h} \in \mathcal{R}$, $\lambda_{j,h} \in \Lambda_\sigma$, and $\delta'_j \in \Delta$), we set

$$\Big(\sum_i \lambda_i \cdot \delta_i\Big) \cdot \Big(\sum_j r_j \cdot \delta'_j\Big) := \sum_{i,j,h} \big(r_{j,h} \otimes (\lambda_i \lambda_{j,h})\big) \cdot (\delta_i \delta'_j).$$

As above, extend $\tilde\kappa \in \widetilde{\mathcal{X}}^{\mathrm{arith}}$ to a continuous $\mathcal{O}$-algebra morphism

$$\tilde\kappa_N : \widetilde{\mathcal{R}}_N \longrightarrow \bar{\mathbb{Q}}_p$$

by setting

$$\tilde\kappa_N\left(\sum_{i=1}^n x_i \cdot \delta_i\right) := \sum_{i=1}^n \tilde\kappa(x_i) \cdot \psi_{\mathcal{R}}(\delta_i)$$

for $x_i \in \widetilde{\mathcal{R}}$ and $\delta_i \in \Delta$, where $\psi_{\mathcal{R}}$ is the character of $\mathcal{R}$. If we denote by $\widetilde{\mathcal{X}}_N$ the $\mathcal{O}$-module of continuous $\mathcal{O}$-linear homomorphisms from $\widetilde{\mathcal{R}}_N$ to $\bar{\mathbb{Q}}_p$, the above correspondence sets up an injective map $\widetilde{\mathcal{X}}^{\mathrm{arith}} \hookrightarrow \widetilde{\mathcal{X}}_N$ and we let $\widetilde{\mathcal{X}}_N^{\mathrm{arith}}$ denote the image of $\widetilde{\mathcal{X}}^{\mathrm{arith}}$. Put $\epsilon_{\tilde\kappa} := \psi_{\tilde\kappa} \cdot \psi_{\mathcal{R}} \cdot \omega^{-n_{\tilde\kappa}}$, which we view as a Dirichlet character of $(\mathbb{Z}/Np^{r_{\tilde\kappa}}\mathbb{Z})^\times$, and call the pair $(\epsilon_{\tilde\kappa}, k_{\tilde\kappa})$ the signature of $\tilde\kappa_N$, where $\tilde\kappa$ is the arithmetic point corresponding to $\tilde\kappa_N$.

We also have a map $p_N : \widetilde{\mathcal{X}}_N \to \mathcal{X}_N$ induced from the map $\mathcal{R}_N \to \widetilde{\mathcal{R}}_N$ taking $r \mapsto r \otimes 1$ by pull-back. The map p_N also restricts to arithmetic points. The maps p and p_N make the following diagram commute:

$$\begin{tikzcd}
\widetilde{\mathcal{X}}^{\mathrm{arith}} \arrow[r, hook] \arrow[d, "p"] & \widetilde{\mathcal{X}}_N^{\mathrm{arith}} \arrow[d, "p_N"] \\
\mathcal{X}^{\mathrm{arith}} \arrow[r, hook] & \mathcal{X}_N^{\mathrm{arith}}
\end{tikzcd}$$

where the projections take a signature (ϵ, k) to $(\epsilon^2, 2k)$.

3.4. The Λ-adic correspondence. In this part, we combine the explicit integral formula of Shimura and the fact that the toric integrals can be p-adically interpolated to show the existence of a Λ-adic Shimura-Shintani-Waldspurger correspondence with the expected interpolation property. This follows very closely [21, §6].

Let $\tilde\kappa_N \in \widetilde{\mathcal{X}}_N^{\mathrm{arith}}$ of signature $(\epsilon_{\tilde\kappa}, k_{\tilde\kappa})$. Let L_r denote the order of $\mathbb{M}_2(F)$ consisting of matrices $\left(\begin{smallmatrix} a & b/Mp^r \\ Mp^r c & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathcal{O}_F$. Define

$$\mathcal{O}_{B,r} := \iota^{-1}\big(i_F^{-1}\big(i_F(B \otimes 1) \cap L_r\big)\big)$$

Then $\mathcal{O}_{B,r}$ is the maximal order introduced in §2.1 (and denoted $\mathcal{O}'_B$ there) defined in terms of the maximal order $\mathcal{O}_B$ and the integer Mp^r. Also,

$$S := \mathcal{O}_B \cap \mathcal{O}_{B,r}$$

is an Eichler order of B of level Mp containing the fixed Eichler order R of level M. With $\alpha \in V^* \cap \mathcal{O}_{B,1}$, we have

$$i_F(\alpha) = \begin{pmatrix} a & b/(Mp) \\ c & -a \end{pmatrix} \tag{7}$$

in $\mathbb{M}_2(F)$ with $a, b, c \in \mathcal{O}_F$ and we can consider the quadratic forms

$$Q_\alpha(x,y) := cx^2 - 2axy - \big(b/(Mp)\big)y^2,$$

and

$$\tilde{Q}_\alpha(x,y) := Mp \cdot Q_\alpha(x,y) = Mpcx^2 - 2Mpaxy - by^2. \tag{8}$$

Then $\tilde{Q}_\alpha(x,y)$ has coefficients in $\mathcal{O}_F$ and, composing with $F \hookrightarrow \mathbb{R}$ and letting $x = z$, $y = 1$, we recover $Q_\alpha(z)$ and $\tilde{Q}_\alpha(z)$ of §2.1 (defined by means of the isomorphism i_∞). Since each prime $\ell | Mp$ is split in F, the elements a, b, c can be viewed as elements in $\mathbb{Z}_\ell$ via our fixed embedding $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_\ell$, for any prime $\ell | Mp$ (we will continue writing a, b, c for these elements, with a slight abuse of notation). So, letting $b_\alpha \in \mathbb{Z}$ such that $i_\ell(\alpha) \equiv \left(\begin{smallmatrix} * & b_\alpha/(Mp) \\ * & * \end{smallmatrix}\right)$ modulo $i_\ell(S \otimes_\mathbb{Z} \mathbb{Z}_\ell)$, for all $\ell | Mp$, we have $b \equiv b_\alpha$ modulo $Mp\mathbb{Z}_\ell$ as elements in $\mathbb{Z}_\ell$, for all $\ell | Mp$, and thus we get

$$\eta_{\epsilon_{\tilde{\kappa}}}(\alpha) = \epsilon_{\tilde{\kappa}}(b_\alpha) = \epsilon_{\tilde{\kappa}}(b) \tag{9}$$

for b as in (7).

For any $\nu \in \mathbb{D}$, we may define an $\mathcal{O}$-valued measure $j_\alpha(\nu)$ on $\mathbb{Z}_p^\times$ by the formula:

$$\int_{\mathbb{Z}_p^\times} f(t)dj_\alpha(\nu)(t) := \int_{\mathbb{Z}_p \times \mathbb{Z}_p^\times} f\big(\tilde{Q}_\alpha(x,y)\big)d\nu(x,y).$$

for any continuous function $f : \mathbb{Z}_p^\times \to \mathbb{C}_p$. Recall that the group of $\mathcal{O}$-valued measures on $\mathbb{Z}_p^\times$ is isomorphic to the Iwasawa algebra $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$, and thus we may view $j_\alpha(\nu)$ as an element in $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$ (see, for example, [1, §3.2]). In particular, for any group-like element $[\lambda] \in \mathcal{O}[\![\mathbb{Z}_p^\times]\!]$ we have:

$$\int_{\mathbb{Z}_p^\times} f(t)d\big([\lambda] \cdot j_\alpha(\nu)\big)(t) = \int_{\mathbb{Z}_p^\times} \left(\int_{\mathbb{Z}_p^\times} f(ts)d[\lambda](s)\right) dj_\alpha(\nu)(t) = \int_{\mathbb{Z}_p^\times} f(\lambda t)dj_\alpha(\nu)(t).$$

On the other hand,

$$\int_{\mathbb{Z}_p \times \mathbb{Z}_p^\times} f\big(\tilde{Q}_\alpha(x,y)\big)d(\lambda \cdot \nu) = \int_{\mathbb{Z}_p \times \mathbb{Z}_p^\times} f\big(\tilde{Q}_\alpha(\lambda x, \lambda y)\big)d\nu = \int_{\mathbb{Z}_p \times \mathbb{Z}_p^\times} f\big(\lambda^2 \tilde{Q}_\alpha(x,y)\big)d\nu$$

and we conclude that $j_\alpha(\lambda \cdot \nu) = [\lambda^2] \cdot j_\alpha(\nu)$. In other words, j_α is a $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$-linear map

$$j_\alpha : \mathbb{D} \longrightarrow \mathcal{O}[\![\mathbb{Z}_p^\times]\!]_\sigma.$$

Before going ahead, let us introduce some notation. Let χ be a Dirichlet character modulo Mp^r, for a positive integer r, which we decompose accordingly

with the isomorphism $(\mathbb{Z}/Np^r\mathbb{Z})^\times \simeq (\mathbb{Z}/N\mathbb{Z})^\times \times (\mathbb{Z}/p^r\mathbb{Z})^\times$ into the product $\chi = \chi_N \cdot \chi_p$ with $\chi_N : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ and $\chi_p : (\mathbb{Z}/p^r\mathbb{Z})^\times \to \mathbb{C}^\times$. Thus, we will write $\chi(x) = \chi_N(x_N) \cdot \chi_p(x_p)$, where x_N and x_p are the projections of $x \in (\mathbb{Z}/Np^r\mathbb{Z})^\times$ to $(\mathbb{Z}/N\mathbb{Z})^\times$ and $(\mathbb{Z}/p^r\mathbb{Z})^\times$, respectively. To simplify the notation, we will suppress the N and p from the notation for x_N and x_p, thus simply writing x for any of the two. Using the isomorphism $(\mathbb{Z}/N\mathbb{Z})^\times \simeq (\mathbb{Z}/M\mathbb{Z})^\times \times (\mathbb{Z}/D\mathbb{Z})^\times$, decompose χ_N as $\chi_N = \chi_M \cdot \chi_D$ with χ_M and χ_D characters on $(\mathbb{Z}/M\mathbb{Z})^\times$ and $(\mathbb{Z}/D\mathbb{Z})^\times$, respectively. In the following, we only need the case when $\chi_D = 1$.
Using the above notation, we may define a $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$-linear map $J_\alpha : \mathbb{D} \to \mathcal{O}[\![\mathbb{Z}_p^\times]\!]$ by

$$J_\alpha(\nu) = \epsilon_{\tilde\kappa,M}(b) \cdot \epsilon_{\tilde\kappa,p}(-1) \cdot j_\alpha(\nu)$$

with b as in (7). Set $\mathbb{D}_N := \mathbb{D} \otimes_{\mathcal{O}[\![\mathbb{Z}_p^\times]\!]} \Lambda_N$, where the map $\mathcal{O}[\![\mathbb{Z}_p^\times]\!] \to \Lambda_N$ is induced from the map $\mathbb{Z}_p^\times \to \mathbb{Z}_p^\times \times (\mathbb{Z}/N\mathbb{Z})^\times$ on group-like elements given by $x \mapsto x \otimes 1$. Then J_α can be extended to a Λ_N-linear map $J_\alpha : \mathbb{D}_N \to \Lambda_{N,\sigma}$. Setting $\mathbb{D}_{\mathcal{R}_N} := \mathcal{R}_N \otimes_{\Lambda_N} \mathbb{D}_N$ and extending by $\mathcal{R}_N$-linearity over Λ_N we finally obtain a $\mathcal{R}_N$-linear map, again denoted by the same symbol,

$$J_\alpha : \mathbb{D}_{\mathcal{R}_N} \longrightarrow \widetilde{\mathcal{R}}_N.$$

For $\nu \in \mathbb{D}_N$ and $r \in \mathcal{R}_N$ we thus have

$$J_\alpha(r \otimes \nu) = \epsilon_{\tilde\kappa,M}(b) \cdot \epsilon_{\tilde\kappa,p}(-1) \cdot r \otimes j_\alpha(\nu).$$

For the next result, for any arithmetic point $\kappa_N \in \mathcal{X}_N^{\mathrm{arith}}$ coming from $\kappa \in \mathcal{X}^{\mathrm{arith}}$, extend ρ_κ in (5) by $\mathcal{R}_N$-linearity over $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$, to get a map

$$\rho_{\kappa_N} : \mathbb{D}_{\mathcal{R}_N} \longrightarrow V_{n_\kappa}$$

defined by $\rho_{\kappa_N}(r \otimes \nu) := \rho_\kappa(\nu) \cdot \kappa_N(r)$, for $\nu \in \mathbb{D}$ and $r \in \mathcal{R}_N$. To simplify the notation, set

$$\langle \nu, \alpha\rangle_{\kappa_N} := \rho_{\kappa_N}(\nu)(\tilde{Q}_\alpha^{n_{\tilde\kappa}/2}). \tag{10}$$

The following is essentially [21, Lemma (6.1)].

Lemma 3.2. *Let $\tilde\kappa_N \in \widetilde{\mathcal{X}}_N^{\mathrm{arith}}$ with signature $(\epsilon_{\tilde\kappa}, k_{\tilde\kappa})$ and define $\kappa_N := p_N(\tilde\kappa_N)$. Then for any $\nu \in \mathbb{D}_{\mathcal{R}_N}$ we have:*

$$\tilde\kappa_N\big(J_\alpha(\nu)\big) = \eta_{\epsilon_{\tilde\kappa}}(\alpha) \cdot \langle \nu, \alpha\rangle_{\kappa_N}.$$

Proof. For $\nu \in \mathbb{D}_N$ and $r \in \mathcal{R}_N$ we have

$$\begin{aligned}\tilde\kappa_N\big(J_\alpha(r \otimes \nu)\big) &= \tilde\kappa_N\big(\epsilon_{\tilde\kappa,M}(b) \cdot \epsilon_{\tilde\kappa,p}(-1) \cdot r \otimes j_\alpha(\nu)\big) \\ &= \epsilon_{\tilde\kappa,M}(b) \cdot \epsilon_{\tilde\kappa,p}(-1) \cdot \tilde\kappa_N(r \otimes 1) \cdot \tilde\kappa_N\big(1 \otimes j_\alpha(\nu)\big) \\ &= \epsilon_{\tilde\kappa,M}(b) \cdot \epsilon_{\tilde\kappa,p}(-1) \cdot \kappa_N(r) \cdot \int_{\mathbb{Z}_p^\times} \tilde\kappa_N(t) dj_\alpha(\nu)\end{aligned}$$

and thus, noticing that $\tilde\kappa_N$ restricted to $\mathbb{Z}_p^\times$ is $\tilde\kappa_N(t) = \epsilon_{\tilde\kappa,p}(t)t^{n_{\tilde\kappa}}$, we have

$$\tilde\kappa_N\big(J_\alpha(r \otimes \nu)\big) = \epsilon_{\tilde\kappa,M}(b) \cdot \epsilon_{\tilde\kappa,p}(-1) \cdot \kappa_N(r) \int_{\mathbb{Z}_p \times \mathbb{Z}_p^\times} \epsilon_{\tilde\kappa,p}(\tilde{Q}_\alpha(x,y))\tilde{Q}_\alpha(x,y)^{n_{\tilde\kappa}/2} d\nu.$$

Recalling (8), and viewing a, b, c as elements in $\mathbb{Z}_p$, we have, for $(x,y) \in \mathbb{Z}_p \times \mathbb{Z}_p^\times$,

$$\epsilon_{\tilde\kappa,p}\big(\tilde Q_\alpha(x,y)\big) = \epsilon_{\tilde\kappa,p}(-by^2) = \epsilon_{\tilde\kappa,p}(-b)\epsilon_{\tilde\kappa,p}(y^2) = \epsilon_{\tilde\kappa,p}(-b)\epsilon_{\tilde\kappa,p}^2(y) = \epsilon_{\tilde\kappa,p}(-b)\epsilon_{\kappa,p}(y).$$

Thus, since $\epsilon_{\tilde\kappa}(-1)^2 = 1$, we get:

$$\tilde\kappa_N\big(J_\alpha(r \otimes \nu)\big) = \kappa_N(r) \cdot \epsilon_{\tilde\kappa,M}(b) \cdot \epsilon_{\tilde\kappa,p}(b) \cdot \rho_\kappa(\nu)(\tilde Q_\alpha^{n_{\tilde\kappa}/2}) = \eta_{\epsilon_\kappa}(\alpha) \cdot \langle \nu, \alpha\rangle_{\kappa_N}$$

where for the last equality use (9) and (10). □

Define

$$\mathbb{W}_{\mathcal{R}_N} := \mathbb{W} \otimes_{\mathcal{O}[\![\mathbb{Z}_p^\times]\!]} \mathcal{R}_N,$$

the structure of $\mathcal{O}[\![\mathbb{Z}_p^\times]\!]$-module of $\mathcal{R}_N$ being that induced by the composition of the two maps $\mathcal{O}[\![\mathbb{Z}_p^\times]\!] \to \Lambda_N \to \mathcal{R}_N$ described above. There is a canonical map

$$\vartheta : \mathbb{W}_{\mathcal{R}_N} \longrightarrow H^1(\Gamma_0, \mathbb{D}_{\mathcal{R}_N})$$

described as follows: if ν_γ is a representative of an element ν in $\mathbb{W}$ and $r \in \mathcal{R}_N$, then $\vartheta(\nu \otimes r)$ is represented by the cocycle $\nu_\gamma \otimes r$.
For $\nu \in \mathbb{W}_{\mathcal{R}_N}$ represented by ν_γ and $\xi \geq 1$ an integer, define

$$\theta_\xi(\nu) := \sum_{\mathcal{C} \in R(\Gamma_1), q(\mathcal{C}) = \xi} \frac{J_{\alpha_\mathcal{C}}(\nu_{\gamma_{\alpha_\mathcal{C}}})}{\mathfrak{t}_{\alpha_\mathcal{C}}}.$$

Definition 3.3. For $\nu \in \mathbb{W}_{\mathcal{R}_N}$, the formal Fourier expansion

$$\Theta(\nu) := \sum_{\xi \geq 1} \theta_\xi(\nu) q^\xi$$

in $\mathcal{R}_N[\![q]\!]$ is called the Λ-adic Shimura-Shintani-Waldspurger lift of ν. For any $\tilde\kappa \in \widetilde{\mathcal{X}}^{\mathrm{arith}}$, the formal power series expansion

$$\Theta(\nu)(\tilde\kappa_N) := \sum_{\xi \geq 1} \tilde\kappa_N\big(\theta_\xi(\nu)\big) q^\xi$$

is called the $\tilde\kappa$-specialization of $\Theta(\nu)$.

There is a natural map

$$\mathbb{W}_\mathcal{R} \longrightarrow \mathbb{W}_{\mathcal{R}_N}$$

taking $\nu \otimes r$ to itself (use that $\mathcal{R}$ has a canonical map to $\mathcal{R}_N \simeq \mathcal{R}[\Delta]$, as described above). So, for any choice of sign, $\Phi^\pm \in \mathbb{W}_\mathcal{R}$ will be viewed as an element in $\mathbb{W}_{\mathcal{R}_N}$.
¿From now on we will use the following notation. Fix $\tilde\kappa_0 \in \widetilde{\mathcal{X}}^{\mathrm{arith}}$ and put $\kappa_0 := p(\tilde\kappa_0) \in \mathcal{X}^{\mathrm{arith}}$. Recall the neighborhood $\mathcal{U}_0$ of κ_0 in Theorem 3.1. Define $\widetilde{\mathcal{U}}_0 := p^{-1}(\mathcal{U}_0)$ and

$$\widetilde{\mathcal{U}}_0^{\mathrm{arith}} := \widetilde{\mathcal{U}}_0 \cap \widetilde{\mathcal{X}}^{\mathrm{arith}}.$$

For each $\tilde\kappa \in \widetilde{\mathcal{U}}_0^{\mathrm{arith}}$ put $\kappa = p(\tilde\kappa) \in \mathcal{U}_0^{\mathrm{arith}}$. Recall that if $(\epsilon_{\tilde\kappa}, k_{\tilde\kappa})$ is the signature of $\tilde\kappa$, then $(\epsilon_\kappa, k_\kappa) := (\epsilon_{\tilde\kappa}^2, 2k_{\tilde\kappa})$ is that of κ_0. For any $\kappa := p(\tilde\kappa)$ as above, we may consider the modular form

$$f_\kappa^{\mathrm{JL}} \in S_{k_\kappa}(\Gamma_{r_\kappa}, \epsilon_\kappa)$$

and its Shimura-Shintani-Waldspurger lift

$$h_\kappa = \sum_\xi a_\xi(h_\kappa)q^\xi \in S_{k_\kappa+1/2}(4Np^{r_\kappa},\chi_\kappa), \quad \text{where } \chi_\kappa(x) := \epsilon_{\tilde\kappa}(x)\left(\frac{-1}{x}\right)^{k_\kappa},$$

normalized as in (2) and (3). For our fixed κ_0, recall the elements $\Phi := \Phi^+$ chosen as in Theorem 3.1 and define $\phi_\kappa := \phi_\kappa^+$ and $\Omega_\kappa := \Omega_\kappa^+$ for $\kappa \in \mathcal{U}_0^{\mathrm{arith}}$.

Proposition 3.4. *For all $\tilde\kappa \in \widetilde{\mathcal{U}}_0^{\mathrm{arith}}$ such that $r_\kappa = 1$ we have*

$$\tilde\kappa_N\big(\theta_\xi(\Phi)\big) = \Omega_\kappa \cdot a_\xi(h_\kappa) \quad \text{and} \quad \Theta(\Phi)(\tilde\kappa_N) = \Omega_\kappa \cdot h_\kappa.$$

Proof. By Lemma 3.2 we have

$$\tilde\kappa_N\big(\theta_\xi(\Phi)\big) = \sum_{\mathcal{C}\in R(\Gamma_1),q(\mathcal{C})=\xi} \frac{\eta_{\epsilon_{\tilde\kappa}}(\alpha_{\mathcal{C}})}{\mathfrak{t}_{\alpha_{\mathcal{C}}}}\rho_{\kappa_N}(\Phi)(\tilde Q_{\alpha_{\mathcal{C}}}^{n_{\tilde\kappa}/2}).$$

Using Theorem 3.1, we get

$$\tilde\kappa_N\big(\theta_\xi(\Phi)\big) = \sum_{\mathcal{C}\in R(\Gamma_1),q(\mathcal{C})=\xi} \frac{\eta_{\epsilon_{\tilde\kappa}}(\alpha_{\mathcal{C}})\cdot\Omega_\kappa}{\mathfrak{t}_{\alpha_{\mathcal{C}}}}\phi_\kappa(\tilde Q_{\alpha_{\mathcal{C}}}^{k_\kappa-1}).$$

Now (2) shows the statement on $\tilde\kappa_N(\theta_\xi(\Phi))$, while that on $\Theta(\Phi)(\tilde\kappa_N)$ is a formal consequence of the previous one. □

Corollary 3.5. *Let a_p denote the image of the Hecke operator T_p in $\mathcal{R}$. Then $\Theta(\Phi)|T_p^2 = a_p \cdot \Theta(\Phi)$.*

Proof. For any $\kappa \in \mathcal{X}^{\mathrm{arith}}$, let $a_p(\kappa) := \kappa(T_p)$, which is a p-adic unit by the ordinarity assumption. For all $\tilde\kappa \in \widetilde{\mathcal{U}}_0^{\mathrm{arith}}$ with $r_\kappa = 1$, we have

$$\Theta(\Phi)(\tilde\kappa_N)|T_p^2 = \Omega_\kappa \cdot h_\kappa|T_p^2 = a_p(\kappa)\cdot\Omega_\kappa\cdot h_\kappa = a_p(\kappa)\cdot\Theta(\Phi)(\tilde\kappa_N).$$

Consequently,

$$\tilde\kappa_N\big(\theta_{\xi p^2}(\Phi)\big) = a_p(\kappa)\cdot\tilde\kappa_N\big(\theta_\xi(\Phi)\big)$$

for all $\tilde\kappa$ such that $r_\kappa = 1$. Since this subset is dense in $\widetilde{\mathcal{X}}_N$, we conclude that $\theta_{\xi p^2}(\Phi) = a_p \cdot \theta_\xi(\Phi)$ and so $\Theta(\Phi)|T_p^2 = a_p\cdot\Theta(\Phi)$. □

For any integer $n \geq 1$ and any quadratic form Q with coefficients in F, write $[Q]_n$ for the class of Q modulo the action of $i_F(\Gamma_n)$. Define $\mathcal{F}_{n,\xi}$ to be the subset of the F-vector space of quadratic forms with coefficients in F consisting of quadratic forms $\tilde Q_\alpha$ such that $\alpha \in V^* \cap \mathcal{O}_{B,n}$ and $-\mathrm{nr}(\alpha) = \xi$. Writing $\delta_{\tilde Q_\alpha}$ for the discriminant of Q_α, the above set can be equivalently described as

$$\mathcal{F}_{n,\xi} := \{\tilde Q_\alpha |\, \alpha \in V^*\cap\mathcal{O}_{B,n},\, \delta_{\tilde Q_\alpha} = Np^n\xi\}.$$

Define $\mathcal{F}_{n,\xi}/\Gamma_n$ to be the set $\{[\tilde Q_\alpha]_n |\, \tilde Q_\alpha \in \mathcal{F}_{n,\xi}\}$ of equivalence classes of $\mathcal{F}_{n,\xi}$ under the action of $i_F(\Gamma_n)$. A simple computation shows that $Q_{g^{-1}\alpha g} = Q_\alpha|g$ for all $\alpha \in V^*$ and all $g \in \Gamma_n$, and thus we find

$$\mathcal{F}_{n,\xi}/\Gamma_n = \{[\tilde Q_{\mathcal{C}_\alpha}]_n |\, \mathcal{C}\in R(\Gamma_n),\, \delta_{\tilde Q_\alpha} = Np^n\xi\}.$$

We also note that, in the notation of §2.1, if f has weight character ψ, defined modulo Np^n, and level Γ_n, the Fourier coefficients $a_\xi(h)$ of the Shimura-Shintani-Waldspurger lift h of f are given by

$$a_\xi(h) = \sum_{[Q]\in\mathcal{F}_{n,\xi}/\Gamma_n} \frac{\psi(Q)}{\mathfrak{t}_Q}\phi_f^+\big(Q(z)^{k-1}\big) \tag{11}$$

and, if $Q = \tilde{Q}_\alpha$, we put $\psi(Q) := \eta_\psi(b_\alpha)$ and $\mathfrak{t}_Q := \mathfrak{t}_\alpha$. Also, if we let

$$\mathcal{F}_n/\Gamma_n := \coprod_\xi \mathcal{F}_{n,\xi}/\Gamma_n$$

we can write

$$h = \sum_{[Q]\in\mathcal{F}_n/\Gamma_n} \frac{\psi(Q)}{\mathfrak{t}_Q}\phi_f^+\big(Q(z)^{k-1}\big)q^{\delta_Q/(Np^n)}. \tag{12}$$

Fix now an integer $m \geq 1$ and let $n \in \{1, m\}$. For any $t \in (\mathbb{Z}/p^n\mathbb{Z})^\times$ and any integer $\xi \geq 1$, define $\mathcal{F}_{n,\xi,t}$ to be the subset of $\mathcal{F}_{n,\xi}$ consisting of forms such that $Np^n b_\alpha \equiv t \mod Np^m$. Also, define $\mathcal{F}_{n,\xi,t}/\Gamma_n$ to be the set of equivalence classes of $\mathcal{F}_{n,\xi,t}$ under the action of $i_F(\Gamma_n)$. If $\alpha \in V^* \cap \mathcal{O}_{B,m}$ and

$$i_F(\alpha) = \begin{pmatrix} a & b \\ c & -a \end{pmatrix},$$

then

$$\tilde{Q}_\alpha(x,y) = Np^n cx^2 - 2Np^n axy - Np^n by^2 \tag{13}$$

from which we see that there is an inclusion $\mathcal{F}_{m,\xi,t} \subseteq \mathcal{F}_{1,\xi p^{m-1},t}$. If $\tilde{Q}_\alpha$ and $\tilde{Q}_{\alpha'}$ belong to $\mathcal{F}_{m,\xi,t}$, and $\alpha' = g\alpha g^{-1}$ for some $g \in \Gamma_m$, then, since $\Gamma_m \subseteq \Gamma_1$, we see that $\tilde{Q}_\alpha$ and $\tilde{Q}_{\alpha'}$ represent the same class in $\mathcal{F}_{1,\xi p^{m-1},t}/\Gamma_1$. This shows that $[\tilde{Q}_\alpha]_m \mapsto [\tilde{Q}_\alpha]_1$ gives a well-defined map

$$\pi_{m,\xi,t} : \mathcal{F}_{m,\xi,t}/\Gamma_m \longrightarrow \mathcal{F}_{1,\xi p^{m-1},t}/\Gamma_1.$$

LEMMA 3.6. *The map $\pi_{m,\xi,t}$ is bijective.*

Proof. We first show the injectivity. For this, suppose $\tilde{Q}_\alpha$ and $\tilde{Q}_{\alpha'}$ are in $\mathcal{F}_{m,\xi,t}$ and $[\tilde{Q}_\alpha]_1 = [\tilde{Q}_{\alpha'}]_1$. So there exists $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ in $i_F(\Gamma_1)$ such that such that $\tilde{Q}_\alpha = \tilde{Q}_{\alpha'}|g$. If $\tilde{Q}_\alpha = cx^2 - 2axy - by^2$, and easy computation shows that $\tilde{Q}_{\alpha'} = c'x^2 - 2a'xy - b'y^2$ with

$$c' = c\alpha^2 - 2a\alpha\gamma - b\gamma^2$$
$$a' = -c\alpha\beta + a\beta\gamma + a\alpha\delta + b\gamma\delta$$
$$b' = -c\beta^2 + 2a\beta\delta + b\delta^2.$$

The first condition shows that $\gamma \equiv 0 \mod Np^m$. We have $b \equiv b' \equiv t \mod Np^m$, so $\delta^2 \equiv 1 \mod Np^m$. Since $\delta \equiv 1 \mod Np$, we see that $\delta \equiv 1 \mod Np^m$ too.

We now show the surjectivity. For this, fix $[\tilde{Q}_{\alpha_{\mathcal{C}}}]_1$ in the target of π, and choose a representative

$$\tilde{Q}_{\alpha_{\mathcal{C}}} = cx^2 - 2axy - by^2$$

(recall $Mp^m\xi|\delta_{\tilde{Q}_{\alpha_{\mathcal{C}}}}$, $Np|c$, $Np|a$, and $b \in \mathcal{O}_F^\times$, the last condition due to $\eta_\psi(\alpha_{\mathcal{C}}) \neq 0$). By the Strong Approximation Theorem, we can find $\tilde{g} \in \Gamma_1$ such that

$$i_\ell(\tilde{g}) \equiv \begin{pmatrix} 1 & 0 \\ ab^{-1} & 1 \end{pmatrix} \mod Np^m$$

for all $\ell|Np$. Take $g := i_F(\tilde{g})$, and put $\alpha := g^{-1}\alpha_{\mathcal{C}}g$. An easy computation, using the expressions for a', b', c' in terms of a, b, c and $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ as above, shows that $\alpha \in V^* \cap \mathcal{O}_{B,m}$, $\eta_\psi(\alpha) = t$ and $\delta_{\tilde{Q}_\alpha} = Np^m\xi$, and it follows that $\tilde{Q}_\alpha \in \mathcal{F}_{m,\xi,t}$. Now

$$\pi\big([\tilde{Q}_\alpha]_m\big) = [\tilde{Q}_\alpha]_1 = [\tilde{Q}_{g^{-1}\alpha_{\mathcal{C}}g}]_1 = [\tilde{Q}_{\alpha_{\mathcal{C}}}]_1$$

where the last equality follows because $g \in \Gamma_1$. □

Proposition 3.7. *For all $\tilde{\kappa} \in \widetilde{\mathcal{U}}_0^{\mathrm{arith}}$ we have*

$$\Theta(\Phi)(\tilde{\kappa}_N)|T_p^{r_\kappa - 1} = \Omega_\kappa \cdot h_\kappa.$$

Proof. For $r_\kappa = 1$, this is Proposition 3.4 above, so we may assume $r_\kappa \geq 2$. As in the proof of Proposition 3.4, combining Lemma 3.2 and Theorem 3.1 we get

$$\Theta(\Phi)(\tilde{\kappa}_N) = \sum_{\xi \geq 1} \left(\sum_{\mathcal{C} \in R(\Gamma_1), q(\mathcal{C}) = \xi} \frac{\eta_{\epsilon_{\tilde{\kappa}}}(\alpha_{\mathcal{C}}) \cdot \Omega_\kappa}{\mathsf{t}_{\alpha_{\mathcal{C}}}} \phi_\kappa(\tilde{Q}_{\alpha_{\mathcal{C}}}^{k_\kappa - 1}) \right) q^\xi$$

which, by (11) and (12) above we may rewrite as

$$\Theta(\Phi)(\tilde{\kappa}_N) = \sum_{[Q] \in \mathcal{F}_1/\Gamma_1} \frac{\epsilon_{\tilde{\kappa}}(Q) \cdot \Omega_\kappa}{\mathsf{t}_Q} \phi_\kappa(Q^{k_\kappa - 1}) q^{\delta_Q/(Np)}$$

By definition of the action of T_p on power series, we have

$$\Theta(\Phi)(\tilde{\kappa}_N)|T_p^{r_\kappa - 1} = \sum_{[Q] \in \mathcal{F}_1/\Gamma_1,\, p^{r_\kappa}|\delta_Q} \frac{\epsilon_{\tilde{\kappa}}(Q) \cdot \Omega_\kappa}{\mathsf{t}_Q} \phi_\kappa(Q^{k_\kappa - 1}) q^{\delta_Q/(Np^{r_\kappa})}.$$

Setting $\mathcal{F}_{n,t}/\Gamma_n := \coprod_{\xi \geq 1} \mathcal{F}_{n,t,\xi}/\Gamma_n$ for $n \in \{1, r_\kappa\}$, Lemma 3.6 shows that

$$\mathcal{F}_{1,t}^* := \{[Q] \in \mathcal{F}_{1,t}/\Gamma_{1,t} \text{ such that } p^{r_\kappa}|\delta_Q\} \text{ is equal to } \mathcal{F}_{r_\kappa,t}.$$

Therefore, splitting the above sum over $t \in (\mathbb{Z}/Np^{r_\kappa}\mathbb{Z})^\times$, we get

$$\begin{aligned}
\Theta(\Phi)(\tilde{\kappa}_N)|T_p^{r_\kappa - 1} &= \sum_{t \in (\mathbb{Z}/p^{r_\kappa - 1}\mathbb{Z})^\times} \sum_{[Q] \in \mathcal{F}_{1,t}^*} \frac{\epsilon_{\tilde{\kappa}}(Q) \cdot \Omega_\kappa}{\mathsf{t}_Q} \phi_\kappa(Q^{k_\kappa - 1}) q^{\delta_Q/(Np^{r_\kappa})} \\
&= \sum_{t \in (\mathbb{Z}/p^{r_\kappa - 1}\mathbb{Z})^\times} \sum_{[Q] \in \mathcal{F}_{m,t}/\Gamma_m} \frac{\epsilon_{\tilde{\kappa}}(Q) \cdot \Omega_\kappa}{\mathsf{t}_Q} \phi_\kappa(Q^{k_\kappa - 1}) q^{\delta_Q/(Np^{r_\kappa})} \\
&= \sum_{[Q] \in \mathcal{F}_m/\Gamma_m} \frac{\epsilon_{\tilde{\kappa}}(Q) \cdot \Omega_\kappa}{\mathsf{t}_Q} \phi_\kappa(Q^{k_\kappa - 1}) q^{\delta_Q/(Np^{r_\kappa})}.
\end{aligned}$$

Comparing this expression with (12) gives the result. □

We are now ready to state the analogue of [21, Thm. 3.3], which is our main result. For the reader's convenience, we briefly recall the notation appearing below. We denote by $\mathcal{X}$ the points of the ordinary Hida Hecke algebra, and by $\mathcal{X}^{\mathrm{arith}}$ its arithmetic points. For $\kappa_0 \in \mathcal{X}^{\mathrm{arith}}$, we denote by $\mathcal{U}_0$ the p-adic neighborhood of κ_0 appearing in the statement of Theorem 3.1 and put $\mathcal{U}_0^{\mathrm{arith}} := \mathcal{U}_0 \cap \mathcal{X}^{\mathrm{arith}}$. We also denote by $\Phi = \Phi^+ \in \mathbb{W}_{\mathcal{R}}^{\mathrm{ord}}$ the cohomology class appearing in Theorem 3.1. The points $\widetilde{\mathcal{X}}$ of the metaplectic Hida Hecke algebra defined in §3.3 are equipped with a canonical map $p : \widetilde{\mathcal{X}}^{\mathrm{arith}} \to \mathcal{X}^{\mathrm{arith}}$ on arithmetic points. Let $\widetilde{\mathcal{U}}_0^{\mathrm{arith}} := \widetilde{\mathcal{U}}_0 \cap \widetilde{\mathcal{X}}^{\mathrm{arith}}$. For each $\tilde\kappa \in \widetilde{\mathcal{U}}_0^{\mathrm{arith}}$, put $\kappa = p(\tilde\kappa) \in \mathcal{U}_0^{\mathrm{arith}}$. Recall that if $(\epsilon_{\tilde\kappa}, k_{\tilde\kappa})$ is the signature of $\tilde\kappa$, then the signature of κ is $(\epsilon_\kappa, k_\kappa) := (\epsilon_{\tilde\kappa}^2, 2k_{\tilde\kappa})$. For any $\kappa := p(\tilde\kappa)$ as above, we may consider the modular form

$$f_\kappa^{\mathrm{JL}} \in S_{k_\kappa}(\Gamma_{r_\kappa}, \epsilon_\kappa)$$

and its Shimura-Shintani-Waldspurger lift

$$h_\kappa = \sum_\xi a_\xi(h_\kappa) q^\xi \in S_{k_\kappa+1/2}(4Np^{r_\kappa}, \chi_\kappa), \quad \text{where } \chi_\kappa(x) := \epsilon_{\tilde\kappa}(x) \left(\frac{-1}{x}\right)^{k_\kappa},$$

normalized as in (2) and (3). Finally, for $\tilde\kappa \in \widetilde{\mathcal{X}}^{\mathrm{arith}}$, we denote by $\tilde\kappa_N$ its extension to the metaplectic Hecke algebra $\widetilde{\mathcal{R}}_N$ defined in §3.3.

Theorem 3.8. *Let $\kappa_0 \in \mathcal{X}^{\mathrm{arith}}$. Then there exists a choice of p-adic periods Ω_κ for $\kappa \in \mathcal{U}_0$ such that the Λ-adic Shimura-Shintani-Waldspurger lift of Φ*

$$\Theta(\Phi) := \sum_{\xi \geq 1} \theta_\xi(\Phi) q^\xi$$

in $\mathcal{R}_N[\![q]\!]$ has the following properties:

(1) $\Omega_{\kappa_0} \neq 0$.

(2) *For any $\tilde\kappa \in \widetilde{\mathcal{U}}_0^{\mathrm{arith}}$, the $\tilde\kappa$-specialization of $\Theta(\Phi)$*

$$\Theta(\nu)(\tilde\kappa_N) := \sum_{\xi \geq 1} \tilde\kappa\big(\theta_\xi(\Phi)\big) q^\xi \text{ belongs to } S_{k_\kappa+1/2}(4Np^{r_\kappa}, \chi'_\kappa),$$

where $\chi'_\kappa(x) := \chi_\kappa(x) \cdot \left(\frac{p}{x}\right)^{k_\kappa - 1}$, and satisfies

$$\Theta(\Phi)(\tilde\kappa_N) = \Omega_\kappa \cdot h_\kappa | T_p^{1-r_\kappa}.$$

Proof. The elements Ω_κ are those Ω_κ^+ appearing in Theorem 3.1, which we used in Propositions 3.4 and 3.7 above, so (1) is clear. Applying $T_p^{r_\kappa - 1}$ to the formula of Proposition 3.7, using Corollary 3.5 and applying $a_p(\kappa)^{1-r_\kappa}$ on both sides gives

$$\Theta(\Phi)(\tilde\kappa_N) = a_p(\kappa)^{1-r_\kappa} \Omega_\kappa \cdot h_\kappa | T_p^{r_\kappa - 1}.$$

By [18, Prop. 1.9], each application of T_p has the effect of multiplying the character by $\left(\frac{p}{\cdot}\right)$, hence

$$h'_\kappa := h_\kappa | T_p^{r_\kappa - 1} \in S_{k_\kappa+1/2}(4Np^{r_\kappa}, \chi'_\kappa)$$

with χ'_κ as in the statement. This gives the first part of (2), while the last formula follows immediately from Proposition 3.7. □

Remark 3.9. Theorem 1.1 is a direct consequence of Theorem 3.8, as we briefly show below.

Recall the embedding $\mathbb{Z}^{\geq 2} \hookrightarrow \mathrm{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times)$ which sends $k \in \mathbb{Z}^{\geq 2}$ to the character $x \mapsto x^{k-2}$. Extending characters by $\mathcal{O}$-linearity gives a map

$$\mathbb{Z}^{\geq 2} \hookrightarrow \mathcal{X}(\Lambda) := \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}\text{-alg}}(\Lambda, \bar{\mathbb{Q}}_p).$$

We denote by $k^{(\Lambda)}$ the image of $k \in \mathbb{Z}^{\geq 2}$ in $\mathcal{X}(\Lambda)$ via this embedding. We also denote by $\varpi : \mathcal{X} \to \mathcal{X}(\Lambda)$ the finite-to-one map obtained by restriction of homomorphisms to Λ. Let $k^{(\mathcal{R})}$ be a point in $\mathcal{X}$ of signature $(k, 1)$ such that $\varpi(k^{(\mathcal{R})}) = k^{(\Lambda)}$. A well-known result by Hida (see [6, Cor. 1.4]) shows that $\mathcal{R}/\Lambda$ is unramified at $k^{(\Lambda)}$. As shown in [21, §1], this implies that there is a section $s_{k^{(\Lambda)}}$ of ϖ which is defined on a neighborhood $\mathcal{U}_{k^{(\Lambda)}}$ of $k^{(\Lambda)}$ in $\mathcal{X}(\Lambda)$ and sends $k^{(\Lambda)}$ to $k^{(\mathcal{R})}$.

Fix now k_0 as in the statement of Theorem 1.1, corresponding to a cuspform f_0 of weight k_0 with trivial character. The form f_0 corresponds to an arithmetic character $k_0^{(\mathcal{R})}$ of signature $(1, k_0)$ belonging to $\mathcal{X}$. Let $\mathcal{U}'_0$ be the inverse image of $\mathcal{U}_0$ under the section $s^{-1}_{k_0^{(\Lambda)}}$, where $\mathcal{U}_0 \subseteq \mathcal{X}$ is the neighborhood of $k_0^{(\mathcal{R})}$ in Theorem 3.8. Extending scalars by $\mathcal{O}$ gives, as above, an injective continuos map $\mathrm{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times) \hookrightarrow \mathcal{X}(\Lambda)$, and we let U_0 be any neighborhood of the character $x \mapsto x^{k_0-2}$ which maps to $\mathcal{U}'_0$ and is contained in the residue class of k_0 modulo $p-1$. Composing this map with the section $\mathcal{U}'_0 \hookrightarrow \mathcal{U}_0$ gives a continuous injective map

$$\varsigma : U_0 \hookrightarrow \mathcal{U}'_0 \hookrightarrow \mathcal{U}_0$$

which takes k_0 to $k_0^{(\mathcal{R})}$, since by construction the image of k_0 by the first map is $k_0^{(\Lambda)}$. We also note that, more generally, $\varsigma(k) = k^{(\mathcal{R})}$ because by construction $\varsigma(k)$ restricts to $k^{(\Lambda)}$ and its signature is $(1, k)$, since the character of $\varsigma(k)$ is trivial. To show the last assertion, recall that the character of $\varsigma(k)$ is $\psi_k \cdot \psi_{\mathcal{R}} \cdot \omega^{-k}$, and note that ψ_k is trivial because $k^{(\Lambda)}(x) = x^{k-1}$, and $\psi_{\mathcal{R}} \cdot \omega^{-k}$ is trivial because the same is true for k_0 and $k \equiv k_0$ modulo $p-1$. In other words, arithmetic points in $\varsigma(U_0)$ correspond to cuspforms with trivial character. This is the Hida family of forms with trivial character that we considered in the Introduction.

We can now prove Theorem 1.1. For all $k \in U_0 \cap \mathbb{Z}^{\geq 2}$, put $\Omega_k := \Omega_{k^{(\mathcal{R})}}$ and define $\Theta := \Theta(\Phi) \circ \varsigma$ with Φ as in Theorem 3.8 for $\kappa_0 = k_0^{(\mathcal{R})}$. Applying Theorem 3.8 to $k_0^{(\mathcal{R})}$, and restricting to $\varsigma(U_0)$, shows that U_0, Ω_k and Θ satisfy the conclusion of Theorem 1.1.

Remark 3.10. For $\tilde\kappa \in \widetilde{\mathcal{U}}_0^{\mathrm{arith}}$ of signature $(\epsilon_{\tilde\kappa}, k_{\tilde\kappa})$ with $r_{\tilde\kappa} = 1$ as in the above theorem, h_κ is trivial if $(-1)^{k_{\tilde\kappa}} \neq \epsilon_{\tilde\kappa}(-1)$. However, since $\phi_{\kappa_0} \neq 0$, it follows that h_{κ_0} is not trivial as long as the necessary condition $(-1)^{k_0} = \epsilon_0(-1)$ is verified.

Remark 3.11. This result can be used to produce a quaternionic Λ-adic version of the Saito-Kurokawa lifting, following closely the arguments in [8, Cor. 1].

References

[1] J. Coates, R. Sujatha, *Cyclotomic fields and zeta values.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2006.

[2] H. Darmon, G. Tornaria, Stark-Heegner points and the Shimura correspondence. *Compositio Math.*, 144 (2008) 1155-1175.

[3] R. Greenberg, G. Stevens, p-adic L-functions and p-adic periods of modular forms. *Invent. Math.* 111 (1993), no. 2, 407–447.

[4] Koblitz, N., *Introduction to elliptic curves and modular forms.* Graduate Texts in Mathematics, 97. Springer-Verlag, New York, 1984. viii+248 pp.

[5] W. Kohnen, Fourier coefficients of modular forms of half-integral weight. *Math. Ann.* 271 (1985), no. 2, 237–268.

[6] Hida H., Galois representations into $\mathrm{GL}_2(\mathbb{Z}_p[[X]])$ attached to ordinary cusp forms. *Invent. Math.* 85 (1986), no. 3, 545–613.

[7] Hida H., On Λ-adic forms of half-integral weight for $\mathrm{SL}(2)/\mathbb{Q}$. Number Theory (Paris 1992-3). Lond. Math. Soc. Lect. Note Ser.

[8] M. Longo, M.-H. Nicole, The Saito-Kurokawa lifting and Darmon points, to appear in Math. Ann. DOI 10.1007/s00208-012-0846-5

[9] M. Longo, S. Vigni, A note on control theorems for quaternionic Hida families of modular forms, Int. J. Number Theory, 2012 (to appear).

[10] M. Longo, S. Vigni, The rationality of quaternionic Darmon points over genus fields of real quadratic fields, preprint 2011.

[11] J. Nekovář, A. Plater, On the parity of ranks of Selmer groups. *Asian J. Math.* 4 (2000), no. 2, 437–497.

[12] J. Park, p-adic family of half-integral weight modular forms via overconvergent Shintani lifting *Manuscripta Mathematica*, Volume 131, 3-4, 2010, 355-384.

[13] A. Popa, Central values of Rankin L-series over real quadratic fields. *Compos. Math.* 142 (2006), no. 4, 811–866.

[14] K. Prasanna, Integrality of a ratio of Petersson norms and level-lowering congruences. *Ann. of Math.* (2) 163 (2006), no. 3, 901–967.

[15] K. Prasanna, Arithmetic properties of the Shimura-Shintani-Waldspurger correspondence. With an appendix by Brian Conrad. *Invent. Math.* 176 (2009), no. 3, 521–600.

[16] K. Prasanna, On the Fourier coefficients of modular forms of half-integral weight. *Forum Math.* 22 (2010), no. 1, 153–177.

[17] Ramsey, N., The overconvergent Shimura lifting, *Int. Math. Res. Not.*, 2009, no. 2, p. 193-220.

[18] G. Shimura, On modular forms of half integral weight. *Ann. of Math.* (2) 97 (1973), 440–481.

[19] G. Shimura, The periods of certain automorphic forms of arithmetic type. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* 28 (1981), no. 3, 605-632 (1982).

[20] T. Shintani, On construction of holomorphic cusp forms of half integral weight. *Nagoya Math. J.* 58 (1975), 83–126.

[21] G. Stevens, Λ-adic modular forms of half-integral weight and a Λ-adic Shintani lifting. Arithmetic geometry (Tempe, AZ, 1993), 129–151, Contemp. Math., 174, Amer. Math. Soc., Providence, RI, 1994.

[22] J.-L. Waldspurger, Correspondances de Shimura et quaternions. *Forum Math.* 3 (1991), no. 3, 219–307.

Matteo Longo
Dipartimento di Matematica
Università di Padova
Via Trieste 63
35121 Padova
Italy
mlongo@math.unipd.it

Marc-Hubert Nicole
Institut de mathématiques
de Luminy
Université d'Aix-Marseille
campus de Luminy, case 907
13288 Marseille cedex 9
France
nicole@iml.univ-mrs.fr

On a Positive Equicharacteristic Variant of the p-Curvature Conjecture

Hélène Esnault, Adrian Langer

Received: October 26, 2012

Communicated by Takeshi Saito

Abstract. Our aim is to formulate and prove a weak form in equal characteristic $p > 0$ of the p-curvature conjecture. We also show the existence of a counterexample to a strong form of it.

2010 Mathematics Subject Classification: 14D05, 14E20, 14F35, 11G10, 11G99
Keywords and Phrases: varieties in positive characteristic, stratified bundles, étale trivializable bundles, monodromy group, abelian varieties

Introduction

If (E,∇) is a vector bundle with an algebraic integrable connection over a smooth complex variety X, then it is defined over a smooth scheme S over $\mathrm{Spec}\,\mathbb{Z}[\frac{1}{N}]$ for some positive integer N, so $(E,\nabla) = (E_S,\nabla_S)\otimes_S \mathbb{C}$ over $X = X_S \otimes_S \mathbb{C}$ for a geometric generic point $\mathbb{Q}(S) \subset \mathbb{C}$. Grothendieck-Katz's p-curvature conjecture predicts that if for all closed points s of some non-trivial open $U \subset S$, the p-curvature of $(E_S,\nabla_S)\times_S s$ is zero, then (E,∇) is trivialized by a finite étale cover of X (see e.g. [An, Conj.3.3.3]). Little is known about it. N. Katz proved it for Gauß-Manin connections [Ka], for S finite over $\mathrm{Spec}\,\mathbb{Z}[\frac{1}{N}]$ (i.e., if X can be defined over a number field), D. V. Chudnovsky and G. V. Chudnovsky in [CC] proved it in the rank 1 case and Y. André in [An] proved it in case the Galois differential Lie algebra of (E,∇) at the generic point of S is solvable (and for extensions of connections satisfying the conjecture). More recently, B. Farb and M. Kisin [FK] proved it for certain locally symmetric varieties X. In general, one is lacking methods to think of the problem.

The first author is supported by the SFB/TR45 and the ERC Advanced Grant 226257. The second author is supported by the Bessel Award of the Humboldt Foundation and a Polish MNiSW grant (contract number N N201 420639).

Y. André in [An, II] and E. Hrushovsky in [Hr, V] formulated the following equal characteristic 0 analog of the conjecture: if $X \to S$ is a smooth morphism of smooth connected varieties defined over a characteristic 0 field k, then if (E_S, ∇_S) is a relative integrable connection such that for all closed points s of some non-trivial open $U \subset S$, $(E_S, \nabla_S) \times_S s$ is trivialized by a finite étale cover of $X \times_S s$, then $(E, \nabla)|_{X_{\bar{\eta}}}$ should be trivialized by a finite étale cover, where $\bar{\eta}$ is a geometric generic point and $X_{\bar{\eta}} = X \times_S \bar{\eta}$. So the characteristic 0 analogy to integrable connections is simply integrable connections, and to the p-curvature condition is the trivialization of the connection by a finite étale cover. André proved it [An, Prop. 7.1.1], using Jordan's theorem and Simpson's moduli of flat connections, while Hrushovsky [Hr, p.116] suggested a proof using model theory.

It is tempting to formulate an equal characteristic $p > 0$ analog of Y. André's theorem. A main feature of integrable connections over a field k of characteristic 0 is that they form an abelian, rigid, k-linear tensor category. In characteristic $p > 0$, the category of bundles with an integrable connection is only $\mathscr{O}_{X^{(1)}}$-linear, where $X^{(1)}$ is the relative Frobenius twist of X, and the notion is too weak. On the other hand, in characteristic 0, the category of bundles with a flat connection is the same as the category of $\mathscr{O}_X$-coherent $\mathscr{D}_X$-modules. In characteristic $p > 0$, $\mathscr{O}_X$-coherent $\mathscr{D}_X$-modules over a smooth variety X defined over a field k form an abelian, rigid, k-linear tensor category (see [Gi]). It is equivalent to the category of stratified bundles. It bears strong analogies with the category of bundles with an integrable connection in characteristic 0. For example, if X is projective smooth over an algebraically closed field, the triviality of the étale fundamental group forces all such $\mathscr{O}_X$-coherent $\mathscr{D}_X$-modules to be trivial ([EM]).

So we raise the QUESTION 1: let $f : X \to S$ be a smooth projective morphism of smooth connected varieties, defined over an algebraically closed characteristic $p > 0$ field, let (E, ∇) be a stratified bundle relative to S, such that for all closed point s of some non-trivial open $U \subset S$, the stratified bundle $(E, \nabla)|_{X_s}$ is trivialized by a finite étale cover of $X_s := X \times_S s$. Is it the case that the stratified bundle $(E, \nabla)|_{X_{\bar{\eta}}}$ is trivialized by a finite étale cover of $X_{\bar{\eta}}$?.

In this form, this is not true. Y. Laszlo [Ls] constructed a one dimensional non-trivial family of bundles over a curve over $\mathbb{F}_2$ which is fixed by the square of Frobenius, as a (negative) answer to a question of J. de Jong concerning the behavior of representations of the étale fundamental group over a finite field $\mathbb{F}_q$, $q = p^a$, with values in $GL(r, \mathbb{F}((t)))$, where $\mathbb{F} \supset \mathbb{F}_2$ is a finite extension. In fact, Laszlo's example yields also a counter-example to the question as stated above. We explain this in Sections 1 and 4 (see Corollary 4.3). We remark that if E is a bundle on X, such that the bundle $E|_{X_s}$ is stable, numerically flat (see Definition 3.2) and moves in the moduli, then $E_{\bar{\eta}}$ cannot be trivialized by a finite étale cover (see Proposition 4.2). In contrast, we show that if the family $X \to S$ is trivial (as it is in Laszlo's example), thus $X = Y \times_k S$, if k is algebraically closed, and if $(F_Y^n \times \text{identity}_s)^*(E)|_{Y \times_k s} \cong E|_{Y \times_k s}$ for all closed points s of some non-trivial open in S and some fixed natural number n, then the moduli points of $E|_{Y \times_k s}$ are constant (see Proposition 4.4). Here $F_Y : Y \to Y$ is the absolute Frobenius of Y. In Laszlo's example, one does have $(F_Y^2 \times \text{identity}_s)^*(E)|_{Y \times_k s} \cong E|_{Y \times_k s}$ but only over $k = \mathbb{F}_2$ (i.e., S is also defined over $\mathbb{F}_2$). When one extends the family to the algebraic closure of $\mathbb{F}_2$, to go from the absolute Frobenius over $\mathbb{F}_2$, that is the relative Frobenius over k, to the absolute one, one needs to replace the power 2 with a higher power $n(s)$, which depends on the field of definition of s, and is not bounded.

So we modify question 1 in QUESTION 2: let $f : X \to S$ be a smooth projective morphism of smooth connected varieties, defined over an algebraically closed characteristic field k of characteristic $p > 0$, let E be a bundle such that for all closed points s of some non-trivial open $U \subset S$, the bundle $E|_{X_s}$ is trivialized by a finite Galois étale cover of $X_s := X \times_S s$ of order prime to p. Is it the case that the bundle $E|_{X_{\bar{\eta}}}$ is trivialized by a finite étale cover of $X_{\bar{\eta}}$?.

The answer is nearly yes: it is the case if k is not algebraic over its prime field (Theorem 5.1 2)). If $k = \bar{\mathbb{F}}_p$, it might be wrong (Remarks 5.4 2), but what remains true is that there exists a finite étale cover of $X_{\bar{\eta}}$ over which the pull-back of E is a direct sum of line bundles (Theorem 5.1 1)). The idea of the proof is borrowed from the proof of Y. André's theorem [An, Thm 7.2.2]. The assumption on the degrees of the Galois covers of X_s trivializing $E|_{X_s}$ is necessary (as follows from Laszlo's example) and it allows us to apply Brauer-Feit's theorem [BF, Theorem] in place of Jordan's theorem used by André. However, there is no direct substitute for Simpson's moduli spaces of flat bundles. Instead, we use the moduli spaces constructed in [La1] and we carefully analyze subloci containing the points of interest, that is the numerically flat bundles. The necessary material needed on moduli is gathered in Section 3.

Finally we raise the general QUESTION 3: let $f : X \to S$ be a smooth projective morphism of smooth connected varieties, defined over an algebraically closed characteristic $p > 0$ field, let (E, ∇) be a stratified bundle relative to S, such that for all closed points s of some non-trivial open $U \subset S$, the stratified bundle $(E, \nabla)|_{X_s}$ is trivialized by a finite Galois étale cover of $X_s := X \times_S s$ of order prime to p. Is it the case that the bundle $(E, \nabla)|_{X_{\bar{\eta}}}$ is trivialized by a finite étale cover of $X_{\bar{\eta}}$?

We give the following not quite complete answer. If the rank of E is 1, (in which case the assumption on the degrees of the Galois covers is *automatically fulfilled*), then the answer is yes provided S is projective, and for any $s \in U$, $\mathrm{Pic}^{\tau}(X_s)$ is reduced (see Theorem 7.1). The proof relies on (a variant of) an idea of M. Raynaud [Ra], using the height function associated to a symmetric line bundle (that is the reason for our assumption on S) on the abelian scheme and its dual, to show that an infinite Verschiebung-divisible point has height equal to 0 (Theorem 6.2). If E has any rank, then the answer is yes if k is not $\bar{\mathbb{F}}_p$ (Theorem 7.2 2)). In general, there is a prime to p-order Galois cover of $X_{\bar{\eta}}$ such that the pull-back of E becomes a sum of stratified line bundles (Theorem 7.2 1)).

Acknowledgements: The first author thanks Michel Raynaud for the fruitful discussions in November 2009, which are reflected in [Ra] and in Section 6. The first author thanks Johan de Jong for a beautiful discussion in November 2010 on the content of [EM], where she suggested question 1 to him, and where he replied that Laszlo's example should contradict this, and that this should be better understood. The second author would like to thank Stefan Schröer for destroying his naive hopes concerning

Néron models of Frobenius twists of an abelian variety. We thank Damian Rössler for discussions on p-torsion on abelian schemes over functions fields. We thank the referee of a first version of the article. He/she explained to us that the dichotomy in Theorem 5.1 2) and in Theorem 7.2 2) should be $\bar{\mathbb{F}}_p$ or not rather that countable or not, thereby improving our result.

1 Preliminaries on relative stratified sheaves

Let S be a scheme of characteristic p (i.e., $\mathscr{O}_S$ is an $\mathbb{F}_p$-algebra). By $F_S^r : S \to S$ we denote the r-th *absolute Frobenius morphism* of S which corresponds to the p^r-th power mapping on $\mathscr{O}_S$.

If X is an S-scheme, we denote by $X_S^{(r)}$ the fiber product of X and S over the r-th Frobenius morphism of S. If it is clear with respect to which structure X is considered, we simplify the notation to $X^{(r)}$. Then the r-th absolute Frobenius morphism of X induces the *relative Frobenius morphism* $F_{X/S}^r : X \to X^{(r)}$. In particular, we have the following commutative diagram:

$$\begin{tikzcd}
X \arrow[r, "F_{X/S}^r"'] \arrow[rr, bend left, "F_X^r"] \arrow[dr] & X^{(r)} \arrow[r, "W_X^r"'] \arrow[d] & X \arrow[d] \\
 & S \arrow[r, "F_S^r"'] & S
\end{tikzcd}$$

which defines $W_{X/S}^r : X^{(r)} \to X$.

Making $r = 1$ and replacing X by $X^{(i)}$, this induces the similar diagram

$$\begin{tikzcd}
X^{(i)} \arrow[r, "F_{X^{(i)}/S}"'] \arrow[rr, bend left, "F_{X^{(i)}}"] \arrow[dr] & X^{(i+1)} \arrow[r, "W_{X^{(i)}}"'] \arrow[d] & X^{(i)} \arrow[d] \\
 & S \arrow[r, "F_S"'] & S
\end{tikzcd}$$

We assume that X/S is smooth. A *relative stratified sheaf* on X/S is a sequence $\{E_i, \sigma_i\}_{i \in \mathbb{N}}$ of locally free coherent $\mathscr{O}_{X^{(i)}}$-modules E_i on $X^{(i)}$ and isomorphisms $\sigma_i : F^*_{X^{(i)}/S} E_{i+1} \to E_i$ of $\mathscr{O}_{X^{(i)}}$-modules. A *morphism of relative stratified sheaves* $\{\alpha_i\} : \{E_i, \sigma_i\} \to \{E_i', \sigma_i'\}$ is a sequence of $\mathscr{O}_{X^{(i)}}$-linear maps $\alpha_i : E_i \to E_i'$ compatible with the σ_i, that is such that $\sigma_i' \circ F^*_{X^{(i)}/S} \alpha_{i+1} = \alpha_i \circ \sigma_i$.

This forms a category $\mathsf{Strat}(X/S)$, which is contravariant for morphisms $\varphi : T \to S$: to $\{E_i, \sigma_i\} \in \mathsf{Start}(X/S)$ one assigns $\varphi^*\{E_i, \sigma_i\} \in \mathsf{Strat}(X \times_S T/T)$ in the obvious way: φ induces $1_{X^{(i)}} \times \varphi : X^{(i)} \times_S T \to X^{(i)}$ and $(\varphi^*\{E_i, \sigma_i\})_i = \{(1_{X^{(i)}} \times \varphi)^* E_i, (1_{X^{(i)}} \times \varphi)^*(\sigma_i)\}$.

If $S = \operatorname{Spec} k$ where k is a field, $\mathsf{Strat}(X/k)$ is an abelian, rigid, tensor category. Giving a rational point $x \in X(k)$ defines a fiber functor via $\omega_x : \mathsf{Strat}(X/k) \to \mathsf{Vec}_k$, $\omega_x(\{E_i, \sigma_i\}) = (E_0)|_x$ in the category of finite dimensional vector spaces over k, thus a k-group scheme $\pi(\mathsf{Strat}(X/k), \omega_x) = \mathrm{Aut}^{\otimes}(\omega_x)$. Tannaka duality implies that $\mathsf{Strat}(X/k)$ is equivalent via ω_x to the representation category of $\pi(\mathsf{Strat}(X/k), \omega_x)$ with values in Vec_k. For any object $\mathbb{E} := \{E_i, \sigma_i\} \in \mathsf{Strat}(X/k)$, we define its *monodromy group* to be the k-affine group scheme $\pi(\langle \mathbb{E} \rangle, \omega_x)$, where $\langle \mathbb{E} \rangle \subset \mathsf{Strat}(X/k)$ is the full subcategory spanned by $\mathbb{E}$. This is the image of $\pi(\mathsf{Strat}(X/k), \omega_x)$ in $GL(\omega_x(\mathbb{E}))$ ([DM, Proposition 2.21 a)]). We denote by $\mathbb{I}_{X/k} \in \mathsf{Strat}(X/k)$ the trivial object, with $E^i = \mathscr{O}_{X^{(i)}}$ and $\sigma_i = $ Identity.

Lemma 1.1. *With the notation above*

1) *If $h : Y \to X$ is a finite étale cover such that $h^*\mathbb{E}$ is trivial, then $h_*\mathbb{I}_{Y/k}$ has finite monodromy group and one has a faithfully flat homomorphism $\pi(\langle h_*\mathbb{I}_{Y/k} \rangle, \omega_x) \to \pi(\langle \mathbb{E} \rangle, \omega_x)$. Thus in particular, $\mathbb{E}$ has finite monodromy group as well.*

2) *If $\mathbb{E} \in \mathsf{Strat}(X/k)$ has finite monodromy group, then there exists a $\pi(\langle \mathbb{E} \rangle, \omega_x)$-torsor $h : Y \to X$ such that $h^*\mathbb{E}$ is trivial in $\mathsf{Strat}(Y/k)$. Moreover, one has an isomorphism $\pi(\langle h_*\mathbb{I}_{Y/k} \rangle, \omega_x) \xrightarrow{\cong} \pi(\langle \mathbb{E} \rangle, \omega_x)$.*

Proof. We first prove 2). Assume $\pi(\langle \mathbb{E} \rangle, \omega_x) =: G$ is a finite group scheme over k. One applies Nori's method [No, Chapter I, II]: the regular representation of G on the affine k-algebra $k[G]$ of regular function defines the Artin k-algebra $k[G]$ as a k-algebra object of the representation category of G on finite dimensional k-vector spaces, (such that $k \subset k[G]$ is the maximal trivial subobject). Thus by Tannaka duality, there is an object $\mathbb{A} = (A^i, \tau_i) \in \mathsf{Strat}(X/k)$, which is an $\mathbb{I}_{X/k}$-algebra object, (such that $\mathbb{I}_{X/k} \subset \mathbb{A}$ is the maximal trivial subobject). We define $h_i : Y_i = \operatorname{Spec}_{X^{(i)}} A^i \to X^{(i)}$. Then the isomorphism τ_i yields an $\mathscr{O}_{X^{(i)}}$-isomorphism between $Y^{(i)} \xrightarrow{h^{(i)}} X^{(i)}$ and $Y_i \xrightarrow{h_i} X^{(i)}$, (see, e.g., [SGA5, Exposé XV, § 1, Proposition 2]), and via this isomorphism, $\mathbb{A}$ is isomorphic to $h_*\mathbb{I}_{Y/k}$. On the other hand, $\omega_x(\mathbb{E})$ is a sub G-representation of $k[G]^{\oplus n}$ for some $n \in \mathbb{N}$, thus $\mathbb{E} \subset \mathbb{A}^{\oplus n}$ in $\mathsf{Strat}(X/k)$, thus there is an inclusion $\mathbb{E} \subset (h_*\mathbb{I}_{Y/k})^{\oplus n}$ in $\mathsf{Strat}(X/k)$, thus $h^*\mathbb{E} \subset (h^*h_*\mathbb{I}_{Y/k})^{\oplus n}$ in $\mathsf{Strat}(Y/k)$. Since $(h^*h_*\mathbb{I}_{Y/k})$ is isomorphic to $\oplus_{\mathrm{length}_k k[G]} \mathbb{I}_{Y/k}$ in $\mathsf{Strat}(Y/k)$ (recall that by [dS, Proposition 13], G is an étale group scheme), then $h^*\mathbb{E}$ is isomorphic to $\oplus_r \mathbb{I}_{Y/k}$, where r is the rank of $\mathbb{E}$. This shows the first part of the statement, and shows the second part as well: indeed, $\mathbb{E}$ is then a subobject of $\oplus_r h_*\mathbb{I}_{Y/k}$, thus $\langle \mathbb{E} \rangle \subset \langle h_*\mathbb{I}_{Y/K} \rangle$ is a full subcategory. One applies [DM, Proposition 2.21 a)] to show that the induced homomorphism $\pi(\langle h_*\mathbb{I}_{Y/k} \rangle, \omega_x) \to \pi(\langle \mathbb{E} \rangle, \omega_x) = G$ is faithfully flat. So $\pi(\langle h_*\mathbb{I}_{Y/k} \rangle, \omega_x)$ acts on $\omega_x(h_*\mathbb{I}_Y) = k[G]$ via its quotient G and the regular representation $G \subset GL(k[G])$. Thus the homomorphism is an isomorphism.

We show 1). Assume that there is a finite étale cover $h : Y \to X$ such that $h^*\mathbb{E}$ is isomorphic in $\mathsf{Strat}(Y/k)$ to $\oplus_r \mathbb{I}_{Y/k}$ where r is the rank of $\mathbb{E}$. Then $\mathbb{E} \subset \oplus_r h_*\mathbb{I}_{Y/k}$, thus $\pi(\langle h_*\mathbb{I}_{Y/k} \rangle, \omega_x) \to \pi(\langle \mathbb{E} \rangle, \omega_x)$ is faithfully flat [DM, loc. cit.], so we are reduced to showing that $\langle h_*\mathbb{I}_{Y/k} \rangle$ has finite monodromy. But, by the same argument as on $\mathbb{E}$, any of its objects of rank r' lies in $\oplus_{r'} h_* \mathbb{I}_{Y/k}$. So we apply [DM, Proposition 2.20 a)] to conclude that the monodromy of $h_* \mathbb{I}_{Y/k}$ is finite. □

Corollary 1.2. *With the notations as in 1.1, if $\mathbb{E} \in \mathsf{Strat}(X/k)$ has finite monodromy group, then for any field extension $K \supset k$, $\mathbb{E} \otimes K \in \mathsf{Strat}(X \otimes K/K)$ has finite monodromy group.*

Let E be an $\mathcal{O}_X$-module. We say that *E has a stratification relative to S* if there exists a relative stratified sheaf $\{E_i, \sigma_i\}$ such that $E_0 = E$.
Let us consider the special case $S = \operatorname{Spec} k$, where k is a perfect field, and X/k is smooth. An *(absolute) stratified sheaf* on X is a sequence $\{E_i, \sigma_i\}_{i \in \mathbb{N}}$ of coherent $\mathcal{O}_X$-modules E_i on X and isomorphisms $\sigma_i : F_X^* E_{i+1} \to E_i$ of $\mathcal{O}_X$-modules.
As k is perfect, the $W_{X^{(i)}}$ are isomorphisms, thus giving an absolute stratified sheaf is equivalent to giving a stratified sheaf relative to $\operatorname{Spec} k$.
We now go back to the general case and we assume that S is an integral k-scheme, where k is a field. Let us set $K = k(S)$ and let $\eta : \operatorname{Spec} K \to S$ be the generic point of S. Let us fix an algebraic closure $\bar{K}$ of K and let $\bar{\eta}$ be the corresponding generic geometric point of S.
By contravariance, a relative stratified sheaf $\{E_i, \sigma_i\}$ on X/S restricts to a relative stratified sheaf $\{E_i, \sigma_i\}|_{X_s}$ in fibers X_s for s a point of S. We are interested in the relation between $\{E_i, \sigma_i\}|_{X_{\bar{\eta}}}$ and $\{E_i, \sigma_i\}|_{X_s}$ for closed points $s \in |S|$. More precisely, we want to understand under which assumptions the finiteness of $\langle \{E_i, \sigma_i\}|_{X_s} \rangle$ for all closed points $s \in |S|$ implies the finiteness of $\langle \{E_i, \sigma_i\}|_{X_{\bar{\eta}}} \rangle$. Recall that finiteness of $\mathbb{E} \subset \mathsf{Strat}(X_s)$ means that all objects of $\langle \mathbb{E} \rangle$ are subquotients in $\mathsf{Strat}(X_s)$ of direct sums of a single object, which is equivalent to saying that after the choice of a rational point, the monodromy group of $\mathbb{E}$ is finite ([DM, Proposition 2.20 (a)]).

Let X be a smooth variety defined over $\mathbb{F}_q$ with $q = p^r$. For all $n \in \mathbb{N} \setminus \{0\}$, one has the commutative diagram

$$\begin{tikzcd} X \arrow[r, "F^{rn}_{X/\mathbb{F}_q}"'] \arrow[rr, bend left, "(F^r_X)^n = F^{rn}_X"] \arrow[dr] & X^{(rn)} \arrow[r, "W^{rn}_{X/\mathbb{F}_q}"'] \arrow[d] & X \arrow[d] \\ & \operatorname{Spec} \mathbb{F}_q \arrow[r, "F^{rn}_{\mathbb{F}_q} = \mathrm{id}"'] & \operatorname{Spec} \mathbb{F}_q \end{tikzcd} \tag{1}$$

which allows us to identify $X^{(rn)}$ with X (as an $\mathbb{F}_q$-scheme).
Let S be an $\mathbb{F}_q$ connected scheme, with field of constants k, i.e. k is the normal closure of $\mathbb{F}_q$ in $H^0(S, \mathcal{O}_X)$. We define $X_S := X \times_{\mathbb{F}_q} S$.

Proposition 1.3. *Let E be a vector bundle on X_S. Assume that there exists a positive integer n such that we have an isomorphism*

$$\tau : ((F^r \times_{\mathbb{F}_q} \mathrm{id}_S)^n)^* E \simeq E. \tag{2}$$

Then E has a natural stratification $\mathbb{E}_\tau = \{E_i, \sigma_i\}$, $E_0 = E$ relative to S.

Proof. We define

$$E_{rn} = (W^{rn}_{X/\mathbb{F}_q} \times_{\mathbb{F}_q} \mathrm{id}_S)^* E. \tag{3}$$

Then we use the factorization

$$\begin{tikzcd}
X \arrow[r, "F_{X/\mathbb{F}_q}"] \arrow[drrrr] & X^{(1)} \arrow[r, "F_{X^{(1)}/\mathbb{F}_q}"] \arrow[drrr] & \cdots \arrow[r] & X^{(rn-1)} \arrow[r, "F_{X^{(rn-1)}/\mathbb{F}_q}"] \arrow[dr] & X^{(rn)} \arrow[d] \\
 & & & & \operatorname{Spec} \mathbb{F}_q
\end{tikzcd} \tag{4}$$

of $F^{rn}_{X/\mathbb{F}_q}$ and we define

$$E_{nr-1} = (F_{X^{(rn-1)}/\mathbb{F}_q} \times_{\mathbb{F}_q} \mathrm{id}_S)^* E_{rn}, \dots, E_1 = (F_{X^{(1)}/\mathbb{F}_q} \times_{\mathbb{F}_q} \mathrm{id}_S)^* E_2 \tag{5}$$

with identity isomorphisms $\sigma_{nr-1}, \dots, \sigma_1$. Then we use the isomorphism τ to define

$$\sigma_0 : E \simeq (F_{X/\mathbb{F}_q} \times_{\mathbb{F}_q} \mathrm{id}_S)^* E_1. \tag{6}$$

Assume we constructed the bundles E_i on $X^{(i)}$ for all $i \leq arn$ for some integer $a \geq 1$. We now replace the diagram (1) by the diagram

$$\begin{tikzcd}
X^{(arn)} \arrow[r, "F^{rn}_{X^{(arn)}/\mathbb{F}_q}"'] \arrow[rr, bend left, "(F^r_{X^{(arn)}})^n"] \arrow[dr] & X^{((a+1)rn)} \arrow[r, "W^{rn}_{X^{(arn)}/\mathbb{F}_q}"'] \arrow[d] & X^{(arn)} \arrow[d] \\
 & \operatorname{Spec} \mathbb{F}_q \arrow[r, "F^{rn}_{\mathbb{F}_q}=1"'] & \operatorname{Spec} \mathbb{F}_q
\end{tikzcd} \tag{7}$$

We then define

$$E_{(a+1)rn} = (W^{rn}_{X^{(arn)}/\mathbb{F}_q} \times_{\mathbb{F}_q} \mathrm{id}_S)^* E_{arn} \tag{8}$$

(which is equal to E under identification of $X^{(arn)}$ with X). Then we use the factorization

$$\begin{tikzcd}
X^{(arn)} \arrow[r, "F_{X^{(arn)}/\mathbb{F}_q}"] \arrow[drrrr] & X^{(arn+1)} \arrow[r, "F_{X^{(arn+1)}/\mathbb{F}_q}"] \arrow[drrr] & \cdots \arrow[r] & X^{((a+1)rn-1)} \arrow[r, "F_{X^{((a+1)rn-1)}/\mathbb{F}_q}"] \arrow[dr] & X^{((a+1)rn)} \arrow[d] \\
 & & & & \operatorname{Spec} \mathbb{F}_q
\end{tikzcd} \tag{9}$$

of $F^{rn}_{X^{(arn)}/\mathbb{F}_q}$ to define

$$E_{(a+1)rn-1} = (F_{X^{((a+1)rn-1)}/\mathbb{F}_q} \times_{\mathbb{F}_q} \mathrm{id}_S)^* E_{(a+1)rn}, \dots,$$
$$E_{arn+1} = (F_{X^{(arn+1)}/\mathbb{F}_q} \times_{\mathbb{F}_q} \mathrm{id}_S)^* E_{arn+2} \tag{10}$$

with identity isomorphisms $\sigma_{(a+1)nr-1},\dots,\sigma_{arn+1}$. Then we again use τ to define

$$\sigma_{arn}: E_{arn} \simeq (F^{rn}_{X^{(arn)}/\mathbb{F}_q})^* E_{arn+1}. \tag{11}$$

□

The above construction and [Gi, Proposition 1.7] imply

Proposition 1.4. *Assume in addition to* (2) *that X is proper and $\mathbb{F}_q \subset k \subset \bar{\mathbb{F}}_q$. Fix a rational point $x \in X_S(k)$. Then for any closed point $s \in |S|$, the Tannaka group scheme $\pi(\mathbb{E}_{\tau_s}, \omega_{x\otimes_k k(s)})$ of $\mathbb{E}_{\tau_s} := \mathbb{E}_\tau|_{X_s}$ over the residue field $k(s)$ of s is finite.*

Proof. The bundle E is base changed of a bundle E^0 defined over $X\times_{\mathbb{F}_q} S_0$ for some form S_0 of S defined over a finite extension $\mathbb{F}_{q^a}$ of $\mathbb{F}_q$ such that x is base change of an $\mathbb{F}_{q^a}$-rational point x_0 of $X\times_{\mathbb{F}_q} S_0$. We can also assume that τ comes by base change from $\tau_0 : ((F^r\times_{\mathbb{F}_q}\mathrm{id}_{S_0})^n)^*E^0 \simeq E^0$. Proposition 1.3 yields then a relative stratification $\mathbb{E}^0_{\tau_0} = (E^0_i, \sigma^0_i)$ of E^0 defined over $\mathbb{F}_{q^a}$, with $E_i = E^0_i \otimes_{\mathbb{F}_q^a} k$. A closed point s of $S = S_0\otimes_{\mathbb{F}_q^a} k$ is a base change of some closed point s_0 of S_0 of degree b say over $\mathbb{F}_{q^a}$. By Corollary 1.2 we just have to show that $\pi(\mathbb{E}_{(\tau_0)_{s_0}}, \omega_{x_0\otimes_{\mathbb{F}_{q^a}} k(s_0)})$ is finite. So we assume that $k = \mathbb{F}_{q^a},\ S = S_0,\ s = s_0$. The underling bundles of $\mathbb{E}_\tau$ and $\mathbb{E}_{\tau^m}$ are by construction all isomorphic for $m = ab$. Thus by [Gi, Proposition 1.7], $\mathbb{E}_\tau \simeq \mathbb{E}_{\tau^m}$ in $\mathsf{Strat}(X/k)$. But this implies that $F^{mn}_{X\times_{\mathbb{F}_{q^a}} s}(\mathbb{E}_{\tau_s}) \cong \mathbb{E}_{\tau_s}$. Thus E is algebraically trivializable on the Lang torsor $h : Y \to X\times_{\mathbb{F}_{q^a}} \mathbb{F}_{q^m}$ and the bundles E_i are trivializable on $Y\times_{X\times_{\mathbb{F}_{q^a}}\mathbb{F}_{q^m}} X^{(i)} = Y^{(i)}/\mathbb{F}_{q^m}$. Thus the stratified bundle $h^*\mathbb{E}_\tau$ on Y relative to $\mathbb{F}_{q^m}$ is trivial. We apply Lemma 1.1 to finish the proof. □

2 Étale trivializable bundles

Let X be a smooth projective variety over an algebraically closed field k. Let $F_X : X \to X$ be the absolute Frobenius morphism.
A locally free sheaf on X is called *étale trivializable* if there exists a finite étale covering of X on which E becomes trivial.
Note that if E is étale trivializable then it is numerically flat (see Definition 3.2 and the subsequent discussion). In particular, stability and semistability for such bundles are independent of a polarization (and Gieseker and slope stability and semistability are equivalent). More precisely, such E is stable if and only if it does not contain any locally free subsheaves of smaller rank and degree 0 (with respect to some or equivalently to any polarization).

Proposition 2.1. (see [LSt]) *If there exists a positive integer n such that $(F^n_X)^*E \simeq E$ then E is étale trivializable. Moreover, if $k = \bar{\mathbb{F}}_p$ then E is étale trivializable if and only if there exists a positive integer n and an isomorphism $(F^n_X)^*E \simeq E$.*

Proposition 2.2. (see [BD]) *If there exists a finite degree d étale Galois covering $f : Y \to X$ such that f^*E is trivial and E is stable, then one has an isomorphism $\alpha : (F^d_X)^*E \simeq E$.*

As a corollary we see that a line bundle on X/k is étale trivializable if and only if it is torsion of order prime to p. One implication follows from the above proposition. The other one follows from the fact that $(F_X^d)^*L \simeq L$ is equivalent to $L^{\otimes(p^d-1)} \simeq \mathscr{O}_X$ and for any integer n prime to p we can find d such that p^d-1 is divisible by n.
We recall that if E is any vector bundle on X such that there is a $d \in \mathbb{N}\setminus\{0\}$ and an isomorphism $\alpha : (F_X^d)^*(E) \cong E$, then E carries an *absolute* stratified structure $\mathbb{E}_\alpha$, i.e. a stratified structure relative to $\mathbb{F}_p$ by the procedure of Proposition 1.3. On the other hand, any stratified stratified structure $\{E_i, \sigma_i\}$ relative to $\mathbb{F}_p$ induces in an obvious way a stratified structure relative to k: the absolute Frobenius $F_X^n : X \to X$ factors through $W^n_{X/k} : X^{(n)} \to X$, so $\{(W^n_{X/k})^*E_n, (W^n_{X/k})^*\sigma_n\}$ is the relative stratified structure, denoted by $\mathbb{E}_{\alpha/k}$. Proposition 2.2 together with Lemma 1.1 2) show

Corollary 2.3. *Under the assumptions of Proposition 2.2, we can take* $d = \mathrm{length}_k k[\pi(\langle \mathbb{E}_{\alpha/k}\rangle, \omega_x)]$.

Let us also recall that there exist examples of étale trivializable bundles such that $(F_X^n)^*E \not\simeq E$ for every positive integer n (see Laszlo's example in [BD]).

Proposition 2.4. (Deligne; see [Ls, 3.2]) *Let X be an $\mathbb{F}_{p^n}$-scheme. If G is a connected linear algebraic group defined over a finite field $\mathbb{F}_{p^n}$ then the embedding $G(\mathbb{F}_{p^n}) \hookrightarrow G$ induces an equivalence of categories between the category of $G(\mathbb{F}_{p^n})$-torsors on X and G-torsors P over X with an isomorphism $(F_X^n)^*P \simeq P$.*

In particular, if G is a connected reductive algebraic group defined over an algebraically closed field k and P is a principal G-bundle on X/k such that there exists an isomorphism $(F_X^n)^*P \simeq P$ for some natural number $n>0$, then there exists a Galois étale cover $f : Y \to X$ with Galois group $G(\mathbb{F}_{p^n})$ such that f^*P is trivial. Indeed, every reductive group has a $\mathbb{Z}$-form so we can use the above proposition.

3 Preliminaries on relative moduli spaces of sheaves

Let S be a scheme of finite type over a universally Japanese ring R. Let $f : X \to S$ be a projective morphism of R-schemes of finite type with geometrically connected fibers and let $\mathscr{O}_X(1)$ be an f-very ample line bundle.
A *family of pure Gieseker semistable sheaves on the fibres of* $X_T = X \times_S T \to T$ is a T-flat coherent $\mathscr{O}_{X_T}$-module E such that for every geometric point t of T the restriction of E to the fibre X_t is pure (i.e., all its associated points have the same dimension) and Gieseker semistable (which is semistability with respect to the growth of the Hilbert polynomial of subsheaves defined by $\mathscr{O}_X(1)$ (see [HL, 1.2]). We introduce an equivalence relation $\sim$ on such families in the following way. $E \sim E'$ if and only if there exist filtrations $0 = E_0 \subset E_1 \subset \ldots \subset E_m = E$ and $0 = E'_0 \subset E'_1 \subset \ldots \subset E'_m = E'$ by coherent $\mathscr{O}_{X_T}$-modules such that $\oplus_{i=0}^m E_i/E_{i-1}$ is a family of pure Gieseker semistable sheaves on the fibres of X_T and there exists an invertible sheaf L on T such that $\oplus_{i=1}^m E'_i/E'_{i-1} \simeq \left(\oplus_{i=1}^m E_i/E_{i-1}\right) \otimes_{\mathscr{O}_T} L$.
Let us define the moduli functor

$$\mathscr{M}_P(X/S) : (\mathrm{Sch}/S)^o \to \mathrm{Sets}$$

from the category of locally noetherian schemes over S to the category of sets by

$$\mathscr{M}_P(X/S)(T) = \left\{ \begin{array}{l} \sim \text{equivalence classes of families of pure Gieseker} \\ \text{semistable sheaves on the fibres of } T \times_S X \to T, \\ \text{which have Hilbert polynomial } P. \end{array} \right\}.$$

Then we have the following theorem (see [La1, Theorem 0.2]).

THEOREM 3.1. *Let us fix a polynomial P. Then there exists a projective S-scheme $M_P(X/S)$ of finite type over S and a natural transformation of functors*

$$\theta : \mathscr{M}_P(X/S) \to \mathrm{Hom}_S(\cdot, M_P(X/S)),$$

which uniformly corepresents the functor $\mathscr{M}_P(X/S)$. For every geometric point $s \in S$ the induced map $\theta(s)$ is a bijection. Moreover, there is an open scheme $M^s_{X/S}(P) \subset M_P(X/S)$ that universally corepresents the subfunctor of families of geometrically Gieseker stable sheaves.

Let us recall that $M_P(X/S)$ *uniformly corepresents* $\mathscr{M}_P(X/S)$ means that for every flat base change $T \to S$ the fiber product $M_P(X/S) \times_S T$ corepresents the fiber product functor $\mathrm{Hom}_S(\cdot, T) \times_{\mathrm{Hom}_S(\cdot, S)} \mathscr{M}_P(X/S)$. For the notion of corepresentability, we refer to [HL, Definition 2.2.1]. In general, for every S-scheme T we have a well defined morphism $M_P(X/S) \times_S T \to M_P(X_T/T)$ which for a geometric point $T = \mathrm{Spec}\, k(s) \to S$ is bijection on points.
The moduli space $M_P(X/S)$ in general depends on the choice of polarization $\mathscr{O}_X(1)$.

DEFINITION 3.2. Let k be a field and let Y be a projective k-variety. A coherent $\mathscr{O}_Y$-module E is called *numerically flat*, if it is locally free and both E and its dual $E^* = \mathscr{H}om(E, \mathscr{O}_Y)$ are numerically effective on $Y \otimes \bar{k}$, where $\bar{k}$ is an algebraic closure of k.

Assume that Y is smooth. Then a numerically flat sheaf is strongly slope semistable of degree 0 with respect to any polarization (see [La2, Proposition 5.1]). But such a sheaf has a filtration with quotients which are numerically flat and slope stable (see [La2, Theorem 4.1]). Let us recall that a slope stable sheaf is Gieseker stable and any extension of Gieseker semistable sheaves with the same Hilbert polynomial is Gieseker semistable. Thus a numerically flat sheaf is Gieseker semistable with respect to any polarization.
Let P be the Hilbert polynomial of the trivial sheaf of rank r. In case S is a spectrum of a field we write $M_X(r)$ to denote the subscheme of the moduli space $M_P(X/k)$ corresponding to locally free sheaves. For a smooth projective morphism $X \to S$ we also define the moduli subscheme $M(X/S, r) \to S$ of the relative moduli space $M_P(X/S)$ as a union of connected components which contains points corresponding to numerically flat sheaves of rank r. Note that in positive characteristic numerical flatness is not an open condition. More precisely, on a smooth projective variety Y with an ample divisor H, a locally free sheaf with *numerically trivial Chern classes, that is with Chern classes c_i in the Chow group of codimension i cycles intersecting trivially $H^{\dim(Y)-i}$*

for all $i \geq 1$, is numerically flat if and only if it is strongly slope semistable (see [La2, Proposition 5.1]).
By definition for every family E of pure Gieseker semistable sheaves on the fibres of X_T we have a well defined morphism $\varphi_E = \theta([E]) : T \to M_P(X/S)$, which we call a *classifying morphism.*

PROPOSITION 3.3. *Let X be a smooth projective variety defined over an algebraically closed field k of positive characteristic. Let S be a k-variety and let E be a rank r locally free sheaf on $X \times_k S$ such that for every $s \in S(k)$ the restriction E_s is Gieseker semistable with numerically trivial Chern classes. Assume that the classifying morphism $\varphi_E : S \to M_X(r)$ is constant and for a dense subset $S' \subset S(k)$ the bundle E_s is étale trivializable for $s \in S'$. Then $E_{\bar{\eta}}$ is étale trivializable.*

Proof. If E_s is stable for some k-point $s \in S$ then there exists an open neighbourhood U of $\varphi_E(s)$, a finite étale morphism $U' \to U$ and a locally free sheaf $\mathscr{U}$ on $X \times_k U'$ such that the pull backs of E and $\mathscr{U}$ to $X \times_k (\varphi_E^{-1}(U) \times_U U')$ are isomorphic (this is called existence of a universal bundle on the moduli space in the étale topology). But $\varphi_E(S)$ is a point, so this proves that there exists a vector bundle on X such that E is its pull back by the projection $X \times_k S \to X$. In this case the assertion is obvious.
Now let us assume that E_s is not stable for all $s \in S(k)$. If $0 = E_0^s \subset E_1^s \subset ... \subset E_m^s = E_s$ is a Jordan–Hölder filtration (in the category of slope semistable torsion free sheaves), then by assumption the isomorphism classes of semi-simplifications $\oplus_{i=1}^m E_i^s/E_{i-1}^s$ do not depend on $s \in S(k)$. Let $(r_1, ..., r_m)$ denote the sequence of ranks of the components E_i^s/E_{i-1}^s for some $s \in S(k)$. Since there is only finitely many such sequences (they differ only by permutation), we choose some permutation that appears for a dense subset $S'' \subset S'$.
Now let us consider the scheme of relative flags $f : \mathrm{Flag}(E/S; P_1, ..., P_m) \to S$, where P_i is the Hilbert polynomial of $\mathscr{O}_X^{r_i}$. By our assumption the image of f contains S''. Therefore by Chevalley's theorem it contains an open subscheme U of S. Let us recall that the scheme of relative flags $\mathrm{Flag}(E|_{X \times_k U}/U; P_1, ..., P_m) \to U$ is projective. In particular, using Bertini's theorem (k is algebraically closed) we can find a generically finite morphism $W \to U$ factoring through this flag scheme. Let us consider pull back of the universal filtration $0 = F_0 \subset F_1 \subset ... \subset F_m = E_W$ to $X \times_k W$. Note that the quotients $F^i = F_i/F_{i-1}$ are W-flat and by shrinking W we can assume that they are families of Gieseker stable locally free sheaves (since by assumption F_s^i is Gieseker stable and locally free for some points $s \in W(k) \cap S'$). This and the first part of the proof implies that $E_{\bar{\eta}}$ has a filtration by subbundles such that the associated graded sheaf is étale trivializable. By Lemma 5.2 this implies that $E_{\bar{\eta}}$ is étale trivializable. □

4 LASZLO'S EXAMPLE

Let us describe Laszlo's example of a line in the moduli space of bundles on a curve fixed by the second Verschiebung morphism (see [Ls, Section 3]).
Let us consider a smooth projective genus 2 curve X over $\mathbb{F}_2$ with affine equation

$$y^2 + x(x+1)y = x^5 + x^2 + x.$$

In this case the moduli space $M_X(2,\mathcal{O}_X)$ of rank 2 vector bundles on X with trivial determinant is an $\mathbb{F}_2$-scheme isomorphic to $\mathbb{P}^3$. The pull back of bundles by the relative Frobenius morphism defines the Verschiebung map

$$V: M_{X^{(1)}}(2,\mathcal{O}_{X^{(1)}}) \simeq \mathbb{P}^3 \dashrightarrow M_X(2,\mathcal{O}_X) \simeq \mathbb{P}^3$$

which in appropriate coordinates can be described as

$$[a:b:c:d] \to [a^2+b^2+c^2+d^2 : ab+cd : ac+bd : ad+bc].$$

The restriction of V to the line $\Delta \simeq \mathbb{P}^1$ given by $b=c=d$ is an involution and it can be described as $[a:b] \to [a+b:b]$.
Using a universal bundle on the moduli space (which exists locally in the étale topology around points corresponding to stable bundles) and taking a finite covering $S \to \Delta$ we obtain the following theorem:

THEOREM 4.1. ([Ls, Corollary 3.2]) *There exist a smooth quasi-projective curve S defined over some finite extension of $\mathbb{F}_2$ and a locally free sheaf E of rank 2 on $X \times S$ such that $(F^2 \times \mathrm{id}_S)^*E \simeq E$, $\det E \simeq \mathcal{O}_{X\times S}$ and the classifying morphism $\varphi_E : S \to M_X(2,\mathcal{O}_X)$ is not constant. Moreover, one can choose S so that E_s is stable for every closed point s in S.*

Now note that the map $(F_X)^* : M_X(2,\mathcal{O}_X) \dashrightarrow M_X(2,\mathcal{O}_X)$ defined by pulling back bundles by the absolute Frobenius morphism can be described on Δ as $[a:b] \to [a^2+b^2 : b^2]$. In particular, the map $(F_X^{2n})^*|_\Delta$ is described as $[a:b] \to [a^{2n}, b^{2n}]$. It follows that if a stable bundle E corresponds to a modular point of $\Delta(\mathbb{F}_2^n)\backslash\Delta(\mathbb{F}_2^{n-1})$ (or, equivalently, E is defined over $\mathbb{F}_{2^n}$) then $(F_X^{2n})^*E \simeq E$ and $(F_X^m)^*E \not\simeq E$ for $0 < m < 2n$.
This implies that for $k = \bar{\mathbb{F}}_2$ and for every $s \in S(k)$, the bundle E_s which is the restriction to $X \times_{\mathbb{F}_2} s$ of the bundle E from Theorem 4.1, is étale trivializable.

Let X, S be varieties defined over an algebraically closed field k of positive characteristic. Assume that X is projective. Let us set $K = k(S)$. Let $\bar{\eta}$ be a generic geometric point of S.

PROPOSITION 4.2. *Let E be a bundle on $X_S = X \times_k S \to S$ which is numerically flat on the closed fibres of $X_S = X \times_k S \to S$. Assume that for some $s \in S$ the bundle E_s is stable and the classifying morphism $\varphi_E : S \to M_X(r)$ is not constant. Then $E_{\bar{\eta}} = E|_{X_{\bar{\eta}}}$ is not étale trivializable.*

Proof. Assume that there exists a finite étale cover $\pi' : Y' \to X_{\bar{\eta}}$ such that $(\pi')^*E_{\bar{\eta}} \simeq \mathcal{O}_{Y'}^r$. As k is algebraically closed, one has the base change $\pi_1(X) \xrightarrow{\cong} \pi_1(X_{\bar{K}})$ for the étale fundamental group ([SGA1, Exp. X, Cor.1.8]), so there exists a finite étale cover $\pi : Y \to X$ such that $\pi' = \pi \otimes \bar{K}$. Hence there exists a finite morphism $T \to U$ over some open subset U of S, such that $\pi_T^*(E_T)$ is trivial where $\pi_T = \pi \times_k \mathrm{id}_T : Y \times_k T \to X \times_k T$ and E_T =pull back by $X \times_k T \to X \times_k U$ of $E|_{X\times_k U}$.
So for any k-rational point $t \in T$, one has $\pi^*E_t \subset \mathcal{O}_Y^r$, where r is the rank of E. Hence $E_t \subset \pi_*\pi^*E_t \subset \pi_*\mathcal{O}_Y^r$, i.e., all the bundles E_t lie in one fixed bundle $\pi_*\mathcal{O}_Y^r$.

Since π is étale, the diagram

$$\begin{tikzcd} Y \arrow[r, "F_Y"] \arrow[d, "\pi"] & Y \arrow[d, "\pi"] \\ X \arrow[r, "F_X"] & X \end{tikzcd}$$

is cartesian (see, e.g., [SGA5, Exp. XIV, §1, Prop. 2]). Since X is smooth, F_X is flat. By flat base change we have isomorphisms $F_X^*(\pi_*\mathcal{O}_Y) \simeq \pi_*(F_Y^*\mathcal{O}_Y) \simeq \pi_*\mathcal{O}_Y$. In particular, this implies that $\pi_*\mathcal{O}_Y$ is strongly semistable of degree 0. Therefore if E_t is stable then it appears as one of the factors in a Jordan–Hölder filtration of $\pi_*\mathcal{O}_Y$. Since the direct sum of factors in a Jordan–Hölder filtration of a semistable sheaf does not depend on the choice of the filtration, there are only finitely many possibilities for the isomorphism classes of stable sheaves E_t for $t \in T(k)$.
It follows that in $U \subset S$ there is an infinite sequence of k-rational points s_i with the property that E_{s_i} is stable (since stability is an open property) and $E_{s_i} \cong E_{s_{i+1}}$. This contradicts our assumption that the classifying morphism φ_E is not constant. □

Corollary 4.3. *There exist smooth curves X and S defined over an algebraic closure k of $\mathbb{F}_2$ such that X is projective and there exists a locally free sheaf E on $X \times_k S \to S$ such that for every $s \in S(k)$, the bundle E_s is étale trivializable but $E_{\bar\eta}$ is not étale trivializable. Moreover, on E there exists a structure of a relatively stratified sheaf $\mathbb{E}$ such that for every $s \in S(k)$, the bundle $\mathbb{E}_s$ has finite monodromy but the monodromy group of $\mathbb{E}_{\bar\eta}$ is infinite.*

The second part of the corollary follows from Proposition 1.3. The above corollary should be compared to the following fact:

Proposition 4.4. *Let X be a projective variety defined over an algebraically closed field k of positive characteristic. Let S be a k-variety and let E be a rank r locally free sheaf on $X \times_k S$. Assume that there exists a positive integer n such that for every $s \in S(k)$ we have $(F_X^n)^*E_s \simeq E_s$, where F_X denotes the absolute Frobenius morphism. Then the classifying morphism $\varphi_E : S \to M_X(r)$ is constant and $E_{\bar\eta}$ is étale trivializable.*

Proof. By Proposition 2.1, if $(F_X^n)^*E_s \simeq E_s$ then there exists a finite étale Galois cover $\pi_s : Y_s \to X$ with Galois group $G = \mathrm{GL}_r(\mathbb{F}_{p^n})$ such that $\pi_s^*E_s$ is trivial (in this case it is essentially due to Lange and Stuhler; see [LSt]). This implies that $E_s \subset (\pi_s)_*\pi_s^*E_s \simeq ((\pi_s)_*\mathcal{O}_Y)^{\oplus r}$ and hence $\mathrm{gr}_{JH} E_s \subset (\mathrm{gr}_{JH}(\pi_s)_*\mathcal{O}_Y)^{\oplus r}$.
Since X is proper, the étale fundamental group of X is topologically finitely generated and hence there exists only finitely many finite étale coverings of X of fixed degree (up to an isomorphism). This theorem is known as the Lang–Serre theorem (see [LS, Théorème 4]). Let $\mathscr{S}$ be the set of all Galois coverings of X with Galois group G. Then for every closed k-point s of S the semi-simplification of E_s is contained in $(\mathrm{gr}_{JH}\,\alpha_*\mathcal{O}_Y)^{\oplus r}$ for some $\alpha \in \mathscr{S}$. Therefore there are only finitely many possibilities for images of k-points s in $M_X(r)$. Since S is connected, it follows that $\varphi_E : S \to M_X(r)$ is constant.

The remaining part of the proposition follows from Proposition 3.3.

□

Note that by Proposition 4.2 together with Corollary 2.3, the monodromy groups of E_s in Theorem 4.1 for $s \in S(k)$ are not uniformly bounded. In fact, only if k is an algebraic closure of a finite field do we know that the monodromy groups of E_s are finite because then E_s can be defined over some finite subfield of k and the isomorphism $(F^2)^*E_s \simeq E_s$ implies that for some n we have $(F_X^n)^*E_s \simeq E_s$ (see the paragraph following Theorem 4.1).
Moreover, the above proposition shows that in Theorem 4.1, we cannot hope to replace F with the absolute Frobenius morphism F_X.

5 ANALOGUE OF THE GROTHENDIECK-KATZ CONJECTURE IN POSITIVE EQUICHARACTERISTIC

As Corollary 4.3 shows, the positive equicharacteristic version of the Grothendieck–Katz conjecture which requests a relatively stratified bundle to have finite monodromy group on the geometric generic fiber once it does on all closed fibers, does not hold in general. But one can still hope that it holds for a family of bundles coming from representations of the prime-to-p quotient of the étale fundamental group. In this section we follow André's approach [An, Théorème 7.2.2] in the equicharacteristic zero case to show that this is indeed the case.

Let k be an algebraically closed field of positive characteristic p. Let $f : X \to S$ be a smooth projective morphism of k-varieties (in particular, integral k-schemes). Let η be the generic point of S. In particular, $X_{\bar\eta}$ is smooth (see [SGA1, Defn 1.1]).

THEOREM 5.1. *Let E be a locally free sheaf of rank r on X. Let us assume that there exists a dense subset $U \subset S(k)$ such that for every s in U, there is a finite Galois étale covering $\pi_s : Y_s \to X_s$ of Galois group of order prime-to-p such that $\pi_s^*(E_s)$ is trivial.*

1) *Then there exists a finite Galois étale covering $\pi_{\bar\eta} : Y_{\bar\eta} \to X_{\bar\eta}$ of order prime-to-p such that $\pi_{\bar\eta}^* E_{\bar\eta}$ is a direct sum of line bundles.*

2) *If k is not algebraic over its prime field and U is open in S, then $E_{\bar\eta}$ is étale trivializable on a finite étale cover $Z_{\bar\eta} \to X_{\bar\eta}$ which factors as a Kummer (thus finite abelian of order prime to p) cover $Z_{\bar\eta} \to Y_{\bar\eta}$ and a Galois cover $Y_{\bar\eta} \to X_{\bar\eta}$ of order prime to p.*

Proof. Without loss of generality, shrinking S if necessary, we may assume that S is smooth. Moreover, by passing to a finite cover of S and replacing U by its inverse image, we can assume that f has a section $\sigma : S \to X$.
By assumption for every $s \in U$ there exists a finite étale Galois covering $\pi_s : Y_s \to X_s$ with Galois group Γ_s of order prime-to-p and such that $\pi_s^* E_s$ is trivial. To these data one can associate a representation $\rho_s : \pi_1^{p'}(X_s, \sigma(s)) \to \Gamma_s \subset \mathrm{GL}_r(k)$ of the prime-to-p quotient of the étale fundamental group.

By the Brauer–Feit version of Jordan's theorem (see [BF, Theorem]) there exist a constant $j(r)$ such that Γ_s contains an abelian normal subgroup A_s of index $\leq j(r)$ (here we use assumption that the p-Sylow subgroup of Γ_s is trivial).
For a k-point s of S we have a homomorphism of specialization

$$\alpha_s : \pi_1(X_{\bar{\eta}}, \sigma(\bar{\eta})) \twoheadrightarrow \pi_1(X_s, \sigma(s)),$$

which induces an isomorphism of the prime-to-p quotients of the étale fundamental groups.
So for every $s \in U$ we can define the composite morphism

$$\tilde{\rho}_s : \pi_1^{p'}(X_{\bar{\eta}}, \sigma(\bar{\eta})) \xrightarrow{\alpha_s} \pi_1^{p'}(X_s, \sigma(s)) \xrightarrow{\rho_s} \Gamma_s \twoheadrightarrow \Gamma_s/A_s.$$

Let K be the kernel of the canonical homomorphism $\pi_* : \pi_1(X, \sigma(\bar{\eta})) \longrightarrow \pi_1(S, \bar{\eta})$, let $K^{p'}$ be its maximal pro-p'-quotient. Then by [SGA1, Exp. XIII, Proposition 4.3 and Exemples 4.4], one has $K^{p'} = \pi_1^{p'}(X_{\bar{\eta}}, \sigma(\bar{\eta}))$, the maximal pro-$p'$-quotient of $\pi_1(X_{\bar{\eta}}, \sigma(\bar{\eta}))$, and one has a short exact sequence

$$\{1\} \to \pi_1^{p'}(X_{\bar{\eta}}, \sigma(\bar{\eta})) \longrightarrow \pi_1'(X, \sigma(\bar{\eta})) \xrightarrow{\pi_*} \pi_1(S, \bar{\eta}) \to \{1\},$$

where $\pi_1'(X, \sigma(\bar{\eta}))$ is defined as the push-out of $\pi_1(X, \sigma(\bar{\eta}))$ by $K \to K^{p'}$.
Since $X_{\bar{\eta}}$ is proper, $\pi_1(X_{\bar{\eta}}, \sigma(\bar{\eta}))$ is topologically finitely generated. Therefore $\pi_1^{p'}(X_{\bar{\eta}}, \sigma(\bar{\eta}))$ is also topologically finitely generated and hence it contains only finitely many subgroups of indices $\leq j(r)$. Let G be the intersection of all such subgroups in $\pi_1^{p'}(X_{\bar{\eta}}, \sigma(\bar{\eta}))$. It is a normal subgroup of finite index. Since $\ker(\tilde{\rho}_s)$ is a normal subgroup of index $\leq j(r)$ in $\pi_1^{p'}(X_{\bar{\eta}}, \sigma(\bar{\eta}))$ we have

$$G \subset \bigcap_{s \in U} \ker(\tilde{\rho}_s).$$

Now let us consider the commutative diagram

$$\begin{tikzcd}
 & \pi_1(X_{\bar{\eta}}, \sigma(\bar{\eta})) \arrow[r] \arrow[d] & \pi_1(X, \sigma(\bar{\eta})) \arrow[r] \arrow[d] & \pi_1(S, \bar{\eta}) \arrow[r] \arrow[d] & \{1\} \\
\{1\} \arrow[r] & \pi_1^{p'}(X_{\bar{\eta}}, \sigma(\bar{\eta})) \arrow[r] & \pi_1'(X, \sigma(\bar{\eta})) \arrow[r] & \pi_1(S, \bar{\eta}) \arrow[r] & \{1\}
\end{tikzcd}$$

Then $G \cdot \sigma_*(\pi_1(S, \bar{\eta})) \subset \pi_1'(X, \sigma(\bar{\eta}))$ is a subgroup of finite index. It is open by the Nikolov–Segal theorem [NS, Theorem 1.1]. So the pre-image H of this subgroup under the quotient homomorphism $\pi_1(X, \sigma(\bar{\eta})) \to \pi_1'(X, \sigma(\bar{\eta}))$ defines a finite étale covering $h : X' \to X$.
Let us take $s \in S(k)$. Since the composition

$$H \subset \pi_1(X, \sigma(\bar{\eta})) \to \pi_1(X, \sigma(s)) \to \pi_1(S, s)$$

is surjective, the geometric fibres of $X' \to S$ are connected. Let us choose a k-point in X' lying over $\sigma(s)$. By abuse of notation we call it $\sigma'(s)$. Similarly, let us choose a geometric point $\sigma'(\bar\eta)$ of $X'_{\bar\eta}$ lying over $\sigma(\bar\eta)$. Then for any $s \in U$ we have the following commutative diagram:

$$\begin{tikzcd}
\pi_1^{p'}(X'_{\bar\eta}, \sigma'(\bar\eta)) \arrow[r, "h_*"] \arrow[rr, bend left, "0"] \arrow[d, "\simeq"] & \pi_1^{p'}(X_{\bar\eta}, \sigma(\bar\eta)) \arrow[r] \arrow[d, "\alpha_s \simeq"] & \pi_1^{p'}(X_{\bar\eta}, \sigma(\bar\eta))/G \arrow[d] \\
\pi_1^{p'}(X'_s, \sigma'(s)) \arrow[r, "h_*"] & \pi_1^{p'}(X_s, \sigma(s)) \arrow[r] & \Gamma_s/A_s
\end{tikzcd}$$

This diagram shows that $\pi_1^{p'}(X'_s, \sigma'(s)) \to \Gamma_s$ factors through A_s and hence $E'_s = (h^*E)_s$ is trivialized by a finite étale Galois covering $\pi'_s : Y'_s \to X'_s$ with an abelian Galois group of order prime to p, which is a subgroup of A_s. Since

$$E'_s \subset (\pi'_s)_*(\pi'_s)^* E'_s \simeq ((\pi'_s)_* \mathscr{O}_{Y'_s})^{\oplus r},$$

and $(\pi'_s)_*\mathscr{O}_{Y'_s}$ is a direct sum of torsion line bundles of orders prime to p, it follows that for every $s \in U$ the bundle E'_s is also a direct sum of torsion line bundles of order prime to p.

We consider the union $M(X'/S, r)$ of the components of $M_P(X'/S)$ containing moduli points of numerically flat bundles, as defined in Section 3. Let us consider the S-morphism $\psi : M(X'/S, 1)^{\times_S r} \to M(X'/S)$ given by $([L_1], \dots, [L_r]) \to [\oplus L_i]$ (in fact we give it by this formula on the level of functors; existence of the morphism follows from the fact that moduli schemes corepresent these functors). The bundle E' gives us a section $\tau : S \to M(X'/S, r)$, and by the above for every k-rational point s of U, the point $\tau(s)$ is contained in the image of ψ. Therefore $\tau(S)$ is contained in the image of ψ as ψ is projective (thus proper).

Let us consider the fibre product

$$\begin{tikzcd}
M(X'/S,1)^{\times_S r} \times_{M(X'/S,r)} S \arrow[r] \arrow[d] & S \arrow[d, "\tau"] \\
M(X'/S,1)^{\times_S r} \arrow[r] & M(X'/S,r)
\end{tikzcd}$$

Let us recall that in positive characteristic the canonical map $M(X' \times_S S'/S', r) \to M(X'/S, r) \times_S S'$ need not be an isomorphism (although it is an isomorphism for $r = 1$). Anyway we can find an étale morphism $S' \to S$ over some non-empty open subset of S, such that there exists a map $\upsilon : S' \to M(X' \times_S S'/S', 1)^{\times_{S'} r}$ which composed with $M(X' \times_S S'/S', 1)^{\times_{S'} r} \to M(X' \times_S S'/S', r) \to M(X'/S, r)$ gives the composition of $S' \to S$ with τ. This shows that the pull back E'' of E' to $X' \times_S S'$ has a filtration whose quotients are line bundles which are of degree 0 on the fibres of $X' \times_S S' \to S'$.

Now let us note the following lemma:

Lemma 5.2. *Let $f : X \to S$ be a projective morphism of k-varieties. Let $0 \to G_1 \to G \to G_2 \to 0$ be a sequence of locally free sheaves on X. Assume that there exists a dense subset $U \subset S(k)$ such that for each $s \in U$ this sequence splits after restricting to X_s. Then it splits on the fibre X_η over the generic point η of S.*

Proof. By shrinking S if necessary, we may assume that S is affine and the relative cohomology sheaf $R^1 p_* \mathscr{H}om(G_2, G_1)$ is locally free. The above short exact sequence defines a class $\lambda \in \operatorname{Ext}^1(G_2, G_1) \simeq H^0(S, R^1 f_* \mathscr{H}om(G_2, G_1))$, such that $\lambda(s) = 0$ for every k-rational point s of U. It follows that $\lambda = 0$ and hence the sequence is split over the generic point of S. □

Now let us note that on a smooth projective variety every short exact sequence of the form $0 \to G_1 \to G \to G_2 \to 0$ in which G is a direct sum of line bundles of degree 0 and G_2 is a line bundle of degree 0 splits. So the filtration of E'' restricted to the closed fibers splits. Therefore the above lemma and easy induction show that $E''_{\eta'}$ is a direct sum of line bundles, where η' is the generic point of S'. This shows the first part of the theorem.

To prove the second part of the theorem, we may assume that $U = S$. Let us take a line bundle L on X such that for every k-rational point s the line bundle L_s is étale trivializable. We need to prove that there exists a positive integer n prime to p and such that $L_\eta^{\otimes n} \simeq \mathscr{O}_{X_\eta}$.

We thank the referee for showing us the following lemma.

Lemma 5.3. *Let $g : A \to S$ be an abelian scheme and let σ be a section of g such that for all $s \in S(k)$, $\sigma(s)$ is torsion of order prime to p. Then σ is torsion of order prime to p.*

Proof. We may assume that S is normal and affine. Let us choose a subfield $k' \subset k$ that is finitely generated and transcendental over $\mathbb{F}_p$ and such that $A \to S$ and σ come by base change $\operatorname{Spec} k \to \operatorname{Spec} k'$ from an abelian scheme $g' : A' \to S'$ and a section σ' defined over k'. Let $m > 1$ be prime to p and let Γ be the subgroup $A'(S') \cap [m]^{-1}(\mathbb{Z}.\sigma')$ of $A'(S')$. Then Γ is a finitely generated group. Note that assumptions of Néron's specialization theorem [L, Chapter 9, Theorem 6.2] are satisfied and therefore there exists a Hilbert set Σ of points $s' \in S'$ for which the specialization map $A'(S') \to A'_{s'}(k(s'))$ is injective on Γ. Since the Hilbert subset $\Sigma \subset S'$ contains infinitely many closed points (see [L, Chapter 9, Theorems 5.1, 5.2 and 4.2]), there is a closed point $s \in S$ the image of which in S' lies in Σ. The specialization of $\mathbb{Z}.\sigma$ at s is injective and hence σ is torsion of order dividing the order of $\sigma(s)$, which is prime to p. □

Let us first assume that $X \to S$ is of relative dimension 1. By passing to a finite cover of S we can assume that f has a section. The relative Picard scheme $A = \operatorname{Pic}^0(X/S) \to S$ is smooth. Using the above lemma to the section corresponding to the line bundle L we see that there exists some positive integer n prime to p and a line bundle M on S such that $L^{\otimes n} \simeq f^* M$. In particular, $L_\eta^{\otimes n} \simeq \mathscr{O}_{X_\eta}$.

Now we use induction on the relative dimension of $f : X \to S$ to prove the theorem in the general case. Note that our assumptions imply that $L_{\bar\eta}$ is numerically flat and therefore the family $\{L_{\bar\eta}^{\otimes n}\}_{n\in\mathbb{Z}}$ is bounded. Thus for any sufficiently ample divisor H on $X_{\bar\eta}$ we have $H^1(X_{\bar\eta}, L_{\bar\eta}^{\otimes n}(-H)) = 0$ for all integers n. We consider such an H which is defined over η.

Using Bertini's theorem we can find a very ample divisor $Y \subset X$ in the linear system $|H|$ such that $f|_Y : Y \to S$ is smooth (possibly after shrinking S) and such that for every positive integer n we have $H^1(X_\eta, L^{\otimes n}(-Y)|_{Y_\eta}) = 0$. Indeed, shrinking S and using semicontinuity of cohomology, we may assume that H is defined over S, that the function $\dim H^0(X_s, \mathscr{O}_{X_s}(H))$ is constant and S is affine. Let us choose a k-rational point s in S. Then by Grauert's theorem (see [Ha, Chapter III, Corollary 12.9]) the restriction map

$$H^0(X, \mathscr{O}_X(H)) \to H^0(X_s, \mathscr{O}_{X_s}(H))$$

is surjective. By Bertini's theorem in the linear system $|\mathscr{O}_{X_s}(H)|$ there exists a smooth divisor. By the above we can lift it to a divisor $Y \subset X$, which after shrinking S is the required divisor.

Applying our induction assumption to $L|_Y$ on $Y \to X$ we see that there exists a positive integer n prime to p such that $(L|_Y)_\eta^{\otimes n} \simeq \mathscr{O}_{Y_\eta}$. Using the short exact sequence

$$0 \to L_\eta^{\otimes n}(-Y_\eta) \to L_\eta^{\otimes n} \to (L_Y^{\otimes n})_\eta \to 0$$

we see that the map

$$H^0(X_\eta, L_\eta^{\otimes n}) \to H^0(Y_\eta, (L_Y^{\otimes n})_\eta)$$

is surjective. In particular, $L_\eta^{\otimes n}$ has a section and hence it is trivial. □

Remarks 5.4. 1. Laszlo's example shows that the first part of the theorem is false if one does not assume that orders of the monodromy groups of E_s are prime to p (in this example $E_{\bar\eta}$ is a stable rank 2 vector bundle). Note that in this example, E has even the richer structure of a relatively stratified bundle (see Proposition 1.3).

2. Let E be a supersingular elliptic curve defined over $k = \bar{\mathbb{F}}_p$. Let M be a line bundle of degree 0 and of infinite order on $E_{\overline{\mathbb{F}_p(t)}}$. Then one can find a smooth curve S defined over k such that there exists a line bundle L on $X = S \times_k E \to S$ such that $L_{\bar\eta} \simeq M$. In this example the line bundle L_s is torsion for every k-rational point s of S as it is defined over a finite field. Since E is a supersingular elliptic curve, there are no torsion line bundles of order divisible by p. So in this case all line bundles L_s for $s \in S(k)$ are étale trivializable (and the monodromy group has order prime to p).

   This shows that the second part of Theorem 5.1 is no longer true if k is an algebraic closure of a finite field.

Let us keep the notation from the beginning of the section, i.e., k is an algebraically closed field of positive characteristic p and $f : X \to S$ is a smooth projective morphism of k-varieties (in particular connected) with geometrically connected fibers. For simplicity, we also assume that f has a section $\sigma : S \to X$.

LEMMA 5.5. *Let E be a locally free sheaf on X. If there exists a point $s_0 \in S(k)$ such that E_{s_0} is numerically flat then $E_{\bar\eta}$ is also numerically flat. In particular, if there exists a point $s_0 \in S(k)$ such that there is a finite covering $\pi_{s_0} : Y_{s_0} \to X_{s_0}$ such that $\pi_{s_0}^*(E_{s_0})$ is trivial, then $E_{\bar\eta}$ is also numerically flat.*

Proof. Let us fix a relatively ample line bundle. If E_{s_0} is numerically flat then it is strongly semistable with numerically trivial Chern classes (see [La2, Proposition 5.1]). Since E is S-flat, the restriction of E to any fiber has numerically trivial Chern classes (as intersection numbers remain constant on fibres). Now note that for any n the sheaf $(F^n_{X_{s_0}/k})^*E_{s_0}$ is slope semistable. Since slope semistability is an open property, it follows that $(F^n_{X_\eta/K})^*E_\eta$ is also slope semistable. By [HL, Corollary 1.3.8] it follows that $(F^n_{X_{\bar\eta}/\bar K})^*E_{\bar\eta}$ is also slope semistable. Thus $E_{\bar\eta}$ is strongly semistable with vanishing Chern classes and hence it is numerically flat by [La2, Proposition 5.1]. □

Let us recall that numerically flat sheaves on a proper k-variety Y form a Tannakian category. A rational point $y \in Y(k)$ neutralizes it. Thus we can define *S-fundamental group scheme of Y at the point y* (see [La2, Definition 6.1]). For a numerically flat sheaf E on Y, we consider the Tannaka k-group $\pi_S(\langle E\rangle, y) := \mathrm{Aut}^{\otimes}(\langle E\rangle, y) \subset \mathrm{GL}(E_y)$, where now $\langle E\rangle$ is the full tensor subcategory of numerically flat bundles spanned by E. We call it the *S-monodromy group scheme*. Using this language we can reformulate Theorem 5.1 in the following way (for simplicity we reformulate only the second part of the theorem).

THEOREM 5.6. *Let E be an S-flat family of numerically flat sheaves on the fibres of $X \to S$. Let us assume that k is not algebraic over its prime field and there exists a non-empty open subset $U \subset S(k)$ such that for every s in U, the S-monodromy group scheme $\pi_S(\langle E_s\rangle, \sigma(s))$ is finite étale of order prime-to-p. Then $\pi_S(\langle E_{\bar\eta}\rangle, \sigma(\bar\eta))$ is also finite étale.*

6 VERSCHIEBUNG DIVISIBLE POINTS ON ABELIAN VARIETIES: ON THE THEOREM BY M. RAYNAUD

Let K be an arbitrary field of positive characteristic p and let A be an abelian variety defined over K. The multiplication by p^n map $[p^n] : A \to A$ factors through the relative Frobenius morphism $F^n_{A/K} : A \to A^{(n)}$ and hence defines the *Verschiebung morphism* $V^n : A^{(n)} \to A$ such that $V^n F^n_{A/K} = [p^n]$.

DEFINITION 6.1. A K-point P of A is said to be *V-divisible* if for every positive integer n there exists a K-point P_n in $A^{(n)}$ such that $V^n(P_n) = P$.

Let T be an integral noetherian separated scheme of dimension 1 with field of rational functions K. Let us recall that a smooth, separated group scheme of finite type $\mathscr{A} \to T$ is called a *Néron model* of A if the general fiber of $\mathscr{A} \to T$ is isomorphic to A and for every smooth morphism $X \to T$, a morphism $X_K \to \mathscr{A}_K$ extends (then uniquely) to a T-morphism $X \to \mathscr{A}$.

Assume that the base field K is the function field of a normal projective variety S defined over a field k of positive characteristic p.
We say that A has *a good reduction* at a codimension 1 point $s \in S$ if the Néron model of A over $\operatorname{Spec} \mathscr{O}_{S,s}$ is an abelian scheme (the usual definition is slightly different as it assumes that the identity component of the special fibre of the Néron model is an abelian variety; it is equivalent to the above one by [BLR, 7.4, Theorem 5]). We say that A has *potential good reduction* at a codimension 1 point $s \in S$ if there exists a finite Galois extension K' of K such that if S' is the normalization of S in K' then $A_{K'}$ has good reduction at every codimension 1 point $s' \in S'$ lying over s.
We say that A has *(potential) good reduction* if it has (potential) good reduction at every codimension 1 point of S. Assume that A has good reduction at every codimension 1 point of S. Then there exists a *big open* subset $U \subset S$ (i.e., the codimension of the complement of U in S is ≥ 2) and an abelian U-scheme $\mathscr{A} \to U$. Note that the group $A(K)$ of K-points of A is isomorphic via the restriction map to the group of rational sections $U \dashrightarrow \mathscr{A}$ of $\mathscr{A} \to U$ defined over some big open subset of U. The section corresponding to $P \in A(K)$ will be denoted by $\tilde{P} : U \dashrightarrow \mathscr{A}$.
Let $c \in \operatorname{Pic} A$ be a class of a line bundle L. By the theorem of the cube c, satisfies the following equality:

$$m_{123}^* c - m_{12}^* c - m_{13}^* c - m_{23}^* c + m_1^* c + m_2^* c + m_3^* c = 0,$$

where m_I for $I \subset \{1,2,3\}$ is the map $A \times_K A \times_K A \to A$ defined by addition over the factors in I. (In particular, m_i is the i-th projection). Combining [MB, Chapter III, 3.1] (relying on [MB, Chapter II, Proposition 1.2.1]), the line bundle $L \in \operatorname{Pic}(A)$ extends uniquely (at least if we fix a rigidification) to a line bundle $\tilde{L}$ over $\mathscr{A}_V$ such that the class $\tilde{c} = [\tilde{L}] \in \operatorname{Pic}(\mathscr{A}_V)$ is cubical, i.e., satisfies the relation

$$\tilde{m}_{123}^* \tilde{c} - \tilde{m}_{12}^* \tilde{c} - \tilde{m}_{13}^* \tilde{c} - \tilde{m}_{23}^* \tilde{c} + \tilde{m}_1^* \tilde{c} + \tilde{m}_2^* \tilde{c} + \tilde{m}_3^* \tilde{c} = 0,$$

where $V \subset U$ is a big open subset and where $\tilde{m}_I$ for $I \subset \{1,2,3\}$ is the map $\mathscr{A} \times_S \mathscr{A} \times_S \mathscr{A} \to \mathscr{A}$ defined by addition over the factors in I.
Now let us choose an ample line bundle H on S. Then the map $\hat{h}_c : A(K) \to \mathbb{Z}$ given by

$$\hat{h}_c(P) = \deg_H (\tilde{P} - \tilde{0})^* \tilde{c}$$

is well defined as $\tilde{P}$ is defined on a big open subset of S and $\tilde{P}^* \tilde{L}$ extends to a rank 1 reflexive sheaf on S. This map is the canonical (Néron–Tate) height of A corresponding to c (see [MB, Chapter III, Section 3]).

The following theorem was suggested to the authors by M. Raynaud (in the good reduction case over a curve S, and with a somewhat different proof).

THEOREM 6.2. *Assume that A has potential good reduction. If $P \in A(K)$ is V-divisible and c is symmetric then $\hat{h}_c(P) = 0$.*

Proof. Let us first assume that A has good reduction. By assumption there exists a K-point P_n of $A^{(n)}$ such that $V^n(P_n) = P$. Since $\mathscr{A} \to U$ is an abelian scheme, so is $\mathscr{A}^{(n)} \to U$, thus P_n is the restriction to $\operatorname{Spec} K$ of $\tilde{P}_n \in \mathscr{A}^{(n)}(U)$.

Let us factor the absolute Frobenius morphism F_A^n into the composition of the relative Frobenius morphism $F_{A/K}^n : A \to A^{(n)}$ and $W_n : A^{(n)} \to A$. Let us set $c_n = W_n^* c$. Its cubical extension $\tilde{c}_n \in \mathrm{Pic}\,(\mathscr{A}_{V_n}^{(n)})$, for some big open $V_n \subset U$, together with H allows one to define $\hat{h}_{c_n}(P_n)$ by the corresponding formula. Since $(F_A^n)^* c = p^n c$, we have $(F_{A/K}^n)^* c_n = p^n c$. On the other hand, since c is symmetric, we have $[p^n]^* c = p^{2n} c$ and hence $(F_{A/K}^n)^*((V^n)^* c) = p^{2n} c$. Therefore

$$(F_{A/K}^n)^*((V^n)^* c - p^n c_n) = 0.$$

Since $F_{A/K}^n$ is an isogeny this implies that the class $d = (V^n)^* c - p^n c_n$ is torsion. By additivity and functoriality of the canonical height (see [Se, Theorem, p. 35]) we have

$$\hat{h}_c(P) = \hat{h}_{(V^n)^* c}(P_n) = \hat{h}_{p^n c_n}(P_n) + \hat{h}_d(P_n) = p^n \cdot \hat{h}_{c_n}(P_n)$$

(note that additivity implies that $\hat{h}_{md} = m\hat{h}_d$, so since $md = 0$ for some m, we get $\hat{h}_d = 0$). Therefore if $\hat{h}_c(P) \neq 0$ then $|\hat{h}_c(P)| \geq p^n$ and we get a contradiction if n is sufficiently large.

Now let us consider the general case. Since there exist only finitely many codimension 1 points $s \in S$ at which A has bad reduction, one can find a finite Galois extension K' of K such that if S' is the normalization of S in K' then $A_{K'}$ has good reduction at every codimension 1 point $s' \in S'$. On the other hand, if $P \in A(K)$ is V-divisible on A, $P \otimes K' \in A(K')$ is V divisible on $A_{K'}$. Then by the above we have $\hat{h}_{\pi^* c}(P') = 0$ and functoriality of the canonical height implies that $\hat{h}_c(P) = 0$. □

Remark 6.3. It is an interesting problem whether Theorem 6.2 holds for an arbitrary abelian variety A/K. Its proof shows that one can use the semiabelian reduction theorem to reduce the general statement to the case when A has semiabelian reduction (see [BLR, 7.4, Theorem 1]).

Now assume that S is geometrically connected. Then the extension $k \subset K$ is regular (i.e., K/k is separable and k is algebraically closed in K). Let (B, τ) be the K/k-trace of the abelian K-variety A, where B is an abelian k-variety and $\tau : B_K \to A$ is a homomorphism of abelian K-varieties (it exists by [Co, Theorem 6.2]). Let us recall that by definition (B, τ) is a final object in the category of pairs consisting of an abelian k-variety and a K-map from the scalar K-extension of this variety to A.

Since the extension $k \subset K$ is regular, the kernel K-group scheme of τ is connected (with connected dual) ([Co, Theorem 6.12]). Therefore τ is injective on K-points and in particular we can treat $B(k)$ as a subgroup of $A(K)$.

Corollary 6.4. *Assume that A has potential good reduction. If $P \in A(K)$ is V-divisible then $[P] \in (A(K)/B(k))_{\mathrm{tors}}$. In particular, if k is algebraically closed then $P \in B(k) + A(K)_{\mathrm{tors}} \subset A(K)$.*

Proof. We can choose the class $c \in \mathrm{Pic}\,(A)$ so that it is ample and symmetric. Then the first part of the corollary follows from Theorem 6.2 and [Co, Theorem 9.15] (which is true for regular extensions K/k).

To prove the second part take positive integer m such that $mP = Q \in B(k)$. Since k is algebraically closed, the set $B(k)$ is divisible and there exists $Q' \in B(k)$ such that $mQ' = Q$. Then $P = Q' + (P - Q')$, where $m(P - Q') = 0$. $\square$

Let us assume that the field k is algebraically closed. It is an interesting question whether a V-divisible K-point P of A can be written as a sum of $Q+R$, where $Q \in B(k)$ and $R \in A(K)_{\mathrm{tors}}$ is torsion *of order prime-to-p*.
By the Lang–Néron theorem ([Co, Theorem 2.1]), the groups $A^{(i)}(K)/B^{(i)}(k)$ are finitely generated. It follows that the groups $G_i = (A^{(i)}(K)/B^{(i)}(k))_{\mathrm{tors}}$ are finite.
Note that the homomorphism $B(k) \to B^{(i)}(k)$ induced by $F^i_{B/k}$ is a bijection. One has a factorization $F^i_{A/K} : A(K^{1/p^i}) \to A^{(i)}(K) \to A^{(i)}(K^{1/p^i})$, inducing a bijection $A(K^{1/p^i}) \to A^{(i)}(K)$. Thus in particular,

$$F_i : A(K)/B(k) \to A^{(i)}(K)/B^{(i)}(k)$$

is injective.
Moreover, the Verschiebung morphism induces the homomorphisms

$$V_i : A^{(i)}(K)/B^{(i)}(k) \to A(K)/B(k)$$

such that $V_iF_i = p^i$ and $F_iV_i = p^i$. This shows that prime-to-p torsion subgroups of groups G_i are isomorphic and in particular have the same order m.
Now let us assume that orders of the p-primary torsion subgroups of the abelian groups G_i are uniformly bounded by some p^e. Then for all $i \geq e$

$$F_i(m[P]) = F_i(V_i(m[P_i])) = p^i m[P_i] = 0.$$

This implies that $m[P] = 0$, so $mP \in B(k)$. Now $B(k)$ is a divisible group so there exists some $Q' \in B(k)$ such that $mP = mQ$. Then $R = P - Q \in A(K)$ is torsion of order prime to p. So we conclude

Lemma 6.5. *If the order of the G_i is bounded as i goes to infinity, under the assumption the Theorem 6.2, there exists a positive integer m, prime to p and such that $m \cdot P_i \in B(k)$ for every integer i.*

Note that the above assumption on G_i is satisfied, e.g., if A is an elliptic curve over the function field K of a smooth curve over $k = \bar{k}$. If A is isotrivial then the assertion is clear. If A is not isotrivial then the j-invariant of A is transcendental over k. In this case $A(K^{\mathrm{perf}})_{\mathrm{tors}}$ is finite (see [Le]) so orders of the groups $G_i = A^{(i)}(K)_{\mathrm{tors}}$ are uniformly bounded.

7 Stratified bundles

In this section we use the height estimate of the previous section and the fact that torsion stratified line bundles on a perfect field have order prime to p (apply Proposition 2.2 together with Lemma 1.1).

Let k be an algebraically closed field of positive characteristic p. Let $f : X \to S$ be a smooth projective morphism of k-varieties with geometrically connected fibres. Assume that S is projective, which surely is a very strong assumption. Indeed, if $k \neq \mathbb{F}_p$, and in the statement of Theorem 7.1, S' is open, then one obtains the stronger Theorem 7.2. For simplicity, let us also assume that f has a section $\sigma : S \to X$. Consider the torsion component $\mathrm{Pic}^{\tau}(X/S) \to S$ of identity of $\mathrm{Pic}(X/S) \to S$. Let $\varphi_n : \mathrm{Pic}(X/S) \to \mathrm{Pic}(X/S)$ be the multiplication by n map. Then there exists an open subgroup scheme $\mathrm{Pic}^{\tau}(X/S)$ of $\mathrm{Pic}(X/S)$ such that every geometric point s of S the fibre of $\mathrm{Pic}^{\tau}(X/S)$ over s is the union

$$\bigcup_{n>0} \varphi_n^{-1}(\mathrm{Pic}^0(X_s)),$$

where $\mathrm{Pic}^0(X_s)$ is the connected component of the identity of $\mathrm{Pic}(X_s/s)$. It is well known that $\mathrm{Pic}^{\tau}(X/S) \to S$ is also a closed subgroup scheme of $\mathrm{Pic}(X/S)$. Moreover, the morphism $\mathrm{Pic}^{\tau}(X/S) \to S$ is projective and the formation of $\mathrm{Pic}^{\tau}(X/S) \to S$ commutes with a base change of S (see, e.g., [Kl, Theorem 6.16 and Exercise 6.18]). We assume that $\mathrm{Pic}^0(X_s)$ is reduced for every point $s \in S$.

Theorem 7.1. *Let $\mathbb{L} = \{L_i, \sigma_i\}$ be a relatively stratified line bundle on X/S. Assume that there exists a dense subset $S' \subset S(k)$ such that for every $s \in S'$ the stratified bundle $\mathbb{L}_s = \mathbb{L}|_{X_s}$ has finite monodromy. Then $\mathbb{L}_{\bar{\eta}}$ has finite monodromy.*

Proof. Replacing $\mathbb{L}$ by a power $\mathbb{L}^{\otimes N}$, where N is sufficiently large, we may assume that $\mathbb{L}_s \in \mathrm{Pic}^0(X_s)$ for all closed points s in S (see [Kl, Corollary 6.17]).
By assumption $\hat{\pi} : \hat{\mathscr{A}} = \mathrm{Pic}^0(X/S) \to S$ is an abelian scheme. Let us consider the dual abelian scheme $\mathscr{A} \to S$. We have a well defined Albanese morphism $g : (X, \sigma) \to (\mathscr{A}, e)$ (see [FGA, Exposé VI, Théorème 3.3]). Moreover, the map $g^* : \mathrm{Pic}^0(\mathscr{A}/S) \to \hat{\mathscr{A}} = \mathrm{Pic}^0(X/S)$ is an isomorphism of S-schemes. Let us set $\hat{A} = \hat{\mathscr{A}}_\eta$.
Let P_i be the K-point of $\hat{A}^{(i)}$ corresponding to $(L_i)_\eta$. Note that the K-point $P_0 \in \hat{A}$ is V-divisible. Indeed, by the definition of a relative stratification we have $V^n(P_n) = P_0$ for all integers n. Similarly, we see that all the points $P_i \in \hat{A}^{(i)}(K)$ are V-divisible. By Corollary 6.4 it follows that $P_i \in \hat{B}^{(i)}(k) + \hat{A}^{(i)}(K)_{\mathrm{tors}}$, where $(\hat{B}/k, \hat{\tau} : \hat{B}_K \to \hat{A})$ is the K/k-trace of $\hat{A}$ (note that $(\hat{B}^{(i)}/k, \hat{\tau}^{(i)})$ is the K/k-trace of $\hat{A}^{(i)}$). So for every $i \geq 0$ we can write $P_i = Q_i + R_i$ for some $Q_i \in \hat{B}^{(i)}(k)$ and $R_i \in \hat{A}^{(i)}(K)_{\mathrm{tors}}$.
Now we transpose the above by duality. Let A be the dual abelian K-variety of $\hat{A}$ and B the dual abelian k-variety of $\hat{B}$. We have the K/k-images $\tau^{(i)} : A_\eta^{(i)} \to B_K^{(i)}$ and an S-morphism $\tau : \mathscr{A} \to B \times_k S$ (possibly after shrinking S). By abuse of notation we can treat L_i as line bundles on $\mathscr{A}$ because $g^* : \mathrm{Pic}^0(\mathscr{A}/S) \to \mathrm{Pic}^0(X/S)$ is an isomorphism. Let M_i be the line bundle on $B^{(i)}$ corresponding to Q_i and let $\pi_i : B^{(i)} \times_k S \to B^{(i)}$ denote the projection. Let us fix a non-negative integer i and take a positive integer n_i such that $n_i R_i = 0$. Then the line bundle $L_i^{\otimes n_i} \otimes \tau^* \pi_i^* M_i^{\otimes -n_i}$ has degree 0 on every fiber of $\mathscr{A} \to S$. Thus it is trivial after restriction to $\mathscr{A}_\eta$. Hence after shrinking S we can assume that $L_i^{\otimes n_i} \simeq \tau^* \pi_i^* M_i^{\otimes n_i}$.
Let us fix a point $s \in S(k)$ and consider the morphism

$$\pi_i' = (\tau^{(i)} \pi_i)_{\mathscr{A}_s^{(i)}} : \mathscr{A}_s^{(i)} \to B^{(i)}.$$

Note that $\tau^{(i)}$ has connected fibres and hence $(\pi_i')_*\mathscr{O}_{\mathscr{A}_s^{(i)}} = \mathscr{O}_{B^{(i)}}$. By assumption there exists a positive integer a_s, such that for every i the order of the line bundle $(L_i)_{\mathscr{A}_s}$ divides a_s. The important point is that a_s is prime to p.
Therefore $(\pi_i')^* M_i^{\otimes a_s n_i} \simeq \mathscr{O}_{A_s}$ and by the projection formula

$$M_i^{\otimes a_s n_i} \simeq (\pi_i')_*(\pi_i')^* M_i^{\otimes a_s n_i} \simeq (\pi_i')_*\mathscr{O}_{A_s} \simeq \mathscr{O}_B.$$

This implies that M_i is a torsion line bundle and hence $Q_i \in \hat{A}^{(i)}(K)_{\mathrm{tors}}$. Therefore

$$P_i = Q_i + R_i \in \hat{A}^{(i)}(K)_{\mathrm{tors}}.$$

Let us recall that the set of p-torsion points of $\hat{A}(K)$ is finite. Assuming it is not empty, we can therefore find a non-empty open subset $U \subset S$ such that for every $s \in U(k)$ and every p-torsion point $T \in \hat{A}(K)$ the section $\tilde{T}$ is defined on U and the point $\tilde{T}(s)$ is non-zero.
Let us write the order of P_i as $m_i p^{e_i}$, where m_i is not divisible by p. If $e_0 \geq 1$ then the point $m_0 p^{e_0-1} P_0$ is p-torsion in $\hat{A}(K)$. If we take $s \in S' \cap U(k)$, then $a_s m_0 p^{e_0-1}\tilde{P}_0(s) = [L_0^{\otimes a_s m_0}]_s = 0$, a contradiction. It follows that $m_0 P_0 = 0$. Similarly, the order of all P_i is prime to p.
As already mentioned in the last section, the homomorphism $\hat{A}(K^{1/p^i}) \to \hat{A}^{(i)}(K)$ induced by $F^i_{A/K}$ is a bijection. So we have an induced injection

$$F_i : \hat{A}(K) \to \hat{A}^{(i)}(K).$$

On the other hand, the Verschiebung morphism induces homomorphisms

$$V_i : \hat{A}^{(i)}(K) \to \hat{A}(K)$$

such that $V_i F_i(P) = p^i P$ and $F_i V_i(Q) = p^i Q$ for all $P \in \hat{A}(K)$ and $Q \in \hat{A}^{(i)}(K)$. Hence

$$p^i m_0 P_i = F_i V_i(m_0 P_i) = F_i(m_0 P_0) = 0$$

and since the order of P_i is prime to p we have $m_0 P_i = 0$ for all $i \geq 0$. Therefore $(L_i)_{\bar{\eta}}^{\otimes m_0} \simeq \mathscr{O}_{X_{\bar{\eta}}}$ for all i and the stratified line bundle $\mathbb{L}_{\bar{\eta}}$ has finite monodromy. ☐

Now we fix the following notation: k is an algebraically closed field of positive characteristic p and $f : X \to S$ is a smooth projective morphism of k-varieties with geometrically connected fibres.

Theorem 7.2. *Let $\mathbb{E} = \{E_i, \sigma_i\}$ be a relatively stratified bundle on X/S. Assume that there exists a dense subset $U \subset S(k)$ such that for every $s \in U$ the stratified bundle $\mathbb{E}_s = \mathbb{E}|_{X_s}$ has finite monodromy of order prime to p.*

1) *Then there exists a finite Galois étale covering $\pi_{\bar{\eta}} : Y_{\bar{\eta}} \to X_{\bar{\eta}}$ of order prime-to-p such that $\pi_{\bar{\eta}}^* \mathbb{E}_{\bar{\eta}}$ is a direct sum of stratified line bundles.*

2) *If $k \neq \bar{\mathbb{F}}_p$ and U is open in $S(k)$, then the monodromy group of $\mathbb{E}_{\bar\eta}$ is finite, and $\mathbb{E}_{\bar\eta}$ trivializes on a finite étale cover $Z_{\bar\eta} \to X_{\bar\eta}$ which factors as a Kummer (thus finite abelian of order prime to p) cover $Z_{\bar\eta} \to Y_{\bar\eta}$ and a Galois cover $Y_{\bar\eta} \to X_{\bar\eta}$ of order prime to p.*

Proof. We prove 1). Let us first remark that the schemes $X_{\bar\eta}^{(i)}$, $i \geq 0$, are all isomorphic (as schemes, not as k-schemes). Therefore the relative Frobenius induces an isomorphism on fundamental groups.
By the first part of Theorem 5.1 we know that there exists a finite Galois étale covering $\pi_i : Y_{\bar\eta,i} \to X_{\bar\eta}^{(i)}$ of degree prime to p such that $\pi_i^*(E_i)$ is a direct sum of line bundles $\oplus_1^r L_{ij}$. Note that from the proof of Theorem 5.1 the degree of π_i depends only on $\pi_1^{p'}(X_{\bar\eta}^{(i)}, \sigma^{(i)}(\bar\eta))$ and the Brauer-Feit constant $j(r)$, and therefore it can be bounded independently of i. Using the Lang–Serre theorem (see [LS, Théorème 4]) we can therefore assume that $Y_{\bar\eta,i} = Y_{\bar\eta}^{(i)}$, where $Y_{\bar\eta} = Y_{\bar\eta,0}$. Now we know that

$$\oplus_{j=1}^r L_{ij} \simeq (F^i_{Y_{\bar\eta}^{(i)}/\bar\eta})^* \big(\oplus_{j'=1}^r L_{i+1,j'}\big).$$

By the Krull-Schmidt theorem, the set of isomorphism classes of line bundles $\{L_{ij}\}_j$ is the same as the set of isomorphism classes of lines bundles which come by pull-back $\{(F^i_{Y_{\bar\eta}^{(i)}/\bar\eta})^*(L_{i+1,j'})\}_{j'}$. So we can reorder the indices j' so that

$$(F^i_{Y_{\bar\eta}^{(i)}/\bar\eta})^*(L_{i+1,j}) \cong L_{i,j}.$$

This finishes the proof of 1).
To prove 2), we do the proof 1) replacing $Y_{\bar\eta} \to X_{\bar\eta}$ by $Z_{\bar\eta} \to X_{\bar\eta}$ of Theorem 5.1 2). This finishes the proof of 2).

□

Remarks 7.3. 1) Case 2) of Theorem 7.2 applied to a line bundle extends Theorem 7.1, where S was assumed to be projective, $\mathrm{Pic}^0(X_s)$ reduced for all $s \in S$ closed, $S' \subset S(k)$ dense, to the case when S is not necessarily projective and $S' \subset S(k)$ is open and dense, but we have to assume that k is not algebraic over its prime field.

2) If $Y_{\bar\eta}$ has a good projective model satisfying assumptions of Theorem 7.1 then it follows that $\mathbb{E}_{\bar\eta}$ has finite monodromy.

References

[An] Y. André: Sur la conjecture des p-courbures de Grothendieck-Katz et un problème de Dwork, Geometric aspects of Dwork theory, Vol. I, II, 55–112, Walter de Gruyter GmbH & Co. KG, Berlin, 2004.

[BD] I. Biswas, L. Ducrohet: An analog of a theorem of Lange and Stuhler for principal bundles, *C. R. Math. Acad. Sci. Paris* 345 (2007), 495–497.

[BLR] S. Bosch, W. Lütkebohmert, M. Raynaud: Néron models, *Ergebnisse der Mathematik und ihrer Grenzgebiete* 21, Springer-Verlag, Berlin, 1990. x+325 pp.

[BF] R. Brauer, W. Feit: An analogue of Jordan's theorem in characteristic p, *Ann. of Math.* 84 (1966) 119–131.

[CC] D. V. Chudnovsky, G. V. Chudnovsky: Applications of Padé approximations to the Grothendieck conjecture on linear differential equations, Number theory (New York, 1983-84), 52–100, *Lecture Notes in Math.* 1135, Springer, Berlin, 1985.

[Co] B. Conrad: Chow's K/k-image and K/k-trace, and the Lang-Néron theorem, *Enseign. Math.* (2) 52 (2006), 37–108.

[DM] P. Deligne, J. Milne: Tannakian categories, *Lecture Notes in Mathematics* 100, 101–228. Springer Verlag.

[dS] J.-P. dos Santos: Fundamental group schemes for stratified sheaves, *J. Algebra* 317 (2007), no. 2, 691–713.

[EM] H. Esnault, V. Mehta: Simply connected projective manifolds in characteristic $p>0$ have no nontrivial stratified bundles, *Inventiones math.* 181 (2010), 449–465.

[FK] B. Farb, M. Kisin: The Grothendieck-Katz conjecture for certain locally symmetric varieties, *Int. Math. Res. Not.* 22 (2009), 4159–4167.

[Gi] D. Gieseker: Flat vector bundles and the fundamental group in non-zero characteristics, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* 2 (1975), 1–31.

[FGA] A. Grothendieck: Fondements de la géométrie algébrique, [Extraits du Séminaire Bourbaki, 1957–1962.] Secrétariat mathématique, Paris 1962.

[SGA1] A. Grothendieck: Revêtements étales et groupe fondamental, Documents Mathématiques (Paris) 3, Séminaire de géométrie algébrique du Bois Marie 1960–61, With two papers by M. Raynaud, Updated and annotated reprint of the 1971 original, Société Mathématique de France, Paris, 2003.

[SGA5] A. Grothendieck: Cohomologie l-adique et fonctions L, Séminaire de Géometrie Algébrique du Bois-Marie 1965–1966, *Lecture Notes in Mathematics*, Vol. 589. Springer-Verlag, Berlin-New York, 1977.

[Ha] R. Hartshorne: Algebraic Geometry, *Graduate Texts in Mathematics* 52, Springer Verlag, 1977.

[Hr] E. Hrushovsky: Computing the Galois group of a linear differential equation, in Differential Galois Theory, *Banach Center Publications* 58, Inst. of Math. Polish Academy of Sciences, Warszawa, 2002, 97–138.

[HL] D. Huybrechts, M. Lehn: The geometry of moduli spaces of sheaves, *Aspects of Mathematics* 31, 1997.

[Ka] N. Katz: Algebraic Solutions of Differential Equations (p-Curvature and the Hodge Filtration), *Invent. Math.* 18 (1972), 1–118.

[Kl] S. L. Kleiman: The Picard scheme, Fundamental algebraic geometry, 235–321, *Math. Surveys Monogr.* 123, Amer. Math. Soc., Providence, RI, 2005.

[L] S. Lang: Fundamentals of Diophantine geometry, Springer-Verlag, New York, 1983.

[LS] S. Lang, J.-P. Serre: Sur les revetements non ramifiés des variétés algébriques, *Amer. J. Math.* 79 (1957), 319–330.

[LSt] H. Lange, U. Stuhler: Vektorbündel auf Kurven und Darstellungen der algebraischen Fundamentalgruppe, *Math. Z.* 156 (1977), 73–83.

[La1] A. Langer: Semistable sheaves in positive characteristic, *Ann. of Math.* 159 (2004), 251–276.

[La2] A. Langer: On the S-fundamental group scheme, *Ann. Inst. Fourier* 61 (2011), 2077–2119.

[Ls] Y. Laszlo: A non-trivial family of bundles fixed by the square of Frobenius, *C. R. Acad. Sci. Paris Sér. I Math.* 333 (2001), 651–656.

[Le] M. Levin: On the group of rational points on elliptic curves over function fields, *Amer. J. Math.* 90 (1968), 456–462.

[MB] L. Moret-Bailly: Pinceaux de variétés abéliennes, *Astérisque* 129 (1985), 266 pp.

[NS] N. Nikolov, D. Segal: On finitely generated profinite groups I. Strong completeness and uniform bounds, *Ann. of Math. (2)* 165 (2007), 171–238.

[No] M. Nori: The fundamental group scheme, *Proc. Indian Acad. Sci.* 91 (1982), 73-122.

[Ra] M. Raynaud: Letter to Hélène Esnault dated November 24, 2009.

[Se] J.-P. Serre: Lectures on the Mordell-Weil theorem, *Aspects of Mathematics* E15, Friedr. Vieweg & Sohn, Braunschweig, 1989.

Hélène Esnault
Freie Universität Berlin
FB Mathematik und Informatik
Arnimallee 3
14195 Berlin, Germany
Helene.Esnault@fu-berlin.de

Adrian Langer
University of Warsaw
Institute of Mathematics
ul. Banacha 2
02-097 Warszawa, Poland
alan@mimuw.edu.pl

Birational Motivic Homotopy Theories and the Slice Filtration

Pablo Pelaez

Received: May 21, 2012
Revised: October 5, 2012

Communicated by Lars Hesselholt

Abstract. We show that there is an equivalence of categories between the orthogonal components for the slice filtration and the birational motivic stable homotopy categories which are constructed in this paper. Relying on this equivalence, we are able to describe the slices for projective spaces (including $\mathbb{P}^\infty$), Thom spaces and blow ups.

2000 Mathematics Subject Classification: 14F42
Keywords and Phrases: Birational Invariants, Motivic Homotopy Theory, Motivic Spectral Sequence, Slice Filtration.

1 Definitions and Notation

Our main result, theorem 3.6, shows that there is an equivalence of categories between the orthogonal components for the slice filtration (see definition 1.1) and the weakly birational motivic stable homotopy categories which are constructed in this paper (see definition 2.9). Relying on this equivalence; we are able to describe over an arbitrary base scheme (see theorems 4.2, 4.4 and 4.6) the slices for projective spaces (including $\mathbb{P}^\infty$), Thom spaces and blow ups. We also construct the birational motivic stable homotopy categories (see definition 2.4), which are a natural generalization of the weakly birational motivic stable homotopy categories, and show (see proposition 2.12) that there exists a Quillen equivalence between them when the base scheme is a perfect field. Our approach was inspired by the work of Kahn-Sujatha [1] on birational motives, where the existence of a connection between the layers of the slice filtration and birational invariants is explicitly suggested. Furthermore, this approach allows to obtain analogues for the slice filtration in the unstable setting (see remark 3.8).

In this paper X will denote a Noetherian separated base scheme of finite Krull dimension, Sch_X the category of schemes of finite type over X and Sm_X the full subcategory of Sch_X consisting of smooth schemes over X regarded as a site with the Nisnevich topology. All the maps between schemes will be considered over the base X. Given $Y \in Sch_X$, all the closed subsets Z of Y will be considered as closed subschemes with the reduced structure.

Let $\mathcal{M}$ be the category of pointed simplicial presheaves in Sm_X equipped with the motivic Quillen model structure [14] constructed by Morel-Voevodsky [8, p. 86 Thm. 3.2], taking the affine line $\mathbb{A}^1_X$ as interval. Given a map $f : Y \to W$ in Sm_X, we will abuse notation and denote by f the induced map $f : Y_+ \to W_+$ in $\mathcal{M}$ between the corresponding pointed simplicial presheaves represented by Y and W respectively.

We define T in $\mathcal{M}$ to be the pointed simplicial presheaf represented by $S^1 \wedge \mathbb{G}_m$, where $\mathbb{G}_m$ is the multiplicative group $\mathbb{A}^1_X - \{0\}$ pointed by 1, and S^1 denotes the simplicial circle. Given an arbitrary integer $r \geq 1$, S^r (respectively $\mathbb{G}_m^r$) will denote the iterated smash product $S^1 \wedge \cdots \wedge S^1$ (respectively $\mathbb{G}_m \wedge \cdots \wedge \mathbb{G}_m$) with r-factors; $S^0 = \mathbb{G}_m^0$ will be by definition equal to the pointed simplicial presheaf X_+ represented by the base scheme X.

Let $Spt(\mathcal{M})$ denote Jardine's category of symmetric T-spectra on $\mathcal{M}$ equipped with the motivic model structure defined in [6, Thm. 4.15] and let $\mathcal{SH}$ denote its homotopy category, which is triangulated. We will follow Jardine's notation [6, p. 506-507] where F_n denotes the left adjoint to the n-evaluation functor

$$\begin{array}{rcl} Spt(\mathcal{M}) & \xrightarrow{ev_n} & \mathcal{M} \\ (X^m)_{m\geq 0} & \longmapsto & X^n \end{array}$$

Notice that $F_0(A)$ is just the usual infinite suspension spectrum $\Sigma^\infty_T A$.

For every integer $q \in \mathbb{Z}$, we consider the following family of symmetric T-spectra

$$C^q_{eff} = \{F_n(S^r \wedge \mathbb{G}_m^s \wedge U_+) \mid n, r, s \geq 0; s - n \geq q; U \in Sm_X\} \tag{1.1}$$

where U_+ denotes the simplicial presheaf represented by U with a disjoint base point. Let $\Sigma^q_T \mathcal{SH}^{eff}$ denote the smallest full triangulated subcategory of $\mathcal{SH}$ which contains C^q_{eff} and is closed under arbitrary coproducts. Voevodsky [16] defines the slice filtration in $\mathcal{SH}$ to be the following family of triangulated subcategories

$$\cdots \subseteq \Sigma^{q+1}_T \mathcal{SH}^{eff} \subseteq \Sigma^q_T \mathcal{SH}^{eff} \subseteq \Sigma^{q-1}_T \mathcal{SH}^{eff} \subseteq \cdots$$

It follows from the work of Neeman [9], [10] that the inclusion

$$i_q : \Sigma^q_T \mathcal{SH}^{eff} \to \mathcal{SH}$$

has a right adjoint $r_q : \mathcal{SH} \to \Sigma_T^q \mathcal{SH}^{eff}$, and that the following functors

$$\begin{aligned} f_q &: \mathcal{SH} \to \mathcal{SH} \\ s_{<q} &: \mathcal{SH} \to \mathcal{SH} \\ s_q &: \mathcal{SH} \to \mathcal{SH} \end{aligned}$$

are triangulated, where f_q is defined as the composition $i_q \circ r_q$; and $s_{<q}, s_q$ are characterized by the fact that for every $E \in \mathcal{SH}$, we have distinguished triangles in $\mathcal{SH}$:

$$f_qE \xrightarrow{\theta_q^E} E \xrightarrow{\pi_{<q}^E} s_{<q}E \longrightarrow S^1 \wedge f_qE$$

$$f_{q+1}E \xrightarrow{\rho_q^E} f_qE \xrightarrow{\pi_q^E} s_qE \longrightarrow S^1 \wedge f_{q+1}E$$

We will refer to f_qE as the $(q-1)$-*connective cover* of E, to $s_{<q}E$ as the q-*orthogonal component* of E, and to s_qE as the q-*slice* of E. It follows directly from the definition that $s_{<q+1}E, s_qE$ satisfy that for every symmetric T-spectrum K in $\Sigma_T^{q+1}\mathcal{SH}^{eff}$:

$$\mathrm{Hom}_{\mathcal{SH}}(K, s_{<q+1}E) = \mathrm{Hom}_{\mathcal{SH}}(K, s_qE) = 0$$

Definition 1.1. Let $E \in Spt(\mathcal{M})$ be a symmetric T-spectrum. We will say that E is n-*orthogonal*, if for all $K \in \Sigma_T^n \mathcal{SH}^{eff}$

$$\mathrm{Hom}_{\mathcal{SH}}(K, E) = 0$$

Let $\mathcal{SH}^{\perp}(n)$ denote the full subcategory of $\mathcal{SH}$ consisting of the n-orthogonal objects.

The slice filtration admits an alternative definition in terms of (left and right) Bousfield localization of $Spt(\mathcal{M})$ [11, 12]. The Bousfield localizations are constructed following Hirschhorn's approach [2]. In order to be able to apply Hirschhorn's techniques, it is necessary to know that $Spt(\mathcal{M})$ is *cellular* [2, Def. 12.1.1] and *proper* [2, Def. 13.1.1].

Theorem 1.2. *The Quillen model category* $Spt(\mathcal{M})$ *is:*

1. cellular *(see [5], [3, Cor. 1.6] or [12, Thm. 2.7.4]).*
2. proper *(see [6, Thm. 4.15]).*

For details and definitions about Bousfield localization we refer the reader to Hirschhorn's book [2]. Let us just mention the following theorem of Hirschhorn, which guarantees the existence of left and right Bousfield localizations.

THEOREM 1.3 (see [2, Thms. 4.1.1 and 5.1.1]). *Let $\mathcal{A}$ be a Quillen model category which is cellular and proper. Let L be a set of maps in $\mathcal{A}$ and let K be a set of objects in $\mathcal{A}$. Then:*

1. *The left Bousfield localization of $\mathcal{A}$ with respect to L exists.*
2. *The right Bousfield localization of $\mathcal{A}$ with respect to the class of K-colocal equivalences exists.*

Now, we can describe the slice filtration in terms of suitable Bousfield localizations of $Spt(\mathcal{M})$.

THEOREM 1.4 (see [12]). 1. *Let $R_{C^q_{eff}}Spt(\mathcal{M})$ be the right Bousfield localization of $Spt(\mathcal{M})$ with respect to the set of objects C^q_{eff} (see Eqn. (1.1)). Then its homotopy category $R_{C^q_{eff}}\mathcal{SH}$ is triangulated and naturally equivalent to $\Sigma^q_T\mathcal{SH}^{eff}$. Moreover, the functor f_q is canonically isomorphic to the following composition of triangulated functors:*

$$\mathcal{SH} \xrightarrow{R} R_{C^q_{eff}}\mathcal{SH} \xrightarrow{C_q} \mathcal{SH}$$

where R is a fibrant replacement functor in $Spt(\mathcal{M})$, and C_q a cofibrant replacement functor in $R_{C^q_{eff}}Spt(\mathcal{M})$.

2. *Let $L_{<q}Spt(\mathcal{M})$ be the left Bousfield localization of $Spt(\mathcal{M})$ with respect to the set of maps*

$$\{F_n(S^r \wedge \mathbb{G}^s_m \wedge U_+) \to * \mid F_n(S^r \wedge \mathbb{G}^s_m \wedge U_+) \in C^q_{eff}\}$$

Then its homotopy category $L_{<q}\mathcal{SH}$ is triangulated and naturally equivalent to $\mathcal{SH}^\perp(q)$. Moreover, the functor $s_{<q}$ is canonically isomorphic to the following composition of triangulated functors:

$$\mathcal{SH} \xrightarrow{Q} L_{<q}\mathcal{SH} \xrightarrow{W_q} \mathcal{SH}$$

where Q is a cofibrant replacement functor in $Spt(\mathcal{M})$, and W_q a fibrant replacement functor in $L_{<q}Spt(\mathcal{M})$.

3. *Let $S^qSpt(\mathcal{M})$ be the right Bousfield localization of $L_{<q+1}Spt(\mathcal{M})$ with respect to the set of objects*

$$\{F_n(S^r \wedge \mathbb{G}^s_m \wedge U_+) \mid n, r, s \geq 0; s - n = q; U \in Sm_X\}$$

Then its homotopy category $S^q\mathcal{SH}$ is triangulated and the identity functor

$$id : R_{C^q_{eff}}Spt(\mathcal{M}) \to S^qSpt(\mathcal{M})$$

is a left Quillen functor. Moreover, the functor s_q is canonically isomorphic to the following composition of triangulated functors:

$$\mathcal{SH} \xrightarrow{R} R_{C^q_{eff}}\mathcal{SH} \xrightarrow{C_q} S^q\mathcal{SH} \xrightarrow{W_{q+1}} R_{C^q_{eff}}\mathcal{SH} \xrightarrow{C_q} \mathcal{SH}$$

Proof. (1) and (3) follow directly from [12, Thms. 3.3.9, 3.3.25, 3.3.50, 3.3.68]. On the other hand, (2) follows from proposition 3.2.27(3) together with theorem 3.3.26; proposition 3.3.30 and theorem 3.3.45 in [12] □

2 Birational and Weakly Birational Cohomology Theories

In this section, we construct the birational and weakly birational motivic stable homotopy categories. These are defined as left Bousfield localizations of $Spt(\mathcal{M})$ with respect to maps which are induced by open immersions with a numerical condition in the codimension of the closed complement (which is assumed to be smooth in the weakly birational case). The existence of the left Bousfield localizations considered in this section follows immediately from theorems 1.2 and 1.3.

Lemma 2.1. *Let $a, a', b, b', p, p' \geq 0$ be integers such that $a - p = a' - p'$ and $b - p = b' - p'$. Assume that $p \geq p'$, then for every $Y \in Sm_X$, there is a weak equivalence in $Spt(\mathcal{M})$, which is natural with respect to Y*

$$g^{a,b}_{p,p'}(Y) : F_p(S^a \wedge \mathbb{G}^b_m \wedge Y_+) \to F_{p'}(S^{a'} \wedge \mathbb{G}^{b'}_m \wedge Y_+)$$

Proof. We have the following adjunction (see [12, Def. 2.6.8])

$$(F_p, ev_p, \varphi) : \mathcal{M} \to Spt(\mathcal{M})$$

Using this adjunction, we define $g^{a,b}_{p,p'}(Y)$ as adjoint to the identity map:

$$\begin{aligned} S^a \wedge \mathbb{G}^b_m \wedge Y_+ \xrightarrow{id} ev_p(F_{p'}(S^{a'} \wedge \mathbb{G}^{b'}_m \wedge Y_+)) &\cong S^{p-p'} \wedge \mathbb{G}^{p-p'}_m \wedge S^{a'} \wedge \mathbb{G}^{b'}_m \wedge Y_+ \\ &\cong S^a \wedge \mathbb{G}^b_m \wedge Y_+ \end{aligned}$$

Thus, it is clear that $g^{a,b}_{p,p'}(Y)$ is natural in Y, and it follows from [12, Prop. 2.4.26] that it is a weak equivalence in $Spt(\mathcal{M})$. □

Definition 2.2 (see [13, section 7.5]). Let $Y \in Sch_X$, and Z a closed subscheme of Y. The *codimension* of Z in Y, $codim_Y Z$ is the infimum (over the generic points z_i of Z) of the dimensions of the local rings $\mathcal{O}_{Y,z_i}$.

Since X is Noetherian of finite Krull dimension and Y is of finite type over X, $codim_Y Z$ is always finite.

Definition 2.3. We fix an arbitrary integer $n \geq 0$, and consider the following set of open immersions which have a closed complement of codimension at least $n + 1$

$$\begin{aligned} B_n = \{\iota_{U,Y} : &U \to Y \text{ open immersion } | \\ &Y \in Sm_X; Y \text{ irreducible}; (codim_Y Y \backslash U) \geq n + 1\} \end{aligned}$$

The letter B stands for birational.

Now we consider the left Bousfield localization of $Spt(\mathcal{M})$ with respect to a suitable set of maps induced by the families of open immersions B_n described above.

Definition 2.4. Let $n \in \mathbb{Z}$ be an arbitrary integer.

1. Let $B_nSpt(\mathcal{M})$ denote the left Bousfield localization of $Spt(\mathcal{M})$ with respect to the set of maps

$$sB_n = \{F_p(\mathbb{G}_m^b \wedge \iota_{U,Y}) : b, p, r \geq 0, b - p \geq n - r; \iota_{U,Y} \in B_r\}.$$

2. Let $b^{(n)}$ denote its fibrant replacement functor and $\mathcal{SH}(B_n)$ its associated homotopy category.

For $n \neq 0$ we will call $\mathcal{SH}(B_n)$ the *codimension* $n+1$*-birational motivic stable homotopy category,* and for $n = 0$ we will call it the *birational motivic stable homotopy category.*

Lemma 2.5. *Let $n \in \mathbb{Z}$ be an arbitrary integer. Then for every $a \geq 0$, the maps*

$$S^a \wedge sB_n = \{F_p(S^a \wedge \mathbb{G}_m^b \wedge \iota_{U,Y}) : b, p, r \geq 0, b - p \geq n - r; \iota_{U,Y} \in B_r\}$$

are weak equivalences in $B_nSpt(\mathcal{M})$.

Proof. Let $F_p(\mathbb{G}_m^b \wedge \iota_{U,Y}) \in sB_n$ with $\iota_{U,Y} \in B_r$. Both $F_p(\mathbb{G}_m^b \wedge U_+)$ and $F_p(\mathbb{G}_m^b \wedge Y_+)$ are cofibrant in $Spt(\mathcal{M})$ (see [12, Props. 2.4.17, 2.6.18 and Thm. 2.6.30]) and hence also in $B_nSpt(\mathcal{M})$. By construction, $F_p(\mathbb{G}_m^b \wedge \iota_{U,Y})$ is a weak equivalence in $B_nSpt(\mathcal{M})$; and [2, Thm. 4.1.1.(4)] implies that $B_nSpt(\mathcal{M})$ is a simplicial model category. Thus, it follows from Ken Brown's lemma (see [4, lemma 1.1.12]) that $F_p(S^a \wedge \mathbb{G}_m^b \wedge \iota_{U,Y})$ is also a weak equivalence in $B_nSpt(\mathcal{M})$ for every $a \geq 0$. □

Proposition 2.6. *Let E be an arbitrary symmetric T-spectrum. Then E is fibrant in $B_nSpt(\mathcal{M})$ if and only if the following conditions hold:*

1. *E is fibrant in $Spt(\mathcal{M})$.*
2. *For every $a, b, p, r \geq 0$ such that $b - p \geq n - r$; and every $\iota_{U,Y} \in B_r$, the induced map*

$$\mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge Y_+), E) \xrightarrow[\cong]{\iota_{U,Y}^*} \mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge U_+), E)$$

is an isomorphism.

Proof. (⇒): Since the identity functor

$$id : Spt(\mathcal{M}) \to B_nSpt(\mathcal{M})$$

is a left Quillen functor, the conclusion follows from the derived adjunction

$$(Q, b^{(n)}, \varphi) : \mathcal{SH} \to \mathcal{SH}(B_n)$$

together with lemma 2.5.

($\Leftarrow$): Assume that E satisfies (1) and (2). Let ω_0, η_0 denote the base points of the pointed simplicial sets $\mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge Y_+), E)$ and $\mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge U_+), E)$ respectively. Since $F_p(\mathbb{G}_m^b \wedge Y_+)$ and $F_p(\mathbb{G}_m^b \wedge U_+)$ are always cofibrant, by [2, Def. 3.1.4(1)(a) and Thm. 4.1.1(2)] it is enough to show that every map in sB_n induces a weak equivalence of simplicial sets:

$$\mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge Y_+), E) \xrightarrow{\iota_{U,Y}^*} \mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge U_+), E)$$

Since $Spt(\mathcal{M})$ is a pointed simplicial model category, we observe that lemma 6.1.2 in [4] and remark 2.4.3(2) in [12] imply that the following diagram is commutative for $a \geq 0$ and all the vertical arrows are isomorphisms

$$\begin{tikzcd}
\pi_{a,\omega_0}\mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge Y_+), E) \arrow[r, "\iota_{U,Y}^*"] \arrow[d, "\cong"] & \pi_{a,\eta_0}\mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge U_+), E) \arrow[d, "\cong"] \\
\mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge Y_+), E) \arrow[r, "\iota_{U,Y}^*"] & \mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge U_+), E)
\end{tikzcd}$$

by hypothesis, the bottom row is an isomorphism, hence the top row is also an isomorphism. This implies that for every map in sB_n, the induced map

$$\mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge Y_+), E) \xrightarrow{\iota_{U,Y}^*} \mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge U_+), E)$$

is a weak equivalence when it is restricted to the path component of $\mathrm{Map}_*(F_p(\mathbb{G}_m^b \wedge Y_+), E)$ containing ω_0. This holds in particular for

$$\mathrm{Map}_*(F_{p+1}(\mathbb{G}_m^{b+1} \wedge Y_+), E) \xrightarrow{\iota_{U,Y}^*} \mathrm{Map}_*(F_{p+1}(\mathbb{G}_m^{b+1} \wedge U_+), E)$$

Therefore, the following map is a weak equivalence of pointed simplicial sets, since taking S^1-loops kills the path components that do not contain the base point

$$\begin{tikzcd}
\mathrm{Map}_*(S^1, \mathrm{Map}_*(F_{p+1}(\mathbb{G}_m^{b+1} \wedge Y_+), E)) \arrow[d] \\
\mathrm{Map}_*(S^1, \mathrm{Map}_*(F_{p+1}(\mathbb{G}_m^{b+1} \wedge U_+), E))
\end{tikzcd}$$

Now, since $Spt(\mathcal{M})$ is a simplicial model category we deduce that the rows in the following commutative diagram are isomorphisms

$$\begin{tikzcd}
\mathrm{Map}_*(S^1, \mathrm{Map}_*(F_{p+1}(\mathbb{G}_m^{b+1} \wedge Y_+), E)) \arrow[d, "\iota_{U,Y}^*"'] \arrow[r, "\cong"] & \mathrm{Map}_*(F_{p+1}(S^1 \wedge \mathbb{G}_m^{b+1} \wedge Y_+), E) \arrow[d, "\iota_{U,Y}^*"] \\
\mathrm{Map}_*(S^1, \mathrm{Map}_*(F_{p+1}(\mathbb{G}_m^{b+1} \wedge U_+), E)) \arrow[r, "\cong"] & \mathrm{Map}_*(F_{p+1}(S^1 \wedge \mathbb{G}_m^{b+1} \wedge U_+), E)
\end{tikzcd}$$

Thus, by the three out of two property for weak equivalences, we conclude that

$$\mathrm{Map}_*(F_{p+1}(S^1 \wedge \mathbb{G}_m^{b+1} \wedge Y_+), E) \xrightarrow{\iota_{U,Y}^*} \mathrm{Map}_*(F_{p+1}(S^1 \wedge \mathbb{G}_m^{b+1} \wedge U_+), E)$$

is also a weak equivalence of pointed simplicial sets. Finally, lemma 2.1 implies that the following diagram is commutative and the vertical arrows are weak equivalences in $Spt(\mathcal{M})$

$$\begin{tikzcd}
\mathrm{Map}_*(F_{p+1}(S^1 \wedge \mathbb{G}_m^{b+1} \wedge Y_+), E) \arrow[r, "\iota_{U,Y}^*"] & \mathrm{Map}_*(F_{p+1}(S^1 \wedge \mathbb{G}_m^{b+1} \wedge U_+), E) \\
\mathrm{Map}_*(F_p(\mathbb{G}_m^{b} \wedge Y_+), E) \arrow[u, "g_{p+1,p}^{1,b+1}(Y)^*"] \arrow[r, "\iota_{U,Y}^*"] & \mathrm{Map}_*(F_p(\mathbb{G}_m^{b} \wedge U_+), E) \arrow[u, "g_{p+1,p}^{1,b+1}(U)^*"']
\end{tikzcd}$$

Thus, we conclude by the two out of three property for weak equivalences that the bottom arrow is also a weak equivalence in $Spt(\mathcal{M})$. $\square$

PROPOSITION 2.7. *The homotopy category $\mathcal{SH}(B_n)$ is a compactly generated triangulated category in the sense of Neeman [9, Def. 1.7].*

Proof. We will prove first that $\mathcal{SH}(B_n)$ is a triangulated category. For this, it is enough to show that the smash product with the simplicial circle induces a Quillen equivalence (see [14, sections I.2, I.3])

$$(- \wedge S^1, \Omega_{S^1} -, \varphi) : B_n Spt(\mathcal{M}) \to B_n Spt(\mathcal{M})$$

It follows from [2, Thm. 4.1.1.(4)] that this adjunction is a Quillen adjunction, and the same argument as in [12, Cor. 3.2.38] (replacing [12, Prop. 3.2.32] with proposition 2.6) allows us to conclude that it is a Quillen equivalence. Finally, since $\mathcal{SH}$ is a compactly generated triangulated category (see [12, Prop. 3.1.5]) and the identity functor is a left Quillen functor

$$id : Spt(\mathcal{M}) \to B_n Spt(\mathcal{M})$$

it follows from the derived adjunction

$$(Q, b^{(n)}, \varphi) : \mathcal{SH} \to \mathcal{SH}(B_n)$$

that $\mathcal{SH}(B_n)$ is also compactly generated, having exactly the same set of generators as $\mathcal{SH}$. □

Definition 2.8. We fix an arbitrary integer $n \geq 0$, and consider the following set of open immersions with smooth closed complement of codimension at least $n+1$

$$\begin{aligned} WB_n = \{\iota_{U,Y} :& U \to Y \text{ open immersion} \mid \\ & Y, Z = Y\backslash U \in Sm_X; Y \text{ irreducible}; (codim_Y Z) \geq n+1\} \end{aligned}$$

Notice that every map in WB_n is also in B_n, but the converse doesn't hold. The reason to consider maps $\iota_{U,Y}$ in WB_n is that if the closed complement is smooth, then the Morel-Voevodsky homotopy purity theorem (see [8, Thm. 2.23]) characterizes the homotopy cofibre of $\iota_{U,Y}$ in terms of the Thom space of the normal bundle for the closed immersion $Y\backslash U \to Y$.

Definition 2.9. Let $n \in \mathbb{Z}$ be an arbitrary integer.

1. Let $WB_nSpt(\mathcal{M})$ denote the left Bousfield localization of $Spt(\mathcal{M})$ with respect to the set of maps

$$sWB_n = \{F_p(\mathbb{G}_m^b \wedge \iota_{U,Y}) : b, p, r \geq 0, b - p \geq n - r; \iota_{U,Y} \in WB_r\}.$$

2. Let $wb^{(n)}$ denote its fibrant replacement functor and $\mathcal{SH}(WB_n)$ its associated homotopy category.

For $n \neq 0$ we will call $\mathcal{SH}(WB_n)$ the *codimension $n+1$-weakly birational motivic stable homotopy category*, and for $n = 0$ we will call it the *weakly birational motivic stable homotopy category.*

Proposition 2.10. *Let E be an arbitrary symmetric T-spectrum. Then E is fibrant in $WB_nSpt(\mathcal{M})$ if and only if the following conditions hold:*

1. *E is fibrant in $Spt(\mathcal{M})$.*
2. *For every $a, b, p, r \geq 0$ such that $b - p \geq n - r$; and every $\iota_{U,Y} \in WB_r$, the induced map*

$$\mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge Y_+), E) \xrightarrow[\cong]{\iota_{U,Y}^*} \mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge U_+), E)$$

is an isomorphism.

Proof. The proof is exactly the same as in proposition 2.6. □

PROPOSITION 2.11. *The homotopy category $\mathcal{SH}(WB_n)$ is a compactly generated triangulated category in the sense of Neeman.*

Proof. The proof is exactly the same as in proposition 2.7. $\square$

PROPOSITION 2.12. *Assume that the base scheme $X = \operatorname{Spec} k$, with k a perfect field, then the Quillen adjunction:*

$$(id, id, \varphi) : WB_nSpt(\mathcal{M}) \to B_nSpt(\mathcal{M})$$

is a Quillen equivalence.

Proof. Consider the following commutative diagram

$$\begin{tikzcd} & Spt(\mathcal{M}) \arrow[dl, "id"'] \arrow[dr, "id"] & \\ WB_nSpt(\mathcal{M}) \arrow[rr, dashed, "id"'] & & B_nSpt(\mathcal{M}) \end{tikzcd}$$

where the solid arrows are left Quillen functors. Clearly, $WB_r \subseteq B_r$ for every $r \geq 0$, so $sWB_n \subseteq sB_n$, and we conclude that every sWB_n-local equivalence is a sB_n-local equivalence. Therefore, the universal property of left Bousfield localizations implies that the horizontal arrow is also a left Quillen functor.
The universal property for left Bousfield localizations also implies that it is enough to show that all the maps in

$$sB_n = \{F_p(\mathbb{G}_m^b \wedge \iota_{U,Y}) : b, p, r \geq 0, b - p \geq n - r; \iota_{U,Y} \in B_r\}$$

become weak equivalences in $WB_nSpt(\mathcal{M})$. Given $F_p(\mathbb{G}_m^b \wedge \iota_{U,Y}) \in sB_n$ with $\iota_{U,Y} \in B_r$, we proceed by induction on the dimension of $Z = Y \backslash U$. If $\dim Z = 0$, then $Z \in Sm_X$ since k is a perfect field (and we are considering Z with the reduced scheme structure), hence $F_p(\mathbb{G}_m^b \wedge \iota_{U,Y}) \in sWB_n$ and then a weak equivalence in $WB_nSpt(\mathcal{M})$.
If $\dim Z > 0$, then we consider the singular locus Z_s of Z over X. We have that $\dim Z_s < \dim Z$ since k is a perfect field. Therefore, by induction on the dimension $F_p(\mathbb{G}_m^b \wedge \iota_{V,Y})$ is a weak equivalence in $WB_nSpt(\mathcal{M})$, where $V = Y \backslash Z_s$. On the other hand, $F_p(\mathbb{G}_m^b \wedge \iota_{U,V})$ is also a weak equivalence in $WB_nSpt(\mathcal{M})$ since $\iota_{U,V}$ is also in B_r and its closed complement $V \backslash U = Z \backslash Z_s$ is smooth over X, by construction of Z_s.
But $F_p(\mathbb{G}_m^b \wedge \iota_{U,Y}) = F_p(\mathbb{G}_m^b \wedge \iota_{V,Y}) \circ F_p(\mathbb{G}_m^b \wedge \iota_{U,V})$, so by the two out of three property for weak equivalences we conclude that $F_p(\mathbb{G}_m^b \wedge \iota_{U,Y})$ is a weak equivalence in $WB_nSpt(\mathcal{M})$. $\square$

3 A CHARACTERIZATION OF THE SLICES

This section contains our main results. We give a characterization of the slices in terms of effectivity and birational conditions (in the sense of definition 3.1), and we also show that there is an equivalence between the notion of orthogonality (see definition 1.1) and weak birationality (see definition 3.1).

Definition 3.1. Let $E \in Spt(\mathcal{M})$ be a symmetric T-spectrum and $n \in \mathbb{Z}$.

1. We will say that E is $n{+}1$-*birational* (respectively *weakly* $n{+}1$-*birational*), if E is fibrant in $B_nSpt(\mathcal{M})$ (respectively $WB_nSpt(\mathcal{M})$). If $n = 0$, we will simply say that E is *birational* (respectively *weakly birational*).

2. We will say that E is an n-*slice* if E is isomorphic in $\mathcal{SH}$ to $s_n(E')$ for some symmetric T-spectrum E'.

Definition 3.2. 1. Let $\iota_{U,Y}$ be an open immersion in Sm_X. Let Y/U denote the pushout of the following diagram in $\mathcal{M}$ (i.e. the homotopy cofibre of $\iota_{U,Y}$ in $\mathcal{M}$)

$$\begin{tikzcd} U_+ \arrow[r, "\iota_{U,Y}"] \arrow[d] & Y_+ \arrow[d] \\ X \arrow[r] & Y/U \end{tikzcd}$$

2. Given a vector bundle $\pi : V \to Y$ with $Y \in Sm_X$, let $Th(V)$ denote the Thom space of V, i.e. $V/(V\backslash\sigma_0(Y))$, where $\sigma_0 : Y \to V$ denotes the zero section of V.

Lemma 3.3. *Let $\iota_{U,Y} \in WB_r$ for some $r \geq 0$, and let $a, b, p \geq 0$ be arbitrary integers such that $b - p \geq n - r$. Then*

$$F_p(S^a \wedge \mathbb{G}_m^b \wedge Y/U) \in \Sigma_T^{n+1}\mathcal{SH}^{eff}$$

Proof. Since $\Sigma_T^{n+1}\mathcal{SH}^{eff}$ is a triangulated category, it is enough to consider the case $a = 0$. It is also clear that it suffices to show that $F_0(Y/U) \in \Sigma_T^{r+1}\mathcal{SH}^{eff}$. Now, it follows from the Morel-Voevodsky homotopy purity theorem (see [8, Thm. 2.23]) that there is an isomorphism in $\mathcal{SH}$

$$F_0(Y/U) \to F_0(Th(N))$$

where Th(N) is the Thom space of the normal bundle N of the (smooth) complement Z of U in Y:

$$e : Y\backslash U = Z \to Y$$

But, $\iota_{U,Y} \in WB_r$; so e is a regular embedding of codimension c at least $r + 1$, hence N is a vector bundle of rank at least $r + 1$. Therefore, if N is a trivial vector bundle we conclude from [8, Prop. 2.17(2)] that

$$F_0(Th(N)) \cong F_0(S^c \wedge \mathbb{G}_m^c \wedge Z_+) \in \Sigma_T^c\mathcal{SH}^{eff} \subseteq \Sigma_T^{r+1}\mathcal{SH}^{eff}$$

Finally, we conclude in the general case by choosing a Zariski cover of Z which trivializes N and using the Mayer-Vietoris property for Zariski covers. □

Lemma 3.4. *Let $U \in Sm_X$. Consider the open immersion in Sm_X*

$$m_U : \mathbb{A}^1_U \backslash U \to \mathbb{A}^1_U$$

given by the complement of the zero section. Then $m_U \in WB_0$, and there exists a weak equivalence in $Spt(\mathcal{M})$ between its homotopy cofibre in $\mathcal{M}$, $\mathbb{A}^1_U/(\mathbb{A}^1_U\backslash U)$ and $S^1 \wedge \mathbb{G}_m \wedge U_+$

$$t_U : \mathbb{A}^1_U/(\mathbb{A}^1_U\backslash U) \to S^1 \wedge \mathbb{G}_m \wedge U_+$$

Proof. Since the zero section $i_0 : U \to \mathbb{A}^1_U$ is a closed embedding of codimension 1 between smooth schemes over X, it follows from the definition of WB_0 that $m_U \in WB_0$. Finally, [8, Prop. 2.17(2)] implies the existence of the weak equivalence t_U. □

Proposition 3.5. *Let $E \in Spt(\mathcal{M})$ be a symmetric T-spectrum and $n \in \mathbb{Z}$. Consider the following conditions:*

1. *E is fibrant in $L_{<n+1}Spt(\mathcal{M})$.*
2. *E is weakly $n+1$-birational (see definition 3.1(1)).*
3. *E is $n+1$-birational (see definition 3.1(1)).*

Then (1) and (2) are equivalent. In addition, if the base scheme $X = \operatorname{Spec} k$, with k a perfect field, then (1), (2) and (3) are equivalent.

Proof. (1)⇒(2): Assume that E is fibrant in $L_{<n+1}Spt(\mathcal{M})$. By proposition 2.10 it suffices to show that for every $a, b, p, r \geq 0$ with $b - p \geq n - r$, and every $\iota_{U,Y} \in WB_r$; the induced map

$$\mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge Y_+), E) \xrightarrow[\cong]{\iota^*_{U,Y}} \mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge U_+), E)$$

is an isomorphism. We observe that

$$F_p(S^a \wedge \mathbb{G}_m^b \wedge -) : \mathcal{M} \to Spt(\mathcal{M})$$

is a left Quillen functor, therefore the following

$$F_p(S^a \wedge \mathbb{G}_m^b \wedge U_+) \xrightarrow{F_p(S^a \wedge \mathbb{G}_m^b \wedge \iota_{U,Y})} F_p(S^a \wedge \mathbb{G}_m^b \wedge Y_+) \twoheadrightarrow F_p(S^a \wedge \mathbb{G}_m^b \wedge Y/U)$$

is a cofibre sequence in $Spt(\mathcal{M})$. However, $\mathcal{SH}$ is a triangulated category and lemma 2.1 implies that

$$F_{p+1}(S^a \wedge \mathbb{G}_m^{b+1} \wedge Y/U) \cong \Omega_{S^1} \circ R \circ F_p(S^a \wedge \mathbb{G}_m^b \wedge Y/U)$$

are isomorphic in $\mathcal{SH}$, where R denotes a fibrant replacement functor in $Spt(\mathcal{M})$. Hence it suffices to show that

$$\mathrm{Hom}_{\mathcal{SH}}(F_{p+1}(S^a \wedge \mathbb{G}_m^{b+1} \wedge Y/U), E) = \mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge Y/U), E) = 0$$

But this follows from lemma 3.3 together with [12, Prop. 3.3.30], since we are assuming that E is fibrant in $L_{<n+1}Spt(\mathcal{M})$.
(2)⇒(1) Assume that E is $n+1$-weakly birational. Then, proposition 3.3.30 in [12] implies that it suffices to show that

$$\mathrm{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge U_+), E) = 0$$

for every $F_p(S^a \wedge \mathbb{G}_m^b \wedge U_+) \in C_{e\!f\!f}^{n+1}$.
The same argument as in lemma 2.5 implies that it is enough to consider the case when $F_p(\mathbb{G}_m^b \wedge U_+) \in C_{e\!f\!f}^{n+1}$. Moreover, we can further reduce to the case where $b, p \geq 1$ and $F_p(S^1 \wedge \mathbb{G}_m^b \wedge U_+) \in C_{e\!f\!f}^{n+1}$. In effect, if $F_p(\mathbb{G}_m^b \wedge U_+) \in C_{e\!f\!f}^{n+1}$, then lemma 2.1 implies that the natural map

$$g_{p+1,p}^{1,b+1}(U) : F_{p+1}(S^1 \wedge \mathbb{G}_m^{b+1} \wedge U_+) \to F_p(\mathbb{G}_m^b \wedge U_+)$$

is a weak equivalence in $Spt(\mathcal{M})$.
Now, it follows from lemma 3.4, that if $b \geq 1$, and $0 - p + (b-1) \geq n$ (i.e. $b - p \geq n+1$); then $F_p(\mathbb{G}_m^{b-1} \wedge m_U) \in sWB_n$, i.e. a weak equivalence in $WB_nSpt(\mathcal{M})$.
Since $\mathcal{SH}(WB_n)$ is a triangulated category, $id : Spt(\mathcal{M}) \to WB_nSpt(\mathcal{M})$ is a left Quillen functor, and $F_p(\mathbb{G}_m^{b-1} \wedge (\mathbb{A}_U^1/(\mathbb{A}_U^1 \backslash U_+)))$ is the homotopy cofibre of $F_p(\mathbb{G}_m^{b-1} \wedge m_U)$; we deduce that E being $n+1$-weakly birational implies that

$$\mathrm{Hom}_{\mathcal{SH}}(F_p(\mathbb{G}_m^{b-1} \wedge (\mathbb{A}_U^1/(\mathbb{A}_U^1 \backslash U_+))), E) = 0$$

Finally, it follows from lemma 3.4 that the following groups are isomorphic

$$\begin{aligned} 0 &= \mathrm{Hom}_{\mathcal{SH}}(F_p(\mathbb{G}_m^{b-1} \wedge (\mathbb{A}_U^1/(\mathbb{A}_U^1 \backslash U_+))), E) \\ &\cong \mathrm{Hom}_{\mathcal{SH}}(F_p(S^1 \wedge \mathbb{G}_m^b \wedge U_+), E) \end{aligned}$$

(2)⇔(3): This follows directly from proposition 2.12. □

Theorem 3.6. *The Quillen adjunction*

$$(id, id, \varphi) : WB_nSpt(\mathcal{M}) \to L_{<n+1}Spt(\mathcal{M})$$

is a Quillen equivalence. In addition, if the base scheme $X = \mathrm{Spec}\, k$, with k a perfect field, then the Quillen adjunction

$$(id, id, \varphi) : B_nSpt(\mathcal{M}) \to L_{<n+1}Spt(\mathcal{M})$$

is also a Quillen equivalence.

Proof. We show first that $WB_nSpt(\mathcal{M})$ and $L_{<n+1}Spt(\mathcal{M})$ are Quillen equivalent. Since $WB_nSpt(\mathcal{M})$, $L_{<n+1}Spt(\mathcal{M})$ are both left Bousfield localizations of $Spt(\mathcal{M})$, we deduce that they are simplicial model categories with the same cofibrant replacement functor Q. Thus, it suffices to show that they have the same class of weak equivalences.

However, proposition 3.5 implies that $WB_nSpt(\mathcal{M})$, and $L_{<n+1}Spt(\mathcal{M})$ also have the same class of fibrant objects. Therefore, it follows from [2, Thm. 9.7.4] that they have exactly the same class of weak equivalences.
Finally, if the base scheme is a perfect field, by proposition 2.12 we conclude that $WB_nSpt(\mathcal{M})$ and $B_nSpt(\mathcal{M})$ are Quillen equivalent. □

Theorem 3.7. *Let E be fibrant in $Spt(\mathcal{M})$. Then E is an n-slice (see definition 3.1(2)) if and only if the following conditions hold:*

S1 *E is n-effective, i.e. $E \in \Sigma_T^n \mathcal{SH}^{eff}$.*

S2 *E is $n+1$-weakly birational.*

In addition, if the base scheme $X = \operatorname{Spec} k$, with k a perfect field, then E is an n-slice if and only if the following conditions hold:

GSS1 *E is n-effective, i.e. $E \in \Sigma_T^n \mathcal{SH}^{eff}$.*

GSS2 *E is $n+1$-birational.*

Proof. Assume that E is an n-slice. Then theorems 1.4(1) and 1.4(3) imply that E is n-effective and fibrant in $L_{<n+1}Spt(\mathcal{M})$. Hence, proposition 3.5 implies that E is also $n+1$-weakly birational.
Now we assume that E satisfies the conditions S1 and S2 above. Then, proposition 3.5 implies that E is fibrant in $L_{<n+1}Spt(\mathcal{M})$. Therefore, theorem 1.4(3) implies that E is isomorphic in $\mathcal{SH}$ to its own n-slice $s_n(E)$.
Finally, if the base scheme is a perfect field, then by proposition 3.5 the conditions S2 and GSS2 are equivalent; hence we can conclude applying the same argument as above. □

Remark 3.8. Notice that theorem 3.6 implies that it is possible to construct the slice filtration directly from the Quillen model categories $WB_nSpt(\mathcal{M})$ described in definition 2.9 without making any reference to the effective categories $\Sigma_T^q \mathcal{SH}^{eff}$. One of the interesting consequences of this fact is that it is possible to obtain analogues of the slice filtration in the unstable setting, since the suspension with respect to T or S^1 does not play an essential role in the construction of $WB_nSpt(\mathcal{M})$, i.e. we could consider the left Bousfield localization of the motivic unstable homotopy category $\mathcal{M}$ with respect to the maps in definition 2.8. We will study the details of this construction in a future work.

4 Some Computations

In this section we use the characterization of the slices obtained in theorem 3.7 to describe the slices of projective spaces, Thom spaces and blow ups.
To simplify the notation, given a simplicial presheaf $K \in \mathcal{M}$ or a map $f \in \mathcal{M}$; let $s_j(K)$, $s_j(f)$ (respectively $s_{<j}(K)$, $s_{<j}(f)$) denote $s_j(F_0(K))$, $s_j(F_0(f))$ (respectively $s_{<j}(F_0(K))$, $s_{<j}(F_0(f))$).

Lemma 4.1. *Let $g : E \to F$ be a map in $\mathcal{SH}$ such that $s_{<n}(g)$ and $s_{<n+1}(g)$ are both isomorphisms in $\mathcal{SH}$. Then the n-slice of g, $s_n(g)$ is also an isomorphism in $\mathcal{SH}$.*

Proof. It follows from [12, Prop. 3.1.19] that the rows in the following commutative diagram are distinguished triangles in $\mathcal{SH}$

$$\begin{tikzcd}
s_n(E) \arrow[r] \arrow[d, "s_n(g)"'] & s_{<n+1}(E) \arrow[r] \arrow[d, "s_{<n+1}(g)"] & s_{<n}(E) \arrow[r] \arrow[d, "s_{<n}(g)"] & S^1 \wedge s_n(E) \arrow[d, "S^1 \wedge s_n(g)"] \\
s_n(F) \arrow[r] & s_{<n+1}(F) \arrow[r] & s_{<n}(F) \arrow[r] & S^1 \wedge s_n(F)
\end{tikzcd}$$

Thus, we conclude that $s_n(g)$ is also an isomorphism in $\mathcal{SH}$. $\square$

Consider $Y \in Sm_X$. Let $\mathbb{P}^n(Y)$ denote the trivial projective bundle of rank n over Y, and let $\mathbb{P}^\infty(Y)$ denote the colimit in $\mathcal{M}$ of the following filtered diagram

$$\mathbb{P}^0(Y) \to \mathbb{P}^1(Y) \to \cdots \to \mathbb{P}^n(Y) \to \cdots \tag{4.1}$$

given by the inclusions of the respective hyperplanes at infinity.

Theorem 4.2. *Let $Y \in Sm_X$. Then for any integer $j \leq n$, the diagram 4.1 induces the following isomorphisms in $\mathcal{SH}$*

$$s_j(\mathbb{P}^n(Y)_+) \xrightarrow{\cong} s_j(\mathbb{P}^{n+1}(Y)_+) \xrightarrow{\cong} \cdots \xrightarrow{\cong} s_j(\mathbb{P}^\infty(Y)_+)$$

Proof. Let $k > n$, and consider the closed embedding induced by the diagram (4.1) $\lambda_n^k : \mathbb{P}^n(Y) \to \mathbb{P}^k(Y)$. It is possible to choose a linear embedding $\mathbb{P}^{k-n-1}(Y) \to \mathbb{P}^k(Y)$ such that its open complement $U_{k,n}$ contains $\mathbb{P}^n(Y)$ and has the structure of a vector bundle over $\mathbb{P}^n(Y)$, with zero section σ_n^k:

$$\begin{tikzcd}
U_{k,n} \arrow[r, "v_n^k"] \arrow[d, shift right] & \mathbb{P}^k(Y) & \mathbb{P}^{k-n-1}(Y) \arrow[l] \\
\mathbb{P}^n(Y) \arrow[u, shift right, "\sigma_n^k"'] \arrow[ur, bend right, "\lambda_n^k"'] & &
\end{tikzcd}$$

By homotopy invariance $s_{<j}(\sigma_n^k)$ is an isomorphism in $\mathcal{SH}$ for every integer j. On the other hand, if $j \leq n$, then $F_0(v_n^k)$ is a weak equivalence in $WB_jSpt(\mathcal{M})$ since the codimension of its closed complement is $n+1$. Thus, theorems 1.4(2) and 3.6 imply that if $j \leq n+1$, then $s_{<j}(v_n^k)$ is also an isomorphism in $\mathcal{SH}$. Therefore, $s_{<j}(\lambda_n^k) = s_{<j}(v_n^k) \circ s_{<j}(\sigma_n^k)$ is an isomorphism in $\mathcal{SH}$ for $j \leq n+1$; and using lemma 4.1 we conclude that the induced map on the slices $s_j(\lambda_n^k)$ is also an isomorphism for $j \leq n$.
Finally, the result for $\mathbb{P}^\infty(Y)$ follows directly from the fact that the slices commute with filtered homotopy colimits. $\square$

Let $H\mathbb{Z}$ denote Voevodsky's Eilenberg-MacLane spectrum (see [15, section 6.1]) representing motivic cohomology in $\mathcal{SH}$.

COROLLARY 4.3. *Assume that the base scheme $X = \operatorname{Spec} k$, with k a perfect field. Then, in the following diagram all the symmetric T-spectra are isomorphic to $H\mathbb{Z}$:*

$$H\mathbb{Z} \xrightarrow{\cong} s_0(\mathbb{P}^0(k)_+) \xrightarrow{\cong} s_0(\mathbb{P}^1(k)_+) \xrightarrow{\cong} \cdots$$
$$\cdots \xrightarrow{\cong} s_0(\mathbb{P}^n(k)_+) \xrightarrow{\cong} \cdots \xrightarrow{\cong} s_0(\mathbb{P}^\infty(k)_+)$$

Proof. This follows immediately from theorem 4.2 together with the computation of Levine [7, Thm. 10.5.1] and Voevodsky [17] for the zero slice of the sphere spectrum. □

THEOREM 4.4. *Let $\iota_{U,Y} \in WB_n$, $\pi : V \to Y$ a vector bundle of rank r together with a trivialization $t : \pi^{-1}(U) \to \mathbb{A}^r_U$ of its restriction to U. Then for every integer $j \leq n$, there exists an isomorphism in $\mathcal{SH}$ (see definition 3.2(2))*

$$s_j(Th(V)) \cong S^r \wedge \mathbb{G}_m^r \wedge s_{j-r}(Y_+)$$

Proof. Let $Z \in Sm_X$ be the closed complement of $\iota_{U,Y}$. Consider the following diagram in Sm_X, where all the squares are cartesian

$$\begin{tikzcd}
\pi^{-1}(Z) \cap (V \backslash \sigma_0(Y)) \arrow[r] \arrow[d] & V \backslash \sigma_0(Y) \arrow[d] & \pi^{-1}(U) \cap (V \backslash \sigma_0(Y)) \arrow[l, "\beta"'] \arrow[d] \\
\pi^{-1}(Z) \arrow[r] \arrow[d] & V \arrow[d, "\pi"] & \pi^{-1}(U) \arrow[l, "\alpha"'] \arrow[d] \\
Z \arrow[r] & Y & U \arrow[l, "\iota_{U,Y}"]
\end{tikzcd}$$

and let $\gamma : Th(\pi^{-1}(U)) \to Th(V)$ be the induced map between the corresponding Thom spaces. We observe that α, β also belong to WB_n; thus, if $j \leq n$ we conclude that $F_0(\iota_{U,Y}), F_0(\alpha), F_0(\beta)$ are all weak equivalences in $WB_jSpt(\mathcal{M})$. Therefore, theorems 1.4(2) and 3.6 imply that if $j \leq n+1$, then $s_{<j}(\iota_{U,Y})$, $s_{<j}(\alpha)$, $s_{<j}(\beta)$ are isomorphisms in $\mathcal{SH}$. We claim that if $j \leq n+1$, then

$$s_{<j}(\gamma) : s_{<j}(Th(\pi^{-1}(U))) \to s_{<j}(Th(V))$$

is also an isomorphism in $\mathcal{SH}$. In effect, by construction of the Thom spaces, we deduce that for any integer $j \in \mathbb{Z}$, the rows in the following commutative diagram in $\mathcal{SH}$ are in fact distinguished triangles

$$\begin{tikzcd}
s_{<j}((\pi^{-1}(U) \cap (V \backslash \sigma_0(Y)))_+) \arrow[r] \arrow[d, "s_{<j}(\beta)"'] & s_{<j}(\pi^{-1}(U)_+) \arrow[r] \arrow[d, "s_{<j}(\alpha)"'] & s_{<j}(Th(\pi^{-1}(U))) \arrow[d, "s_{<j}(\gamma)"'] \\
s_{<j}((V \backslash \sigma_0(Y))_+) \arrow[r] & s_{<j}(V_+) \arrow[r] & s_{<j}(Th(V))
\end{tikzcd}$$

Since $s_{<j}(\alpha), s_{<j}(\beta)$ are isomorphisms in $\mathcal{SH}$ for $j \leq n+1$, we conclude that for $j \leq n+1$, $s_{<j}(\gamma)$ is also an isomorphism in $\mathcal{SH}$.
Thus, lemma 4.1 implies that for $j \leq n$, $s_j(\iota_{U,Y}), s_j(\gamma)$ are isomorphisms in $\mathcal{SH}$.
Now, we use the trivialization t to obtain the following commutative diagram in Sm_X where the rows are isomorphisms

$$\begin{tikzcd}
\mathbb{A}^r_U \backslash U \arrow[d] & & \pi^{-1}(U) \cap (V \backslash \sigma_0(Y)) \arrow[ll, "\tilde{t}", "\cong"'] \arrow[d] \\
\mathbb{A}^r_U \arrow[dr] & & \pi^{-1}(U) \arrow[ll, "t", "\cong"'] \arrow[dl, "\pi_U"] \\
& U &
\end{tikzcd}$$

The same argument as above, shows that for every integer $j \in \mathbb{Z}$, there is an isomorphism in $\mathcal{SH}$

$$s_j(\bar{t}) : s_j(Th(\pi^{-1}(U))) \to s_j(Th(\mathbb{A}^r_U))$$

On the other hand, [8, Prop. 2.17(2)] implies that there is a weak equivalence $w : F_0(Th(\mathbb{A}^r_U)) \to S^r \wedge \mathbb{G}^r_m \wedge F_0(U_+)$ in $Spt(\mathcal{M})$. Thus, for $j \leq n$ there exist isomorphisms in $\mathcal{SH}$

$$\begin{tikzcd}
s_j(Th(\pi^{-1}(U))) \arrow[r, "s_j(\bar{t})"] \arrow[d, "s_j(\gamma)"'] & s_j(Th(\mathbb{A}^r_U)) \arrow[d, "s_j(w)"] \\
s_j(Th(V)) & s_j(S^r \wedge \mathbb{G}^r_m \wedge U_+)
\end{tikzcd}$$

However, there exists a canonical isomorphism in $\mathcal{SH}$

$$s_j(S^r \wedge \mathbb{G}^r_m \wedge U_+) \to S^r \wedge \mathbb{G}^r_m \wedge s_{j-r}(U_+)$$

Finally, we conclude by using the isomorphism $s_{j-r}(\iota_{U,Y})$ (notice that if $j \leq n$ then certainly $j - r \leq n$, since $r \geq 0$). □

COROLLARY 4.5. *Assume that the base scheme $X = \operatorname{Spec} k$, with k a perfect field. Let $\iota_{U,Y} \in B_n$, $\pi : V \to Y$ a vector bundle of rank r together with a trivialization $t : \pi^{-1}(U) \to \mathbb{A}^r_U$ of its restriction to U. Then for every integer $j \leq n$, there exists an isomorphism in $\mathcal{SH}$*

$$s_j(Th(V)) \cong S^r \wedge \mathbb{G}^r_m \wedge s_{j-r}(Y_+)$$

Proof. Proposition 2.12 implies that $F_0(\iota_{U,Y})$ is a weak equivalence in $WB_jSpt(\mathcal{M})$ for $j \leq n$. Hence, the result follows using exactly the same argument as in theorem 4.4. □

Given a closed embedding $Z \to Y$ of smooth schemes over X, let $\mathcal{B}\ell_Z Y$ denote the blowup of Y with center in Z.

THEOREM 4.6. *Let $\iota_{U,Y} \in WB_n$ with closed complement Z, and $j \in \mathbb{Z}$ an arbitrary integer. Consider the following cartesian square in Sm_X*

$$\begin{tikzcd} D \arrow[r, "d"] \arrow[d, "q"'] & \mathcal{B}\ell_Z Y \arrow[d, "p"] & U \arrow[l, "u"'] \arrow[d, equal] \\ Z \arrow[r, "i"'] & Y & U \arrow[l, "\iota_{U,Y}"] \end{tikzcd} \tag{4.2}$$

and let q_j, d_j, p_j, i_j denote $s_j(q), s_j(d), s_j(p), s_j(i)$ respectively. Then the cartesian square (4.2) induces the following distinguished triangle in $\mathcal{SH}$

$$s_j(D_+) \xrightarrow{\binom{-d_j}{q_j}} s_j(\mathcal{B}\ell_Z Y_+) \oplus s_j(Z_+) \xrightarrow{(p_j, i_j)} s_j(Y_+) \tag{4.3}$$

If $j \leq n$, then $s_j(\iota_{U,Y})$ is an isomorphism in $\mathcal{SH}$, and the following distinguished triangles in $\mathcal{SH}$ split

$$\begin{tikzcd} s_j(D_+) \arrow[r, shift left, "\binom{-d_j}{q_j}"] & s_j(\mathcal{B}\ell_Z Y_+) \oplus s_j(Z_+) \arrow[l, shift left] \arrow[r, shift left, "(p_j, i_j)"] & s_j(Y_+) \arrow[l, shift left, "\binom{r_j}{0}"] \end{tikzcd} \tag{4.4}$$

$$\begin{tikzcd} s_j(Y_+) \arrow[r, shift left, "r_j"] & s_j(\mathcal{B}\ell_Z Y_+) \arrow[l, shift left, "p_j"] \arrow[r, shift left] & s_j(Th(\mathcal{O}_D(1))) \arrow[l, shift left] \end{tikzcd} \tag{4.5}$$

where $r_j = s_j(u) \circ (s_j(\iota_{U,Y}))^{-1}$, and $\mathcal{O}_D(1)$ denotes the canonical line bundle of the projective bundle $q : D \to Z$.

Proof. It follows from [8, Prop. 2.29 and Rmk. 2.30] that the following square is homotopy cocartesian in $\mathcal{M}$

$$\begin{tikzcd} S^1 \wedge D_+ \arrow[r, "id \wedge d"] \arrow[d, "id \wedge q"'] & S^1 \wedge \mathcal{B}\ell_Z Y_+ \arrow[d, "id \wedge p"] \\ S^1 \wedge Z_+ \arrow[r, "id \wedge i"'] & S^1 \wedge Y_+ \end{tikzcd}$$

Thus, we deduce that the following diagram is a distinguished triangle in $\mathcal{SH}$

$$F_0(D_+) \xrightarrow{\binom{-F_0(d)}{F_0(q)}} F_0(\mathcal{B}\ell_Z Y_+) \oplus F_0(Z_+) \xrightarrow{(F_0(p), F_0(i))} F_0(Y_+)$$

Since the slices s_j are triangulated functors, it follows that diagram (4.3) is a distinguished triangle in $\mathcal{SH}$.

Now, we prove that $s_j(\iota_{U,Y})$ is an isomorphism for $j \leq n$. By lemma 4.1, it suffices to show that $s_{<j}(\iota_{U,Y})$ is an isomorphism in $\mathcal{SH}$ for $j \leq n+1$. But this follows directly from theorems 3.6 and 1.4(2) since $F_0(\iota_{U,Y})$ is clearly a weak equivalence in $WB_jSpt(\mathcal{M})$ for $j \leq n$.
Thus, r_j is well defined for $j \leq n$, and the following diagram shows that it gives a splitting for the distinguished triangle (4.4)

$$\begin{tikzcd} s_j(U_+) \arrow[r, "s_j(u)"] \arrow[d, equal] & s_j(\mathcal{B}\ell_Z Y_+) \arrow[d, "p_j"] \\ s_j(U_+) \arrow[r, "s_j(\iota_{U,Y})"'] & s_j(Y_+) \end{tikzcd} \tag{4.6}$$

Finally, since the normal bundle of the closed embedding $d : D \to \mathcal{B}\ell_Z Y$ is given by $\mathcal{O}_D(1)$, we deduce from the Morel-Voevodsky homotopy purity theorem (see [8, Thm. 2.23]) that the following diagram is a distinguished triangle in $\mathcal{SH}$

$$s_j(U_+) \xrightarrow{s_j(u)} s_j(\mathcal{B}\ell_Z Y_+) \longrightarrow s_j(Th(\mathcal{O}_D(1)))$$

Combining this distinguished triangle with diagram (4.6) above, we conclude that diagram (4.5) is a split distinguished triangle in $\mathcal{SH}$ for $j \leq n$. $\square$

Acknowledgements

The author would like to thank Marc Levine for bringing to his attention the connection between slices and birational cohomology theories, as well as for all his stimulating comments and questions, and also thank Chuck Weibel for his interest in this work and several suggestions which helped to improve the exposition.

References

[1] Bruno Kahn and R. Sujatha. Birational motives, I. Preprint, October 30, 2002, K-theory Preprint Archives, http://www.math.uiuc.edu/K-theory/0596/.

[2] P. S. Hirschhorn. *Model categories and their localizations*, volume 99 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2003.

[3] J. Hornbostel. Localizations in motivic homotopy theory. *Math. Proc. Cambridge Philos. Soc.*, 140(1):95–114, 2006.

[4] M. Hovey. *Model categories*, volume 63 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 1999.

[5] M. Hovey. Spectra and symmetric spectra in general model categories. *J. Pure Appl. Algebra*, 165(1):63–127, 2001.

[6] J. F. Jardine. Motivic symmetric spectra. *Doc. Math.*, 5:445–553 (electronic), 2000.

[7] M. Levine. The homotopy coniveau tower. *J. Topol.*, 1(1):217–267, 2008.

[8] F. Morel and V. Voevodsky. $\mathbf{A}^1$-homotopy theory of schemes. *Inst. Hautes Études Sci. Publ. Math.*, (90):45–143 (2001), 1999.

[9] A. Neeman. The Grothendieck duality theorem via Bousfield's techniques and Brown representability. *J. Amer. Math. Soc.*, 9(1):205–236, 1996.

[10] A. Neeman. *Triangulated categories*, volume 148 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2001.

[11] P. Pelaez. Mixed motives and the slice filtration. *C. R. Math. Acad. Sci. Paris*, 347(9-10):541–544, 2009.

[12] P. Pelaez. Multiplicative properties of the slice filtration. *Astérisque*, (335):xvi+289, 2011.

[13] D. Quillen. Higher algebraic K-theory. I. In *Algebraic K-theory, I: Higher K-theories (Proc. Conf., Battelle Memorial Inst., Seattle, Wash., 1972)*, pages 85–147. Lecture Notes in Math., Vol. 341. Springer, Berlin, 1973.

[14] D. G. Quillen. *Homotopical algebra*. Lecture Notes in Mathematics, No. 43. Springer-Verlag, Berlin, 1967.

[15] V. Voevodsky. $\mathbf{A}^1$-homotopy theory. In *Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998)*, number Extra Vol. I, pages 579–604 (electronic), 1998.

[16] V. Voevodsky. Open problems in the motivic stable homotopy theory. I. In *Motives, polylogarithms and Hodge theory, Part I (Irvine, CA, 1998)*, volume 3 of *Int. Press Lect. Ser.*, pages 3–34. Int. Press, Somerville, MA, 2002.

[17] V. Voevodsky. On the zero slice of the sphere spectrum. *Tr. Mat. Inst. Steklova*, 246(Algebr. Geom. Metody, Svyazi i Prilozh.):106–115, 2004.

Pablo Pelaez
Department of Mathematics
Rutgers University, U.S.A.
pablo.pelaez@rutgers.edu

Coherence for Weak Units

André Joyal and Joachim Kock

Received: September 21, 2011

Communicated by Stefan Schwede

Abstract. We define weak units in a semi-monoidal 2-category $\mathscr{C}$ as cancellable pseudo-idempotents: they are pairs (I, α) where I is an object such that tensoring with I from either side constitutes a biequivalence of $\mathscr{C}$, and $\alpha : I \otimes I \to I$ is an equivalence in $\mathscr{C}$. We show that this notion of weak unit has coherence built in: Theorem A: α has a canonical associator 2-cell, which automatically satisfies the pentagon equation. Theorem B: every morphism of weak units is automatically compatible with those associators. Theorem C: the 2-category of weak units is contractible if non-empty. Finally we show (Theorem E) that the notion of weak unit is equivalent to the notion obtained from the definition of tricategory: α alone induces the whole family of left and right maps (indexed by the objects), as well as the whole family of Kelly 2-cells (one for each pair of objects), satisfying the relevant coherence axioms.

2010 Mathematics Subject Classification: 18D05; 18D10
Keywords and Phrases: Monoidal 2-categories, units, coherence.

Introduction

The notion of tricategory, introduced by Gordon, Power, and Street [2] in 1995, seems still to represent the highest-dimensional explicit weak categorical structure that can be manipulated by hand (i.e. without methods of homotopy theory), and is therefore an important test bed for higher-categorical ideas. In this work we investigate the nature of weak units at this level. While coherence for weak associativity is rather well understood, thanks to the geometrical insight provided by the Stasheff associahedra [12], coherence for unit structures is more mysterious, and so far there seems to be no clear geometric pattern for the coherence laws for units in higher dimensions. Specific interest in weak units stems from Simpson's conjecture [11], according to which strict n-groupoids with weak units should model all homotopy n-types.

In the present paper, working in the setting of a strict 2-category $\mathscr{C}$ with a strict tensor product, we define a notion of weak unit by simple axioms that involve only the notion of equivalence, and hence in principle make sense in all dimensions. Briefly, a weak unit is a cancellable pseudo-idempotent. We work out the basic theory of such units, and compare with the notion extracted from the definition of tricategory. In the companion paper *Weak units and homotopy 3-types* [4] we employ this notion of unit to prove a version of Simpson's conjecture for 1-connected homotopy 3-types, which is the first nontrivial case. The strictness assumptions of the present paper should be justified by that result.

By cancellable pseudo-idempotent we mean a pair (I, α) where I is an object in $\mathscr{C}$ such that tensoring with I from either side is an equivalence of 2-categories, and $\alpha : I \otimes I \xrightarrow{\sim} I$ is an equi-arrow (i.e. an arrow admitting a pseudo-inverse). The remarkable fact about this definition is that α, viewed as a multiplication map, comes with canonical higher order data built in: it possesses a canonical associator A which automatically satisfies the pentagon equation. This is our Theorem A. The point is that the arrow α alone, thanks to the cancellability of I, induces all the usual structure of left and right constraints with all the 2-cell data that goes into them and the axioms they must satisfy.

As a warm-up to the various constructions and ideas, we start out in Section 1 by briefly running through the corresponding theory for cancellable-idempotent units in monoidal 1-categories. This theory has been treated in detail in [8].

The rest of the paper is dedicated to the case of monoidal 2-categories. In Section 2 we give the definitions and state the main results: Theorem A says that there is a canonical associator 2-cell for α, and that this 2-cell automatically satisfies the pentagon equation. Theorem B states that unit morphisms automatically are compatible with the associators of Theorem A. Theorem C states that the 2-category of units is contractible if non-empty. Hence, 'being unital' is, up to homotopy, a property rather than a structure.

Next follow three sections dedicated to proofs of each of these three theorems. In Section 3 we show how the map $\alpha : II \xrightarrow{\sim} I$ alone induces left and right constraints, which in turn are used to construct the associator and establish the pentagon equation. The left and right constraints are not canonical, but surprisingly the associator does not depend on the choice of them. In Section 4 we prove Theorem B by interpreting it as a statement about units in the 2-category of arrows, where it is possible to derive it from Theorem A. In Section 5 we prove Theorem C. The key ingredient is to use the left and right constraints to link up all the units, and to show that the unit morphisms are precisely those compatible with the left and right constraints; this makes them 'essentially unique' in the required sense.

In Section 6 we go through the basic theory of classical units (i.e. as extracted from the definition of tricategory [2]). Finally, in Section 7 we show that the two notions of unit are equivalent. This is our Theorem E. A curiosity implied by the arguments in this section is that the left and right axioms for the 2-cell data in the Gordon-Power-Street definition (denoted TA2 and TA3 in [2]) imply each other.
(We have no Theorem D.)

This notion of weak units as cancellable idempotents is precisely what can be extracted from the more abstract, Tamsamani-style, theory of fair n-categories [7] by making an arbitrary choice of a fixed weak unit. In the theory of fair categories, the key object is a contractible space of all weak units, rather than any particular point in that space, and handling this space as a whole bypasses coherence issues. However, for the sake of understanding what the theory entails, and for the sake of concrete computations, it is interesting to make a choice and study the ensuing coherence issues, as we do in this paper. The resulting approach is very much in the spirit of the classical theory of monoidal categories, bicategories, and tricategories, and provides some new insight to these theories. To stress this fact we have chosen to formulate everything from scratch in such classical terms, without reference to the theory of fair categories.
In the case of monoidal 1-categories, the cancellable-idempotent viewpoint on units goes back to Saavedra [10]. The importance of this viewpoint in higher categories was first suggested by Simpson [11], in connection with his weak-unit conjecture. He gave an ad hoc definition in this style, as a mere indication of what needed to be done, and raised the question of whether higher homotopical data would have to be specified. The surprising answer is, at least here in dimension 3, that specifying α is enough, then the higher homotopical data is automatically built in.

Acknowledgements. We thank Georges Maltsiniotis for pointing out to us that the cancellable-idempotent notion of unit in dimension 1 goes back to Saavedra [10], and we thank Josep Elgueta for catching an error in an earlier version of our comparison with tricategories. The first-named author was supported by the NSERC. The second-named author was very happy to be a CIRGET postdoc at the UQAM in 2004, and currently holds support from grants MTM2009-10359 and MTM2010-20692 of Spain.

1 Units in monoidal categories

It is helpful first briefly to recall the relevant results for monoidal categories, referring the reader to [8] for further details of this case.

1.1. Semi-monoidal categories. A *semi-monoidal category* is a category $\mathscr{C}$ equipped with a tensor product (which we denote by plain juxtaposition), i.e. an associative functor

$$\begin{array}{ccc} \mathscr{C} \times \mathscr{C} & \longrightarrow & \mathscr{C} \\ (X,Y) & \longmapsto & XY. \end{array}$$

For simplicity we assume strict associativity, $X(YZ) = (XY)Z$.

1.2. Monoidal categories. (Mac Lane [9].) A semi-monoidal category $\mathscr{C}$ is a *monoidal category* when it is furthermore equipped with a distinguished object I and natural isomorphisms

$$IX \xrightarrow{\lambda_X} X \xleftarrow{\rho_X} XI$$

obeying the following rules (cf. [9]):

$$\lambda_I = \rho_I \tag{1}$$
$$\lambda_{XY} = \lambda_X Y \tag{2}$$
$$\rho_{XY} = X\rho_Y \tag{3}$$
$$X\lambda_Y = \rho_X Y \tag{4}$$

Naturality of λ and ρ implies

$$\lambda_{IX} = I\lambda_X, \qquad \rho_{XI} = \rho_X I, \tag{5}$$

independently of Axioms (1)–(4).

1.3 Remark. Tensoring with I from either side is an equivalence of categories.

1.4 Lemma. (Kelly [5].) *Axiom (4) implies axioms (1), (2), and (3).*

Proof. (4) implies (2): Since tensoring with I on the left is an equivalence, it is enough to prove $I\lambda_{XY} = I\lambda_X Y$. But this follows from Axiom (4) applied twice (swap λ out for a ρ and swap back again only on the nearest factor):

$$I\lambda_{XY} = \rho_I XY = I\lambda_X Y.$$

Similarly for ρ, establishing (3).
(4) and (2) implies (1): Since tensoring with I on the right is an equivalence, it is enough to prove $\lambda_I I = \rho_I I$. But this follows from (2), (5), and (4):

$$\lambda_I I = \lambda_{II} = I\lambda_I = \rho_I I. \qquad \square$$

The following alternative notion of unit object goes back to Saavedra [10]. A thorough treatment of the notion was given in [8].

1.5. Units as cancellable pseudo-idempotents. An object I in a semi-monoidal category $\mathscr{C}$ is called *cancellable* if the two functors $\mathscr{C} \to \mathscr{C}$

$$\begin{array}{rcl} X & \longmapsto & IX \\ X & \longmapsto & XI \end{array}$$

are fully faithful. By definition, a *pseudo-idempotent* is an object I equipped with an isomorphism $\alpha : II \xrightarrow{\sim} I$. Finally we define a *unit object* in $\mathscr{C}$ to be a cancellable pseudo-idempotent.

1.6 Lemma. [8] *Given a unit object (I, α) in a semi-monoidal category $\mathscr{C}$, for each object X there are unique arrows*

$$IX \xrightarrow{\lambda_X} X \xleftarrow{\rho_X} XI$$

such that

$$\begin{array}{ll} (\mathsf{L}) & I\lambda_X = \alpha X \\ (\mathsf{R}) & \rho_X I = X\alpha. \end{array}$$

The λ_X and ρ_X are isomorphisms and natural in X.

Proof. Let $\mathbb{L} : \mathscr{C} \to \mathscr{C}$ denote the functor defined by tensoring with I on the left. Since $\mathbb{L}$ is fully faithful, we have a bijection

$$\mathrm{Hom}(IX, X) \to \mathrm{Hom}(IIX, IX).$$

Now take λ_X to be the inverse image of αX; it is an isomorphism since αX is. Naturality follows by considering more generally the bijection

$$\mathrm{Nat}(\mathbb{L}, \mathrm{id}_{\mathscr{C}}) \to \mathrm{Nat}(\mathbb{L} \circ \mathbb{L}, \mathbb{L});$$

let λ be the inverse image of the natural transformation whose components are αX. Similarly on the right. $\square$

1.7 Lemma. [8] *For λ and ρ as above, the Kelly axiom* (4) *holds:*

$$X\lambda_Y = \rho_X Y.$$

Therefore, by Lemma 1.6 a semi-monoidal category with a unit object is a monoidal category in the classical sense.

Proof. In the commutative square

$$\begin{tikzcd} XIIY \arrow[r, "XI\lambda_Y"] \arrow[d, "\rho_X IY"'] & XIY \arrow[d, "\rho_X Y"] \\ XIY \arrow[r, "X\lambda_Y"'] & XY \end{tikzcd}$$

the top arrow is equal to $X\alpha Y$, by X tensor (L), and the left-hand arrow is also equal to $X\alpha Y$, by (R) tensor Y. Since $X\alpha Y$ is an isomorphism, it follows that $X\lambda_Y = \rho_X Y$. □

1.8 Lemma. *For a unit object (I, α) we have: (i) The map $\alpha : II \to I$ is associative. (ii) The two functors $X \mapsto IX$ and $X \mapsto XI$ are equivalences.*

Proof. Since α is invertible, associativity amounts to the equation $I\alpha = \alpha I$, which follows from the previous proof by setting $X = Y = I$ and applying L and R once again. To see that $\mathbb{L}$ is an equivalence, just note that it is isomorphic to the identity via λ. □

1.9. Uniqueness of units. Just as in a semi-monoid a unit element is unique if it exists, one can show [8, 2.20] that in a semi-monoidal category, between any two units there is a unique isomorphism of units. This statement does not involve λ and ρ, but the proof does: the canonical isomorphism $I \overset{\sim}{\to} J$ is the composite $I \xrightarrow{\rho_I^{-1}} IJ \xrightarrow{\lambda_J} J$.

2 Units in monoidal 2-categories: definition and main results

In this section we set up the necessary terminology and notation, give the main definition, and state the main results.

2.1. 2-categories. We work in a strict 2-category $\mathscr{C}$. We use the symbol $\#$ to denote composition of arrows and horizontal composition of 2-cells in $\mathscr{C}$, always written from the left to the right, and occasionally decorating the symbol $\#$ by the name of the object where the two arrows or 2-cells are composed. By an *equi-arrow* in $\mathscr{C}$ we understand an arrow f admitting an (unspecified) pseudo-inverse, i.e. an arrow g in the opposite direction such that $f\#g$ and $g\#f$ are isomorphic to the respective identity arrows, and such that the comparison 2-cells satisfy the usual triangle equations for adjunctions. (The usual word for 'equi-arrow' is of course 'equivalence'; we reserve the latter word for equivalence of categories and 2-categories. We find it useful to have a different word for the equivalences inside a 2-category.) It is worth pointing out that it is not necessary to insist on the triangle equations. If the 2-cells exist but do not satisfy the triangle equation, they can always be replaced by 2-cells that do. We shall make extensive use of arguments with pasting diagrams [6]. Our drawings of 2-cells should be read from bottom to top, so that for example

$$\begin{tikzcd}
X \arrow[rr, "h"] \arrow[dr, "f"'] & {} \arrow[d, phantom, "\mathsf{U}"] & Z \\
& Y \arrow[ur, "g"'] &
\end{tikzcd}$$

denotes $\mathsf{U} : f \underset{Y}{\#} g \Rightarrow h$. The symbol © will denote identity 2-cells.

The few 2-functors we need all happen to be strict. By *natural transformation* we always mean pseudo-natural transformation. Hence a natural transformation $u : F \Rightarrow G$ between two 2-functors from $\mathscr{D}$ to $\mathscr{C}$ is given by an arrow $u_X : FX \to GX$ for each object $X \in \mathscr{D}$, and an invertible 2-cell

$$\begin{tikzcd} FX \arrow[r, "u_X"] \arrow[d, "F(x)"'] \arrow[dr, phantom, "u_x"] & GX \arrow[d, "G(x)"] \\ FX' \arrow[r, "u_{X'}"'] & GX' \end{tikzcd}$$

for each arrow $x : X \to X'$ in $\mathscr{D}$, subject to the usual compatibility conditions [6]. The modifications we shall need will happen to be invertible.

2.2. Semi-monoidal 2-categories. By *semi-monoidal* 2*-category* we mean a 2-category $\mathscr{C}$ equipped with a tensor product, i.e. an associative 2-functor

$$\begin{array}{rcl} \otimes : \mathscr{C} \times \mathscr{C} & \longrightarrow & \mathscr{C} \\ (X,Y) & \longmapsto & XY, \end{array}$$

denoted by plain juxtaposition. We already assumed $\mathscr{C}$ to be a strict 2-category, and we also require $\otimes$ to be a strict 2-functor and to be strictly associative: $(XY)Z = X(YZ)$. This is mainly for convenience, to keep the focus on unit issues.

Note that the tensor product of two equi-arrows is again an equi-arrow, since its pseudo-inverse can be taken to be the tensor product of the pseudo-inverses.

2.3. Semi-monoids. A *semi-monoid* in $\mathscr{C}$ is a triple $(X, \alpha, \widetilde{\mathsf{A}})$ consisting of an object X, a multiplication map $\alpha : XX \to X$, and an invertible 2-cell $\widetilde{\mathsf{A}}$ called the *associator*,

$$\begin{tikzcd} XXX \arrow[r, "\alpha X"] \arrow[d, "X\alpha"'] \arrow[dr, phantom, "\widetilde{\mathsf{A}}"] & XX \arrow[d, "\alpha"] \\ XX \arrow[r, "\alpha"'] & X \end{tikzcd}$$

required to satisfy the 'pentagon equation':

$$\begin{tikzcd} XXXX \arrow[r, "\alpha XX"] \arrow[d, "XX\alpha"'] \arrow[dr, "X\alpha X"] & XXX \arrow[dr, "\alpha X"] \arrow[d, phantom, "\widetilde{\mathsf{A}}X"] & \\ XXX \arrow[dr, "X\alpha"] \arrow[r, phantom, "X\widetilde{\mathsf{A}}"] & XXX \arrow[r, "\alpha X"] \arrow[d, "X\alpha"] \arrow[dr, phantom, "\widetilde{\mathsf{A}}"] & XX \arrow[d, "\alpha"] \\ & XX \arrow[r, "\alpha"'] & X \end{tikzcd} \quad = \quad \begin{tikzcd} XXXX \arrow[r, "\alpha XX"] \arrow[d, "XX\alpha"'] \arrow[dr, phantom, "\copyright"] & XXX \arrow[d, "X\alpha"] \arrow[dr, "\alpha X"] & \\ XXX \arrow[r, "\alpha X"'] \arrow[dr, "X\alpha"] & XX \arrow[dr, "\alpha"] \arrow[d, phantom, "\widetilde{\mathsf{A}}"] \arrow[r, phantom, "\widetilde{\mathsf{A}}"] & XX \arrow[d, "\alpha"] \\ & XX \arrow[r, "\alpha"'] & X \end{tikzcd}$$

In the applications, α will be an equi-arrow, and hence we will have

$$\widetilde{\mathsf{A}} = \mathsf{A} \underset{XX}{\#} \alpha$$

for a some unique invertible

$$\mathsf{A} : X\alpha \Rightarrow \alpha X,$$

which it will more convenient to work with. In this case, the pentagon equation is equivalent to the more compact equation

$$\begin{tikzcd}
XXXX \arrow[r, "\alpha XX"] \arrow[d, bend right=50, "XX\alpha"'] \arrow[d, bend left=50, "X\alpha X"] \arrow[d, phantom, "X\mathsf{A}"] \arrow[dr, phantom, "\widetilde{\mathsf{A}}X"] & XXX \arrow[d, bend left=30, "\alpha X"] \\
XXX \arrow[r, "\alpha X"'] & XX
\end{tikzcd}
=
\begin{tikzcd}
XXXX \arrow[r, "\alpha XX"] \arrow[d, bend right=30, "XX\alpha"'] \arrow[dr, phantom, "\copyright"] & XXX \arrow[d, bend right=50, "X\alpha"'] \arrow[d, bend left=50, "\alpha X"] \arrow[d, phantom, "\mathsf{A}"] \\
XXX \arrow[r, "\alpha X"'] & XX
\end{tikzcd} \tag{6}$$

which we shall also make use of.

2.4. Semi-monoid maps. A *semi-monoid map* $f : (X, \alpha, \widetilde{\mathsf{A}}) \to (Y, \beta, \widetilde{\mathsf{B}})$ is the data of an arrow $f : X \to Y$ in $\mathscr{C}$ together with an invertible 2-cell

$$\begin{tikzcd}
XX \arrow[r, "ff"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{F}"] & YY \arrow[d, "\beta"] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}$$

such that this cube commutes:

$$\begin{tikzcd}
 & YYY \arrow[r, "\beta Y"] \arrow[d, phantom, "\mathsf{F}f"] & YY \arrow[d, "\beta"] \\
XXX \arrow[ur, "fff"] \arrow[r, "\alpha X"] \arrow[d, "X\alpha"'] \arrow[dr, phantom, "\widetilde{\mathsf{A}}"] & XX \arrow[ur, "ff"] \arrow[d, "\alpha"] \arrow[r, phantom, "\mathsf{F}"] & Y \\
XX \arrow[r, "\alpha"'] & X \arrow[ur, "f"] &
\end{tikzcd}
=
\begin{tikzcd}
 & YYY \arrow[r, "\beta Y"] \arrow[d, "Y\beta"] \arrow[dr, phantom, "\widetilde{\mathsf{B}}"] & YY \arrow[d, "\beta"] \\
XXX \arrow[ur, "fff"] \arrow[d, "X\alpha"'] \arrow[r, phantom, "f\mathsf{F}"] & YY \arrow[r, "\beta"'] \arrow[d, phantom, "\mathsf{F}"] & Y \\
XX \arrow[ur, "ff"] \arrow[r, "\alpha"'] & X \arrow[ur, "f"] &
\end{tikzcd}$$

When β is an equi-arrow, the cube equation is equivalent to the simpler equation:

$$
\begin{tikzcd}
XXX \arrow[r, "fff"] \arrow[d, bend right, "X\alpha"'] \arrow[d, bend left, "\alpha X"] \arrow[d, phantom, "\mathsf{A}"] \arrow[dr, phantom, "\mathsf{F}f"] & YYY \arrow[d, bend left, "\beta Y"] \\
XX \arrow[r, "ff"'] & YY
\end{tikzcd}
=
\begin{tikzcd}
XXX \arrow[r, "fff"] \arrow[d, bend right, "X\alpha"'] \arrow[dr, phantom, "f\mathsf{F}"] & YYY \arrow[d, bend right, "Y\beta"'] \arrow[d, bend left, "\beta Y"] \arrow[d, phantom, "\mathsf{B}"] \\
XX \arrow[r, "ff"'] & YY
\end{tikzcd}
\qquad (7)
$$

which will be useful.

2.5. Semi-monoid transformations. A *semi-monoid transformation* between two parallel semi-monoid maps (f, F) and (g, G) is a 2-cell $\mathsf{T} : f \Rightarrow g$ in $\mathscr{C}$ such that this cylinder commutes:

$$
\begin{tikzcd}
XX \arrow[r, bend left, "gg"] \arrow[r, bend right, "ff"'] \arrow[r, phantom, "\mathsf{TT}"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{F}"] & YY \arrow[d, "\beta"] \\
X \arrow[r, bend right, "f"'] & Y
\end{tikzcd}
=
\begin{tikzcd}
XX \arrow[r, bend left, "gg"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{G}"] & YY \arrow[d, "\beta"] \\
X \arrow[r, bend left, "g"] \arrow[r, bend right, "f"'] \arrow[r, phantom, "\mathsf{T}"] & Y
\end{tikzcd}
$$

2.6 Lemma. *Let $f : X \to Y$ be a semi-monoid map. If f is an equi-arrow (as an arrow in $\mathscr{C}$) with quasi-inverse $g : Y \to X$, then there is a canonical 2-cell G such that (g, G) is a semi-monoid map.*

Proof. The 2-cell G is defined as the mate [6] of the 2-cell F^{-1}. It is routine to check the cube equation in 2.4. $\square$

2.7. Pseudo-idempotents. A *pseudo-idempotent* is a pair (I, α) where $\alpha : II \to I$ is an equi-arrow. A *morphism of pseudo-idempotents* from (I, α) to (J, β) is a pair (u, U) consisting of an arrow $u : I \to J$ in $\mathscr{C}$ and an invertible 2-cell

$$
\begin{tikzcd}
II \arrow[r, "uu"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{U}"] & JJ \arrow[d, "\beta"] \\
I \arrow[r, "u"'] & J.
\end{tikzcd}
$$

If (u, U) and (v, V) are morphisms of pseudo-idempotents from (I, α) to (J, β), a 2-*morphism of pseudo-idempotents* from (u, U) to (v, V) is a 2-cell $\mathsf{T} : u \Rightarrow v$ satisfying the cylinder equation of 2.5.

2.8. Cancellable objects. An object I in $\mathscr{C}$ is called *cancellable* if the two 2-functors $\mathscr{C} \to \mathscr{C}$

$$\begin{array}{ccc} X & \longmapsto & IX \\ X & \longmapsto & XI \end{array}$$

are fully faithful. (Fully faithful means that the induced functors on hom categories are equivalences.) A *cancellable morphism* between cancellable objects I and J is an equi-arrow $u : I \to J$. (Equivalently it is an arrow such that the functors on hom cats defined by tensoring with u on either side are equivalences, cf. 5.1.) A *cancellable* 2*-morphism* between cancellable arrows is any invertible 2-cell.

We are now ready for the main definition and the main results.

2.9. Units. A *unit object* is by definition a cancellable pseudo-idempotent. Hence it is a pair (I, α) consisting of an object I and an equi-arrow $\alpha : II \to I$, with the property that tensoring with I from either side define fully faithful 2-functors $\mathscr{C} \to \mathscr{C}$.
A *morphism* of units is a cancellable morphism of pseudo-idempotents. In other words, a unit morphism from (I, α) to (J, β) is a pair (u, U) where $u : I \to J$ is an equi-arrow and U is an invertible 2-cell

$$\begin{tikzcd} II \arrow[r, "uu"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{U}"] & JJ \arrow[d, "\beta"] \\ I \arrow[r, "u"'] & J. \end{tikzcd}$$

A 2*-morphism* of units is a cancellable 2-morphism of pseudo-idempotents. Hence a 2-morphism from (u, U) to (v, V) is an invertible 2-cell $\mathsf{T} : u \Rightarrow v$ such that

$$\begin{tikzcd} II \arrow[r, bend left, "vv"] \arrow[r, bend right, "uu"'] \arrow[r, phantom, "\mathsf{TT}"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{U}"] & JJ \arrow[d, "\beta"] \\ I \arrow[r, bend right, "u"'] & J \end{tikzcd} \quad = \quad \begin{tikzcd} II \arrow[r, bend left, "vv"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{V}"] & JJ \arrow[d, "\beta"] \\ I \arrow[r, bend left, "v"] \arrow[r, bend right, "u"'] \arrow[r, phantom, "\mathsf{T}"] & J \end{tikzcd}$$

This defines the 2*-category of units.*

In the next section we'll see how the notion of unit object induces left and right constraints familiar from standard notions of monoidal 2-category. It will then turn out (Lemmas 5.1 and 5.2) that unit morphisms and 2-morphisms can be characterised as those morphisms and 2-morphisms compatible with the left and right constraints.

Theorem A (Associativity). *Given a unit object (I, α), there is a canonical invertible 2-cell*

$$
\begin{tikzcd}
III \arrow[r, "\alpha I"] \arrow[d, "I\alpha"'] \arrow[dr, phantom, "\widetilde{\mathsf{A}}"] & II \arrow[d, "\alpha"] \\
II \arrow[r, "\alpha"'] & I
\end{tikzcd}
$$

which satisfies the pentagon equation

$$
\begin{tikzcd}
IIII \arrow[r, "\alpha II"] \arrow[d, "II\alpha"'] \arrow[dr, "I\alpha I"] & III \arrow[dr, "\alpha I"] \arrow[d, phantom, "\widetilde{\mathsf{A}}I"] & \\
III \arrow[dr, "I\alpha"'] \arrow[r, phantom, "I\widetilde{\mathsf{A}}"] & III \arrow[r, "\alpha I"] \arrow[d, "I\alpha"'] \arrow[dr, phantom, "\widetilde{\mathsf{A}}"] & II \arrow[d, "\alpha"] \\
& II \arrow[r, "\alpha"'] & I
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
IIII \arrow[r, "\alpha II"] \arrow[d, "II\alpha"'] \arrow[dr, phantom, "©"] & III \arrow[d, "I\alpha"] \arrow[dr, "\alpha I"] & \\
III \arrow[r, "\alpha I"'] \arrow[dr, "I\alpha"'] & II \arrow[r, phantom, "\widetilde{\mathsf{A}}"] \arrow[dr, "\alpha"] \arrow[d, phantom, "\widetilde{\mathsf{A}}"] & II \arrow[d, "\alpha"] \\
& II \arrow[r, "\alpha"'] & I
\end{tikzcd}
\tag{8}
$$

In other words, a unit object is automatically a semi-monoid. The 2-cell A is characterised uniquely in 3.7.

Theorem B. *A unit morphism $(u, \mathsf{U}) : (I, \alpha) \to (J, \beta)$ is automatically a semi-monoid map, when I and J are considered semi-monoids in virtue of Theorem A.*

Theorem C (Contractibility). *The 2-category of units in $\mathscr{C}$ is contractible, if non-empty.*

In other words, between any two units there exists a unit morphism, and between any two parallel unit morphisms there is a unique unit 2-morphism. Theorem C shows that units objects are unique up to homotopy, so in this sense 'being unital' is a property not a structure.

The proofs of these three theorems rely on the auxiliary structure of left and right constraints which we develop in the next section, and which also displays the relation with the classical notion of monoidal 2-category: in Section 7 we show that the cancellable-idempotent notion of unit is equivalent to the notion extracted from the definition of tricategory of Gordon, Power, and Street [2]. This is our Theorem E.

3 Left and right actions, and associativity of the unit (Theorem A)

Throughout this section we fix a unit object (I, α).

3.1 Lemma. *For each object X there exists pairs $(\lambda_X, \mathsf{L}_X)$ and (ρ_X, R_X),*

$$\begin{aligned} \lambda_X : IX \to X, \qquad & \mathsf{L}_X : I\lambda_X \Rightarrow \alpha X \\ \rho_X : XI \to X, \qquad & \mathsf{R}_X : X\alpha \Rightarrow \rho_X I \end{aligned}$$

where λ_X and ρ_X are equi-arrows, and L_X are R_X are invertible 2-cells. For every such family, there is a unique way to assemble the λ_X into a natural transformation (this involves defining 2-cells λ_f for every arrow f in $\mathscr{C}$) in such a way that L is a natural modification. Similarly for the ρ_X and R_X.

The λ_X is an action of I on each X, and the 2-cell L_X expresses an associativity constraint on this action. Using these structures we will construct the associator for α, and show it satisfies the pentagon equation. Once that is in place we will see that the actions λ and ρ are coherent too (satisfying the appropriate pentagon equations).

We shall treat the left action. The right action is of course equivalent to establish.

3.2. Construction of the left action. Since tensoring with I is a fully faithful 2-functor, the functor

$$\mathrm{Hom}(IX, X) \to \mathrm{Hom}(IIX, IX)$$

is an equivalence of categories. In the second category there is the canonical object αX. Hence there is a pseudo pre-image which we denote $\lambda_X : IX \to X$, together with an invertible 2-cell $\mathsf{L}_X : I\lambda_X \Rightarrow \alpha X$:

$$\begin{tikzcd} IIX \arrow[r, bend left, "\alpha X"] \arrow[r, bend right, "I\lambda_X"'] \arrow[r, phantom, "\mathsf{L}_X"] & IX \end{tikzcd}$$

Since α is an equi-arrow, also αX is equi, and since L_X is invertible, we conclude that also $I\lambda_X$ is an equi-arrow. Finally since the 2-functor 'tensoring with I' is fully faithful, it reflects equi-arrows, so already λ_X is an equi-arrow.

3.3. Naturality. A slight variation in the formulation of the construction gives directly a natural transformation λ and a modification L: Let $\mathbb{L} : \mathscr{C} \to \mathscr{C}$ denote the 2-functor 'tensoring with I on the left'. Since $\mathbb{L}$ is fully faithful, there is an equivalence of categories

$$\mathrm{Nat}(\mathbb{L}, \mathrm{Id}_{\mathscr{C}}) \to \mathrm{Nat}(\mathbb{L} \circ \mathbb{L}, \mathbb{L}).$$

Now in the second category we have the canonical natural transformation whose X-component is αX (and with trivial components on arrows). Hence there is a pseudo pre-image natural transformation $\lambda : \mathbb{L} \to \mathrm{id}_{\mathscr{C}}$, together with a modification L whose X-component is $\mathsf{L}_X : I\lambda_X \Rightarrow \alpha X$.

However, we wish to stress the fact that the construction is completely objectwise. This fact is of course due to the presence of the isomorphism L_X: something isomorphic to a natural transformation is again natural. More precisely, to provide the 2-cell data λ_f needed to make λ into a natural transformation, just pull back the 2-cell data from the natural transformation αX. In detail, we need invertible 2-cells

$$\begin{tikzcd} IX \arrow[r, "\lambda_X"] \arrow[d, "If"'] \arrow[dr, phantom, "\lambda_f"] & X \arrow[d, "f"] \\ IY \arrow[r, "\lambda_Y"'] & Y. \end{tikzcd}$$

To say that the L_X constitute a modification (from λ to the identity) is to have this compatibility for every arrow $f : X \to Y$:

$$\begin{tikzcd} IIX \arrow[r, bend left, "\alpha X"] \arrow[r, bend right, "I\lambda_X"'] \arrow[r, phantom, "\mathsf{L}_X"] \arrow[d, "IIf"'] \arrow[dr, phantom, "I\lambda_f"] & IX \arrow[d, "If"] \\ IIY \arrow[r, bend right, "I\lambda_Y"'] & IY \end{tikzcd} = \begin{tikzcd} IIX \arrow[r, bend left, "\alpha X"] \arrow[d, "IIf"'] \arrow[dr, phantom, "©"] & IX \arrow[d, "If"] \\ IIY \arrow[r, bend left, "\alpha Y"] \arrow[r, bend right, "I\lambda_Y"'] \arrow[r, phantom, "\mathsf{L}_Y"] & IY \end{tikzcd}$$

(Here the commutative cell is actually the 2-cell part of the natural transformation αX.) Now the point is that each 2-cell λ_f is uniquely defined by this compatibility: indeed, since the other three 2-cells in the diagram are invertible, there is a unique 2-cell that can fill the place of $I\lambda_f$, and since I is cancellable this 2-cell comes from a unique 2-cell λ_f. The required compatibilities of λ_f with composition, identities, and 2-cells now follows from its construction: λ_f is just the translation via L of the identity 2-cell αX.

3.4. Uniqueness of the left constraints. There may be many choices for λ_X, and even for a fixed λ_X, there may be many choices for L_X. However, between any two pairs $(\lambda_X, \mathsf{L}_X)$ and $(\lambda'_X, \mathsf{L}'_X)$ there is a unique invertible 2-cell $\mathsf{U}^{\text{left}}_X : \lambda_X \Rightarrow \lambda'_X$ such that this compatibility holds:

$$\begin{tikzcd} I\lambda_X \arrow[rr, Rightarrow, "I\mathsf{U}^{\text{left}}_X"] \arrow[dr, Rightarrow, "\mathsf{L}_X"'] & & I\lambda'_X \arrow[dl, Rightarrow, "\mathsf{L}'_X"] \\ & \alpha X & \end{tikzcd}$$

Indeed, this diagram defines uniquely an invertible 2-cell $I\lambda_X \Rightarrow I\lambda'_X$, and since I is cancellable, this 2-cell comes from a unique 2-cell $\lambda_X \Rightarrow \lambda'_X$ which we then call $\mathsf{U}^{\text{left}}_X$.

There is of course a completely analogous statement for right constraints.

3.5. CONSTRUCTION OF THE ASSOCIATOR. We define $\mathsf{A} : I\alpha \Rightarrow \alpha I$ as the unique 2-cell satisfying the equation

$$
\begin{tikzcd}
IIII \arrow[dr, "I\alpha I"] & & \\
& III \arrow[dr, bend left, "\alpha I"] \arrow[dr, bend right, "I\alpha"'] \arrow[dr, phantom, "\mathsf{A}"] & \\
& & II
\end{tikzcd}
\quad = \quad
\begin{tikzcd}[column sep=huge, row sep=huge]
IIII \arrow[r, bend left, "I\alpha I"] \arrow[r, bend right, "\rho II"'] \arrow[r, phantom, "\mathsf{R}_I^{-1} I"] \arrow[d, bend right, "I\alpha I"'] \arrow[d, bend left, "II\lambda"] \arrow[d, phantom, "I\mathsf{L}_I^{-1}"] \arrow[dr, phantom, "\copyright"] & III \arrow[d, bend right, "I\lambda"'] \arrow[d, bend left, "\alpha I"] \arrow[d, phantom, "\mathsf{L}_I"] \\
III \arrow[r, bend left, "\rho I"] \arrow[r, bend right, "I\alpha"'] \arrow[r, phantom, "\mathsf{R}_I"] & II
\end{tikzcd}
\tag{9}
$$

This definition is meaningful: since $I\alpha I$ is an equi-arrow, pre-composing with $I\alpha I$ is a 2-equivalence, hence gives a bijection on the level of 2-cells, so A is determined by the right-hand side of the equation. Note that A is invertible since all the 2-cells in the construction are.

The associator $\widetilde{\mathsf{A}}$ is defined as A-followed-by-α:

$$\widetilde{\mathsf{A}} := \mathsf{A} \underset{II}{\#} \alpha,$$

but it will be more convenient to work with A.

3.6 PROPOSITION. *The definition of* A *does not depend on the choices of left constraint* (λ, L) *and right constraint* (ρ, R).

Proof. Write down the right-hand side of (9) in terms of different left and right constraints. Express these cells in terms of the original L_I and R_I, using the comparison 2-cells $\mathsf{U}_I^{\text{left}}$ and $\mathsf{U}_I^{\text{right}}$ of 3.4. Finally observe that these comparison cells can be moved across the commutative square to cancel each other pairwise. $\square$

3.7. UNIQUENESS OF A. Equation (9) may not appear familiar, but it is equivalent to the following 'pentagon' equation (after post-whiskering with α):

$$
\begin{tikzcd}
IIII \arrow[r, "\rho II"] \arrow[d, "II\lambda"'] \arrow[dr, "I\alpha I"] & III \arrow[dr, "\alpha I"] \arrow[d, phantom, "(\mathsf{R}I)\#(\alpha I)"] & \\
III \arrow[dr, "I\alpha"'] \arrow[r, phantom, "(I\mathsf{L})\#(I\alpha)"] & III \arrow[r, "\alpha I"] \arrow[d, "I\alpha"'] \arrow[dr, phantom, "\mathsf{A}\#\alpha"] & II \arrow[d, "\alpha"] \\
 & II \arrow[r, "\alpha"'] & I
\end{tikzcd}
=
\begin{tikzcd}
IIII \arrow[r, "\rho II"] \arrow[d, "II\lambda"'] \arrow[dr, phantom, "©"] & III \arrow[d, "I\lambda"] \arrow[dr, "\alpha I"] & \\
III \arrow[r, "\rho I"'] \arrow[dr, "I\alpha"'] & II \arrow[r, phantom, "\mathsf{L}\#\alpha"] \arrow[dr, "\alpha"] \arrow[d, phantom, "\mathsf{R}\#\alpha"] & II \arrow[d, "\alpha"] \\
 & II \arrow[r, "\alpha"'] & I
\end{tikzcd}
\tag{10}
$$

From this pentagon equation we shall derive the pentagon equation for A, asserted in Theorem A. To this end we need comparison between α, λ_I, and ρ_I, which we now establish, in analogy with Axiom (1) of monoidal category: the left and right constraints coincide on the unit object, up to a canonical 2-cell:

3.8 LEMMA. *There are unique invertible* 2*-cells*

$$\rho_I \overset{\mathsf{E}}{\Rightarrow} \alpha \overset{\mathsf{D}}{\Rightarrow} \lambda_I,$$

such that

$$
\begin{tikzcd}
III \arrow[r, bend left=50, "\alpha I"] \arrow[r, "I\lambda" description] \arrow[r, bend right=50, "I\alpha"'] \arrow[r, phantom, shift left=3, "\mathsf{L}"] \arrow[r, phantom, shift right=3, "I\mathsf{D}"] & II
\end{tikzcd}
=
\begin{tikzcd}
III \arrow[r, bend left, "\alpha I"] \arrow[r, bend right, "I\alpha"'] \arrow[r, phantom, "\mathsf{A}"] & II
\end{tikzcd}
=
\begin{tikzcd}
III \arrow[r, bend left=50, "\alpha I"] \arrow[r, "\rho I" description] \arrow[r, bend right=50, "I\alpha"'] \arrow[r, phantom, shift left=3, "\mathsf{E}I"] \arrow[r, phantom, shift right=3, "\mathsf{R}"] & II
\end{tikzcd}
\tag{11}
$$

Proof. The left-hand equation defines uniquely a 2-cell $I\alpha \Rightarrow I\lambda_I$, and since I is cancellable, this cell comes from a unique 2-cell $\alpha \Rightarrow \lambda_I$ which we then call D. Same argument for E. □

THEOREM A (ASSOCIATIVITY). *Given a unit object* (I, α)*, there is a canonical invertible* 2*-cell*

$$
\begin{tikzcd}
III \arrow[r, "\alpha I"] \arrow[d, "I\alpha"'] \arrow[dr, phantom, "\widetilde{\mathsf{A}}"] & II \arrow[d, "\alpha"] \\
II \arrow[r, "\alpha"'] & I
\end{tikzcd}
$$

which satisfies the pentagon Equation (8).

Proof. On each side of the cube equation (10), paste the cell $\mathsf{E}II$ on the top, and the cell $II\mathsf{D}$ on the left. On the left-hand side of the equation we can use Equations (11) directly, while on the right-hand side we first need to move those cells across the commutative square before applying (11). The result is precisely the pentagon cube for $\widetilde{\mathsf{A}} = \mathsf{A}\#\alpha$. $\square$

3.9. Coherence of the actions. We have now established that $(I, \alpha, \widetilde{\mathsf{A}})$ is a semi-monoid, and may observe that the left and right constraints are coherent actions, i.e. that their 'associators' L and R satisfy the appropriate pentagon equations. For the left action this equation is:

$$
\begin{tikzcd}
IIIX \arrow[r, "\alpha IX"] \arrow[d, "II\lambda"'] \arrow[dr, "I\alpha X"] & IIX \arrow[dr, "\alpha X"] \arrow[d, phantom, "\widetilde{\mathsf{A}}X"] & \\
IIX \arrow[r, phantom, "(I\mathsf{L})\#(I\lambda)"] \arrow[dr, "I\lambda"'] & IIX \arrow[r, "\alpha X"] \arrow[d, "I\lambda"] \arrow[dr, phantom, "\mathsf{L}\#\lambda"] & IX \arrow[d, "\lambda"] \\
& IX \arrow[r, "\lambda"'] & X
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
IIIX \arrow[r, "\alpha IX"] \arrow[d, "II\lambda"'] \arrow[dr, phantom, "©"] & IIX \arrow[d, "I\lambda"] \arrow[dr, "\alpha X"] & \\
IIX \arrow[r, "\alpha X"'] \arrow[dr, "I\lambda"'] & IX \arrow[r, phantom, "\mathsf{L}\#\lambda"] \arrow[d, phantom, "\mathsf{L}\#\lambda"] \arrow[dr, "\lambda"] & IX \arrow[d, "\lambda"] \\
& IX \arrow[r, "\lambda"'] & X
\end{tikzcd}
$$

Establishing this (and the analogous equation for the right action) is a routine calculation which we omit since we will not actually need the result. We also note that the two actions are compatible—i.e. constitute a two-sided action. Precisely this means that there is a canonical invertible 2-cell

$$
\begin{tikzcd}
IXI \arrow[r, "\lambda_X I"] \arrow[d, "I\rho_X"'] \arrow[dr, phantom, "\mathsf{B}"] & XI \arrow[d, "\rho_X"] \\
IX \arrow[r, "\lambda_X"'] & X.
\end{tikzcd}
$$

This 2-cell satisfies two pentagon equations, one for $IIXI$ and one for $IXII$.

4 Units in the 2-category of arrows in $\mathscr{C}$, and Theorem B

In this section we prove Theorem B, which asserts that a morphism of units $(u, \mathsf{U}) : (I, \alpha) \to (J, \beta)$ is automatically a semi-monoid map (with respect to the canonical associators A and B of the two units). We have to establish the cube equation of 2.4, or in fact the reduced version (7). The strategy to establish Equation (7) is to interpret everything in the 2-category of arrows of $\mathscr{C}$. The key point is to prove that a morphism of units is itself a unit in the 2-category of arrows. Then we invoke Theorem A to get an associator for this unit, and a pentagon equation, whose short form (6) will be the sought equation.

4.1. The 2-category of arrows. The 2-*category of arrows* in $\mathscr{C}$, denoted $\mathscr{C}^{\mathbf{2}}$, is the 2-category described as follows. The objects of $\mathscr{C}^{\mathbf{2}}$ are the arrows of $\mathscr{C}$,

$$X_0 \xrightarrow{x} X_1 .$$

The arrows from (X_0, X_1, x) to (Y_0, Y_1, y) are triples (f_0, f_1, F) where $f_0 : X_0 \to Y_0$ and $f_1 : X_1 \to Y_1$ are arrows in $\mathscr{C}$ and F is a 2-cell

$$\begin{tikzcd}
X_0 \arrow[r, "f_0"] \arrow[d, "x"'] \arrow[dr, phantom, "F"] & Y_0 \arrow[d, "y"] \\
X_1 \arrow[r, "f_1"'] & Y_1.
\end{tikzcd}$$

If (g_0, g_1, G) is another arrow from (X_0, X_1, x) to (Y_0, Y_1, y), a 2-cell from (f_0, f_1, F) to (g_0, g_1, G) is given by a pair (m_0, m_1) where $m_0 : f_0 \Rightarrow g_0$ and $m_1 : f_1 \Rightarrow g_1$ are 2-cells in $\mathscr{C}$ compatible with F and G in the sense that this cylinder commutes:

$$\begin{tikzcd}
X_0 \arrow[r, bend left, "g_0"] \arrow[r, bend right, "f_0"'] \arrow[r, phantom, "m_0"] \arrow[d, "x"'] \arrow[dr, phantom, "F"] & Y_0 \arrow[d, "y"] \\
X_1 \arrow[r, bend right, "f_1"'] & Y_1
\end{tikzcd}
=
\begin{tikzcd}
X_0 \arrow[r, bend left, "g_0"] \arrow[d, "x"'] \arrow[dr, phantom, "G"] & Y_0 \arrow[d, "y"] \\
X_1 \arrow[r, bend left, "g_1"] \arrow[r, bend right, "f_1"'] \arrow[r, phantom, "m_1"] & Y_1
\end{tikzcd}$$

Composition of arrows in $\mathscr{C}^{\mathbf{2}}$ is just pasting of squares. Vertical composition of 2-cells is just vertical composition of the components (the compatibility is guaranteed by pasting of cylinders along squares), and horizontal composition of 2-cells is horizontal composition of the components (compatibility guaranteed by pasting along the straight sides of the cylinders). Note that $\mathscr{C}^{\mathbf{2}}$ inherits a tensor product from $\mathscr{C}$: this follows from functoriality of the tensor product on $\mathscr{C}$.

4.2 Lemma. *If I_0 and I_1 are cancellable objects in $\mathscr{C}$ and $i : I_0 \to I_1$ is an equi-arrow, then i is cancellable in $\mathscr{C}^{\mathbf{2}}$.*

Proof. We have to show that for given arrows $x : X_0 \to X_1$ and $y : Y_0 \to Y_1$, the functor

$$\mathrm{Hom}_{\mathscr{C}^{\mathbf{2}}}(x, y) \to \mathrm{Hom}_{\mathscr{C}^{\mathbf{2}}}(ix, iy)$$

defined by tensoring with i on the left is an equivalence of categories (the check for tensoring on the right is analogous).

Let us first show that this functor is essentially surjective. Let

$$\begin{tikzcd} I_0X_0 \arrow[r, "s_0"] \arrow[d, "ix"'] \arrow[dr, phantom, "S"] & I_0Y_0 \arrow[d, "iy"] \\ I_1X_1 \arrow[r, "s_1"'] & I_1Y_1 \end{tikzcd}$$

be an object in $\mathrm{Hom}_{\mathscr{C}^{\mathbf{2}}}(ix, iy)$. We need to find a square

$$\begin{tikzcd} X_0 \arrow[r, "k_0"] \arrow[d, "x"'] \arrow[dr, phantom, "K"] & Y_0 \arrow[d, "y"] \\ X_1 \arrow[r, "k_1"'] & Y_1 \end{tikzcd}$$

and an isomorphism (m_0, m_1) from (s_0, s_1, S) to (I_0k_0, I_1k_1, iK), i.e. a cylinder

$$\begin{tikzcd} I_0X_0 \arrow[r, bend left, "I_0k_0"] \arrow[r, bend right, "s_0"'] \arrow[r, phantom, "m_0"] \arrow[d, "ix"'] \arrow[dr, phantom, "S"] & I_0Y_0 \arrow[d, "iy"] \\ I_1X_1 \arrow[r, bend right, "s_1"'] & I_1Y_1 \end{tikzcd} \quad = \quad \begin{tikzcd} I_0X_0 \arrow[r, bend left, "I_0k_0"] \arrow[d, "ix"'] \arrow[dr, phantom, "iK"] & I_0Y_0 \arrow[d, "iy"] \\ I_1X_1 \arrow[r, bend left, "I_1k_1"] \arrow[r, bend right, "s_1"'] \arrow[r, phantom, "m_1"] & I_1Y_1 \end{tikzcd}$$

Since I_0 is a cancellable object, the arrow s_0 is isomorphic to I_0k_0 for some $k_0 : X_0 \to Y_0$. Let the connecting invertible 2-cell be denoted $m_0 : s_0 \Rightarrow I_0k_0$. Similarly we find k_1 and $m_1 : s_1 \Rightarrow I_1k_1$. Since m_0 and m_1 are invertible, there is a unique 2-cell

$$\begin{tikzcd} I_0X_0 \arrow[r, "I_0k_0"] \arrow[d, "ix"'] \arrow[dr, phantom, "T"] & I_0Y_0 \arrow[d, "iy"] \\ I_1X_1 \arrow[r, "I_1k_1"'] & I_1Y_1 \end{tikzcd}$$

that can take the place of iK in the cylinder equation; it remains to see that T is of the form iK for some K. But this follows since the map

$$\begin{array}{rcl} 2\mathrm{Cell}_{\mathscr{C}}(k_0\#y, x\#k_1) & \longrightarrow & 2\mathrm{Cell}_{\mathscr{C}}(i(k_0\#y), i(x\#k_1)) \\ K & \longmapsto & iK \end{array} \tag{12}$$

is a bijection. Indeed, the map factors as 'tensoring with I_0 on the left' followed by 'post-composing with iY_1'; the first is a bijection since I_0 is cancellable, the second is a bijection since i (and hence iY_1) is an equi-arrow).

Now for the fully faithfulness of $\mathrm{Hom}_{\mathscr{C}^{\mathbf{2}}}(x,y) \to \mathrm{Hom}_{\mathscr{C}^{\mathbf{2}}}(ix,iy)$. Fix two objects in the left-hand category, P and Q:

$$\begin{tikzcd} X_0 \arrow[r, "p_0"] \arrow[d, "x"'] \arrow[dr, phantom, "P"] & Y_0 \arrow[d, "y"] \\ X_1 \arrow[r, "p_1"'] & Y_1 \end{tikzcd} \qquad \begin{tikzcd} X_0 \arrow[r, "q_0"] \arrow[d, "x"'] \arrow[dr, phantom, "Q"] & Y_0 \arrow[d, "y"] \\ X_1 \arrow[r, "q_1"'] & Y_1. \end{tikzcd}$$

The arrows from P to Q are pairs (m_0, m_1) consisting of

$$m_0 : p_0 \Rightarrow q_0 \qquad m_1 : p_1 \Rightarrow q_1$$

cylinder-compatible with the 2-cells P and Q. The image of these two objects are

$$\begin{tikzcd} I_0X_0 \arrow[r, "I_0p_0"] \arrow[d, "ix"'] \arrow[dr, phantom, "iP"] & I_0Y_0 \arrow[d, "iy"] \\ I_1X_1 \arrow[r, "I_1p_1"'] & I_1Y_1 \end{tikzcd} \qquad \begin{tikzcd} I_0X_0 \arrow[r, "I_0q_0"] \arrow[d, "ix"'] \arrow[dr, phantom, "iQ"] & I_0Y_0 \arrow[d, "iy"] \\ I_1X_1 \arrow[r, "I_1q_1"'] & I_1Y_1. \end{tikzcd}$$

The possible 2-cells from iP to iQ are pairs (n_0, n_1) consisting of

$$n_0 : I_0p_0 \Rightarrow I_0q_0 \qquad n_1 : I_1p_1 \Rightarrow I_1q_1$$

cylinder-compatible with the 2-cells iP and iQ. Now since I_0 is cancellable, every 2-cell n_0 like this is uniquely of the form I_0n_0 for some n_0. Hence there is a bijection between the possible m_0 and the possible n_0. Similarly for m_1 and n_1. So there is a bijection between pairs (m_0, m_1) and pairs (n_0, n_1). Now by functoriality of tensoring with i, all images of compatible (m_0, m_1) are again compatible. It remains to rule out the possibility that some (n_0, n_1) pair could be compatible without (m_0, m_1) being so, but this follows again from the argument that 'tensoring with i on the left' is a bijection on hom sets, just like argued for (12). $\square$

4.3 Lemma. *An arrow in $\mathscr{C}^{\mathbf{2}}$,*

$$\begin{tikzcd} X_0 \arrow[r, "f_0"] \arrow[d, "x"'] \arrow[dr, phantom, "F"] & Y_0 \arrow[d, "y"] \\ X_1 \arrow[r, "f_1"'] & Y_1 \end{tikzcd}$$

is an equi-arrow in $\mathscr{C}^{\mathbf{2}}$ if the components f_0 and f_1 are equi-arrows in $\mathscr{C}$ and F is invertible.

Proof. We can construct an explicit quasi-inverse by choosing quasi-inverses to the components. ☐

4.4 Corollary. *If (I_0, α_0) and (I_1, α_1) are units in $\mathscr{C}$, and $(u, \mathsf{U}) : I_0 \to I_1$ is a unit map between them, then*

$$u : I_0 \to I_1$$

is a unit object in $\mathscr{C}^{\mathbf{2}}$ with structure map

$$\begin{tikzcd}
I_0 I_0 \arrow[r, "\alpha_0"] \arrow[d, "uu"'] \arrow[dr, phantom, "\mathsf{U}^{-1}"] & I_0 \arrow[d, "u"] \\
I_1 I_1 \arrow[r, "\alpha_1"'] & I_1 .
\end{tikzcd}$$

Proof. The object u is cancellable by Lemma 4.2, and the morphism $(\alpha_0, \alpha_1, \mathsf{U}^{-1})$ from uu to u is an equi-arrow by Lemma 4.3. ☐

Theorem B. *Let (I_0, α_0) and (I_1, α_1) be units, with canonical associators A_0 and A_1, respectively. If (u, U) is a unit map from I_0 to I_1 then it is automatically a semi-monoid map. That is,*

$$\begin{tikzcd}
I_0 I_0 I_0 \arrow[r, "uuu"] \arrow[d, bend right=50, "I_0\alpha_0"'] \arrow[d, bend left=50, "\alpha_0 I_0"] \arrow[d, phantom, "\mathsf{A}_0"] \arrow[dr, phantom, "\mathsf{U}u"] & I_1 I_1 I_1 \arrow[d, bend left=30, "\alpha_1 I_1"] \\
I_0 I_0 \arrow[r, "uu"'] & I_1 I_1
\end{tikzcd}
=
\begin{tikzcd}
I_0 I_0 I_0 \arrow[r, "uuu"] \arrow[d, bend right=30, "I_0\alpha_0"'] \arrow[dr, phantom, "u\mathsf{U}"] & I_1 I_1 I_1 \arrow[d, bend right=50, "I_1\alpha_1"'] \arrow[d, bend left=50, "\alpha_1 I_1"] \arrow[d, phantom, "\mathsf{A}_1"] \\
I_0 I_0 \arrow[r, "uu"'] & I_1 I_1
\end{tikzcd}$$

Proof. By the previous Corollary, (u, U^{-1}) is a unit object in $\mathscr{C}^{\mathbf{2}}$. Hence there is a canonical associator

$$\mathsf{B} : u\mathsf{U}^{-1} \Leftrightarrow \mathsf{U}^{-1}u.$$

By definition of 2-cells in $\mathscr{C}^{\mathbf{2}}$, this is a pair of 2-cells in $\mathscr{C}$

$$\mathsf{B}_0 : I_0\alpha_0 \Rightarrow \alpha_0 I_0 \qquad \mathsf{B}_1 : I_1\alpha_1 \Rightarrow \alpha_1 I_1,$$

fitting the cylinder equation

$$\begin{tikzcd}
I_0 I_0 I_0 \arrow[r, bend left, "\alpha_0 I_0"] \arrow[r, bend right, "I_0\alpha_0"'] \arrow[r, phantom, "\mathsf{B}_0"] \arrow[d, "uuu"'] \arrow[dr, phantom, "u\mathsf{U}^{-1}"] & I_0 I_0 \arrow[d, "uu"] \\
I_1 I_1 I_1 \arrow[r, bend right, "I_1\alpha_1"'] & I_1 I_1
\end{tikzcd}
=
\begin{tikzcd}
I_0 I_0 I_0 \arrow[r, bend left, "\alpha_0 I_0"] \arrow[d, "uuu"'] \arrow[dr, phantom, "\mathsf{U}^{-1}u"] & I_0 I_0 \arrow[d, "uu"] \\
I_1 I_1 I_1 \arrow[r, bend left, "\alpha_1 I_1"] \arrow[r, bend right, "I_1\alpha_1"'] \arrow[r, phantom, "\mathsf{B}_1"] & I_1 I_1
\end{tikzcd}$$

This is precisely the cylinder diagram we are looking for—provided we can show that $\mathsf{B}_0 = \mathsf{A}_0$ and $\mathsf{B}_1 = \mathsf{A}_1$. But this is a consequence of the characterising property of the associator of a unit: first note that as a unit object in $\mathscr{C}^{\mathbf{2}}$, u induces left and right constraints: for each object $x : X_0 \to X_1$ in $\mathscr{C}^{\mathbf{2}}$ there is a left action of the unit u, and this left action will induce a left action of (I_0, α_0) on X_0 and a left action of (I_1, α_1) on X_1 (the ends of the cylinders). Similarly there is a right action of u which induces right actions at the ends of the cylinder. Now the unique B that exists as associator for the unit object u compatible with the left and right constraints induces B_0 and B_1 at the ends of the cylinder, and these will of course be compatible with the induced left and right constraints. Hence, by uniqueness of associators compatible with left and right constraints, these induced associators B_0 and B_1 must coincide with A_0 and A_1. Note that this does not dependent on choice of left and right constraints, cf. Proposition 3.6. $\square$

5 Contractibility of the space of weak units (Theorem C)

The goal of this section is to prove Theorem C, which asserts that the 2-category of units in $\mathscr{C}$ is contractible if non-empty. First we describe the unit morphisms and unit 2-morphisms in terms of compatibility with left and right constraints. This will show that there are not too many 2-cells. Second we use the left and right constraints to connect any two units.

The following lemma shows that just as the single arrow α induces all the λ_X and ρ_X, the single 2-cell U of a unit map induces families $\mathsf{U}_X^{\text{left}}$ and $\mathsf{U}_X^{\text{right}}$ expressing compatibility with λ_X and ρ_X.

5.1 Lemma. *Let (I, α) and (J, β) be units, and let (u, U) be a morphism of pseudo-idempotents from (I, α) to (J, β). The following are equivalent.*

- *(i) u is an equi-arrow (i.e. u is a morphism of units).*
- *(ii) u is left cancellable, i.e. tensoring with u on the left is an equivalence of categories* $\operatorname{Hom}(X, Y) \to \operatorname{Hom}(IX, JY)$.
- *(ii') u is right cancellable, i.e. tensoring with u on the right is an equivalence of categories* $\operatorname{Hom}(X, Y) \to \operatorname{Hom}(XI, YJ)$.
- *(iii) For fixed left actions $(\lambda_X, \mathsf{L}_X)$ for the unit (I, α) and (ℓ_X, L'_X) for the unit (J, β), there is a unique invertible 2-cell $\mathsf{U}_X^{\text{left}}$, natural in X:*

$$\begin{tikzcd} IX \arrow[r, "uX"] \arrow[d, "\lambda_X"'] \arrow[dr, phantom, "\mathsf{U}_X^{\text{left}}"] & JX \arrow[d, "\ell_X"] \\ X \arrow[r, "X"'] & X \end{tikzcd}$$

such that this compatibility holds:

$$
\begin{tikzcd}
IIX \arrow[r, "uuX"] \arrow[d, bend right=60, "I\lambda_X"'] \arrow[d, bend left=60, "\alpha X"] \arrow[d, phantom, "\mathsf{L}_X"] \arrow[dr, phantom, "\mathsf{U}X"] & JJX \arrow[d, "\beta X"] \\
IX \arrow[r, "uX"'] & JX
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
IIX \arrow[r, "uuX"] \arrow[d, "I\lambda_X"'] \arrow[dr, phantom, "u\mathsf{U}^{\mathrm{left}}_X"] & JJX \arrow[d, bend right=60, "J\ell_X"'] \arrow[d, bend left=60, "\beta X"] \arrow[d, phantom, "\mathsf{L}'_X"] \\
IX \arrow[r, "uX"'] & JX
\end{tikzcd}
\tag{13}
$$

(iii') For fixed right actions (ρ_X, R_X) for the unit (I, α) and (r_X, R'_X) for the unit (J, β), there is a unique invertible 2-cell $\mathsf{U}^{\mathrm{right}}_X$, natural in X:

$$
\begin{tikzcd}
XI \arrow[r, "Xu"] \arrow[d, "\rho_X"'] \arrow[dr, phantom, "\mathsf{U}^{\mathrm{right}}_X"] & XJ \arrow[d, "r_X"] \\
X \arrow[r, "X"'] & X
\end{tikzcd}
$$

such that this compatibility holds:

$$
\begin{tikzcd}
XII \arrow[r, "Xuu"] \arrow[d, bend right=60, "X\alpha"'] \arrow[d, bend left=60, "\rho_X I"] \arrow[d, phantom, "\mathsf{R}_X"] \arrow[dr, phantom, "\mathsf{U}^{\mathrm{right}}_X u"] & XJJ \arrow[d, "r_X J"] \\
XI \arrow[r, "Xu"'] & XJ
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
XII \arrow[r, "Xuu"] \arrow[d, "X\alpha"'] \arrow[dr, phantom, "X\mathsf{U}"] & XJJ \arrow[d, bend right=60, "X\beta"'] \arrow[d, bend left=60, "r_X J"] \arrow[d, phantom, "\mathsf{R}'_X"] \\
XI \arrow[r, "Xu"'] & XJ
\end{tikzcd}
\tag{14}
$$

Proof. (i) implies (ii): 'tensoring with u' can be done in two steps: given an arrow $X \to Y$, first tensor with I to get $IX \to IY$, and then post-compose with uY to get $IX \to JY$. The first step is an equivalence because I is a unit, and the second step is an equivalence because u is an equi-arrow.
(ii) implies (iii): In Equation (13), the 2-cell labelled $u\mathsf{U}^{\mathrm{left}}_X$ is uniquely defined by the three other cells, and it is invertible since the three other cells are. Since tensoring with u on the left is an equivalence, this cell comes from a unique invertible cell $\mathsf{U}^{\mathrm{left}}_X$, justifying the label $u\mathsf{U}^{\mathrm{left}}_X$.
(iii) implies (i): The invertible 2-cell $\mathsf{U}^{\mathrm{left}}_X$ shows that uX is isomorphic to an equi-arrow, and hence is an equi-arrow itself. Now take X to be a right cancellable object (like for example I) and conclude that already u is an equi-arrow.
Finally, the equivalence (i)⇒(ii')⇒(iii')⇒(i) is completely analogous. □

Note that for (u, U) the identity morphism on (I, α), we recover Observation 3.4.

5.2 Lemma. *Let (I, α) and (J, β) be units; let (u, U) and (v, V) be morphisms of pseudo-idempotents from I to J; and consider a 2-cell $\mathsf{T} : u \Rightarrow v$. Then the following are equivalent.*

(i) T *is an invertible* 2*-morphism of pseudo-idempotents.*

(ii) T *is a left cancellable* 2*-morphism of pseudo-idempotents (i.e., induces a bijection on hom sets (of hom cats) by tensoring with* T *from the left).*

(ii') T *is a right cancellable* 2*-morphism of pseudo-idempotents (i.e., induces a bijection on hom sets (of hom cats) by tensoring with* T *from the right).*

(iii) For fixed left actions $(\lambda_X, \mathsf{L}_X)$ for (I, α) and (ℓ_X, L'_X) for (J, β), with induced canonical 2*-cells $\mathsf{U}^{\text{left}}_X$ and $\mathsf{V}^{\text{left}}_X$ as in 5.1, we have:*

$$\begin{tikzcd}
IX \arrow[r, bend left, "vX"] \arrow[r, bend right, "uX"'] \arrow[r, phantom, "\mathsf{T}X"] \arrow[d, "\lambda_X"'] \arrow[dr, phantom, "\mathsf{U}^{\text{left}}_X"] & JX \arrow[d, "\ell_X"] \\
X \arrow[r, bend right, "X"'] & X
\end{tikzcd}
=
\begin{tikzcd}
IX \arrow[r, bend left, "vX"] \arrow[d, "\lambda_X"'] \arrow[dr, phantom, "\mathsf{V}^{\text{left}}_X"] & JX \arrow[d, "\ell_X"] \\
X \arrow[r, bend left, "X"] \arrow[r, bend right, "X"'] \arrow[r, phantom, "©"] & X
\end{tikzcd}
\tag{15}$$

(iii') For fixed right actions (ρ_X, R_X) for (I, α) and (r_X, R'_X) for (J, β), with induced canonical 2*-cells $\mathsf{U}^{\text{right}}_X$ and $\mathsf{V}^{\text{right}}_X$ as in 5.1, we have:*

$$\begin{tikzcd}
XI \arrow[r, bend left, "Xv"] \arrow[r, bend right, "Xu"'] \arrow[r, phantom, "X\mathsf{T}"] \arrow[d, "\rho_X"'] \arrow[dr, phantom, "\mathsf{U}^{\text{right}}_X"] & XJ \arrow[d, "r_X"] \\
X \arrow[r, bend right, "X"'] & X
\end{tikzcd}
=
\begin{tikzcd}
XI \arrow[r, bend left, "Xv"] \arrow[d, "\rho_X"'] \arrow[dr, phantom, "\mathsf{V}^{\text{right}}_X"] & XJ \arrow[d, "r_X"] \\
X \arrow[r, bend left, "X"] \arrow[r, bend right, "X"'] \arrow[r, phantom, "©"] & X
\end{tikzcd}
\tag{16}$$

Proof. It is obvious that (i) implies (ii). Let us prove that (ii) implies (iii), so assume that tensoring with T on the left defines a bijection on the level of 2-cells. Start with the cylinder diagram for compatibility of tensor 2-cells (cf. 2.5). Tensor this diagram with X on the right to get

$$
\begin{tikzcd}
IIX \arrow[r, bend left, "vvX"] \arrow[r, bend right, "uuX"'] \arrow[r, phantom, "\mathsf{TT}X"] \arrow[d, "\alpha X"'] \arrow[dr, phantom, "\mathsf{U}X"] & JJX \arrow[d, "\beta X"] \\
IX \arrow[r, bend right, "uX"'] & JX
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
IIX \arrow[r, bend left, "vvX"] \arrow[d, "\alpha X"'] \arrow[dr, phantom, "\mathsf{V}X"] & JJX \arrow[d, "\beta X"] \\
IX \arrow[r, bend left, "vX"] \arrow[r, bend right, "uX"'] \arrow[r, phantom, "\mathsf{T}X"] & JX
\end{tikzcd}
$$

On each side of this equation, paste an L_X along αX, apply Equation (13) on each side, and cancel the L'_X that appear on the other side of the square. The resulting diagram

$$
\begin{tikzcd}
IIX \arrow[r, bend left, "vvX"] \arrow[r, bend right, "uuX"'] \arrow[r, phantom, "\mathsf{TT}X"] \arrow[d, "I\lambda_X"'] \arrow[dr, phantom, "u\mathsf{U}_X^{\text{left}}"] & JJX \arrow[d, "J\ell_X"] \\
IX \arrow[r, bend right, "uX"'] & JX
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
IIX \arrow[r, bend left, "vvX"] \arrow[d, "I\lambda_X"'] \arrow[dr, phantom, "v\mathsf{V}_X^{\text{left}}"] & JJX \arrow[d, "J\ell_X"] \\
IX \arrow[r, bend left, "vX"] \arrow[r, bend right, "uX"'] \arrow[r, phantom, "\mathsf{T}X"] & JX
\end{tikzcd}
$$

is the tensor product of T with the promised equation (15). Since T is cancellable, we can cancel it away to finish.

(iii) implies (i): the arguments in (ii)⇒(iii) can be reversed: start with (15), tensor with T on the left, and apply (13) to arrive at the axiom for being a 2-morphism of pseudo-idempotents. Since both $\mathsf{U}_X^{\text{left}}$ and $\mathsf{V}_X^{\text{left}}$ are invertible, so is $\mathsf{T}X$. Now take X to be a right cancellable object, and cancel it away to conclude that already T is invertible.

Finally, the equivalence (i)⇒(ii')⇒(iii')⇒(i) is completely analogous. □

5.3 Corollary. *Given two parallel morphisms of units, there is a unique unit 2-morphism between them.*

Proof. Choose left actions for (I, α) and (J, β) as in Lemma 5.2 (iii), and take X to be a right cancellable object. For given morphisms of units u and v as in Lemma 5.2, Equation (15) defines the 2-cell T uniquely, since λ_X is an equi-arrow and X is right cancellable. □

Next we aim at proving that there is a unit morphism between any two units. The strategy is to use the left and right constraints to produce a unit morphism

$$I \longrightarrow IJ \longrightarrow J.$$

As a first step towards this goal we have:

5.4 Lemma. *Let I and J be units, and pick a left constraint λ for I and a right constraint r for J. Put*

$$\gamma := r_I \lambda_J : IJIJ \to IJ$$

Then (IJ, γ) is a unit.

Proof. Since I and J are cancellable, clearly IJ is cancellable too. Since λ_J and r_I are equi-arrows, γ is too. □

5.5 Lemma. *There is an invertible 2-cell*

$$\begin{tikzcd}
IJIJ \arrow[r, "\lambda_J \lambda_J"] \arrow[d, "\gamma"'] \arrow[dr, phantom, "\mathsf{Z}"] & JJ \arrow[d, "\beta"] \\
IJ \arrow[r, "\lambda_J"'] & J.
\end{tikzcd}$$

Hence (λ_J, Z) is a unit map. (And there is another 2-cell making r_I a unit map.)

Proof. The 2-cell Z is defined like this:

$$\begin{tikzcd}
IJIJ \arrow[d, "IJ\lambda_J"'] \arrow[dr, bend left, "\lambda_J \lambda_J"] & \\
IJJ \arrow[r, bend left, "\lambda_J J"] \arrow[r, bend right, "\lambda_{JJ}"'] \arrow[r, phantom, "\mathsf{K}^\lambda"] \arrow[d, bend right, "r_I J"'] \arrow[d, bend left, "I\beta"] \arrow[d, phantom, "\mathsf{R}^{-1}"] \arrow[dr, phantom, "\lambda_\beta"] & JJ \arrow[d, "\beta"] \\
IJ \arrow[r, "\lambda_J"'] & J
\end{tikzcd}$$

where the 2-cell K^λ is constructed in Lemma 7.2. □

5.6 Corollary. *Given two units, there exists a unit morphism between them.*

Proof. Continuing the notation from above, by Lemma 5.4, (IJ, γ) is a unit, and by Lemma 5.5, $\lambda : IJ \to J$ is a morphism of units. Similarly, $r : IJ \to I$ is a unit morphism, and by Lemma 2.6 any chosen pseudo-inverse $r^{-1} : I \to IJ$ is again a unit morphism. Finally we take

$$I \xrightarrow{r^{-1}} IJ \xrightarrow{\lambda} J.$$

□

Theorem C (Contractibility). *The 2-category of units in $\mathscr{C}$ is contractible, if non-empty. In other words, between any two units there exists a unit morphism, and between any two parallel unit morphisms there is a unique unit 2-morphism.*

Proof. By Lemma 5.6 there is a unit morphism between any two units (an equi-arrow by definition), and by Corollary 5.3 there is a unique unit 2-morphism between any two parallel unit morphisms. □

6 Classical units

In this section we review the classical theory of units in a monoidal 2-category, as extracted from the definition of tricategory of Gordon, Power, and Street [2]. In the next section we compare this notion with the cancellable-idempotent approach of this work. The equivalence is stated explicitly in Theorem E.

6.1. Tricategories. The notion of tricategory introduced by Gordon, Power, and Street [2] is roughly a weak category structure enriched over bicategories: this means that the structure maps (composition and unit) are weak 2-functors satisfying weak versions of associativity and unit constraints. For the associativity, the pentagon equation is replaced by a specified pentagon 3-cell (TD7), required to satisfy an equation corresponding to the 3-dimensional associahedron. This equation (TA1) is called the nonabelian 4-cocycle condition. For the unit structure, three families of 3-cells are specified (TD8): one corresponding to the Kelly axiom, one left variant, and one right variant (those two being the higher-dimensional analogues of Axioms (2) and (3) of monoidal category). Two axioms are imposed on these three families of 3-cells: one (TA2) relating the left family with the middle family, and one (TA3) relating the right family with the middle family. These are called left and right normalisation. (These two axioms are the higher-dimensional analogues of the first argument in Kelly's lemma 1.6.) It is pointed out in [2] that the middle family together with the axioms (TA2) and (TA3) completely determine the left and right families if they exist.

6.2. Monoidal 2-categories. By specialising the definition of tricategory to the one-object case, and requiring everything strict except the units, we arrive at the following notion of monoidal 2-category: a *monoidal 2-category* is a semi-monoidal 2-category (cf. 2.2) equipped with an object I, two natural transformations λ and ρ with equi-arrow components

$$\lambda_X : IX \to X$$

$$\rho_X : XI \to X$$

and (invertible) 2-cell data

$$
\begin{tikzcd}
IX \arrow[r, "\lambda_X"] \arrow[d, "If"'] \arrow[dr, phantom, "\lambda_f"] & X \arrow[d, "f"] \\
IY \arrow[r, "\lambda_Y"'] & Y
\end{tikzcd}
\qquad
\begin{tikzcd}
XI \arrow[r, "\rho_X"] \arrow[d, "fI"'] \arrow[dr, phantom, "\rho_f"] & X \arrow[d, "f"] \\
YI \arrow[r, "\rho_Y"'] & Y,
\end{tikzcd}
$$

together with three natural modifications K, K^λ, and K^ρ, with invertible components

$$
\begin{aligned}
\mathsf{K}\ &: X\lambda_Y \Rightarrow \rho_X Y \\
\mathsf{K}^\lambda &: \lambda_{XY} \Rightarrow \lambda_X Y \\
\mathsf{K}^\rho &: X\rho_Y \Rightarrow \rho_{XY}.
\end{aligned}
$$

We call K the *Kelly cell*.
These three families are subject to the following two equations:

$$
\begin{tikzcd}
X\lambda_{YZ} \arrow[rr, Rightarrow, "X\mathsf{K}^\lambda_{Y,Z}"] \arrow[dr, Rightarrow, "\mathsf{K}_{X,YZ}"'] & & X\lambda_Y Z \arrow[dl, Rightarrow, "\mathsf{K}_{X,Y}Z"] \\
& \rho_X YZ &
\end{tikzcd}
\tag{17}
$$

$$
\begin{tikzcd}
X\rho_Y Z \arrow[rr, Rightarrow, "\mathsf{K}^\rho_{X,Y}Z"] & & \rho_{XY} Z \\
& XY\lambda_Z \arrow[ul, Rightarrow, "X\mathsf{K}_{Y,Z}"] \arrow[ur, Rightarrow, "\mathsf{K}_{XY,Z}"'] &
\end{tikzcd}
\tag{18}
$$

6.3 Remark. We have made one change compared to [2], namely the direction of the arrow ρ_X: from the viewpoint of α it seems more practical to work with $\rho_X : XI \to X$ rather than with the convention of $\rho_X : X \to XI$ chosen in [2]. Since in any case it is an equi-arrow, the difference is not essential. (Gurski in his thesis [3] has studied a version of tricategory where all the equi-arrows in the definition are equipped with specified pseudo-inverses. This has the advantage that the definition becomes completely algebraic, in a technical sense.)

6.4 Lemma. *The object I is cancellable (independently of the existence of* K*,* K^λ*, and* K^ρ*.)*

Proof. We need to establish that 'tensoring with I on the left',

$$\mathbb{L} : \mathrm{Hom}(X,Y) \to \mathrm{Hom}(IX,IY),$$

is an equivalence of categories. But this follows since the diagram

$$\begin{tikzcd}
\mathrm{Hom}(X,Y) \arrow[r, "\mathbb{L}"] \arrow[d, "\mathrm{Id}"'] & \mathrm{Hom}(IX,IY) \arrow[d, "_ \,\#\, \lambda_Y"] \\
\mathrm{Hom}(X,Y) \arrow[r, "\lambda_X \,\#\, _"'] & \mathrm{Hom}(IX,Y)
\end{tikzcd}$$

is commutative up to isomorphism: the component at $f : X \to Y$ of this isomorphism is just the naturality square λ_f. Since the functors $\lambda_X \# \_$ and $\_ \# \lambda_Y$ are equivalences, it follows from this isomorphism that $\mathbb{L}$ is too. $\square$

6.5. Coherence of the Kelly cell. As remarked in [2], if the K^λ and K^ρ exist, they are determined uniquely from K and the two axioms. Indeed, the two equations

$$\begin{tikzcd}
I\lambda_{YZ} \arrow[rr, Rightarrow, "I\mathsf{K}^\lambda_{Y,Z}"] \arrow[dr, Rightarrow, "\mathsf{K}_{I,YZ}"'] & & I\lambda_Y Z \arrow[dl, Rightarrow, "\mathsf{K}_{I,Y}Z"] \\
& \rho_I YZ &
\end{tikzcd}
\qquad
\begin{tikzcd}
X\rho_Y I \arrow[rr, Rightarrow, "\mathsf{K}^\rho_{X,Y}I"] & & \rho_{XY} I \\
& XY\lambda_I \arrow[ul, Rightarrow, "X\mathsf{K}_{Y,I}"] \arrow[ur, Rightarrow, "\mathsf{K}_{XY,I}"'] &
\end{tikzcd} \tag{19}$$

which are just special cases of (17) and (18) uniquely determine K^λ and K^ρ, by cancellability of I. But these two special cases of the axioms do not imply the general case.
We shall take the Kelly cell K as the main structure, and say that K is *coherent on the left* (resp. *on the right*) if Axiom (17) (resp. (18)) holds for the induced cell K^λ (resp. K^ρ). We just say *coherent* if both hold. We shall see (7.8) that in fact coherence on the left implies coherence on the right and vice versa.

6.6. Naturality. The Kelly cell is a modification. For future reference we spell out the naturality condition satisfied: given arrows $f : X \to X'$ and $g : Y \to Y'$, we have

$$\begin{tikzcd}
XIY \arrow[r, bend left, "\rho_X Y"] \arrow[r, bend right, "X\lambda_Y"'] \arrow[r, phantom, "\mathsf{K}_{X,Y}"] \arrow[d, "fIg"'] \arrow[dr, phantom, "f\lambda_g"] & XY \arrow[d, "fg"] \\
X'IY' \arrow[r, bend right, "X'\lambda_{Y'}"'] & X'Y'
\end{tikzcd}
=
\begin{tikzcd}
XIY \arrow[r, bend left, "\rho_X Y"] \arrow[d, "fIg"'] \arrow[dr, phantom, "\rho_f g"] & XY \arrow[d, "fg"] \\
X'IY' \arrow[r, bend left, "\rho_{X'}Y'"] \arrow[r, bend right, "X'\lambda_{Y'}"'] \arrow[r, phantom, "\mathsf{K}_{X',Y'}"] & X'Y'
\end{tikzcd}$$

6.7 Remark. Particularly useful is naturality of λ with respect to λ_X and naturality of ρ with respect to ρ_X. In these cases, since λ_X and ρ_X are equi-arrows, we can cancel them and find the following invertible 2-cells:

$$\mathsf{N}^\lambda : I\lambda_X \Rightarrow \lambda_{IX}$$
$$\mathsf{N}^\rho : \rho_{XI} \Rightarrow X\rho_I,$$

in analogy with Observation (5) of monoidal categories.

The following lemma holds for K independently of Axioms (17) and (18):

6.8 Lemma. *The Kelly cell* K *satisfies the equation*

$$\begin{tikzcd} XIIY \arrow[rr, bend left=50, "\rho_X IY"] \arrow[rr, "X\lambda_{IY}"] \arrow[rr, bend right=50, "XI\lambda_Y"'] \arrow[rr, phantom, bend left=25, "\mathsf{K}_{X,IY}"] \arrow[rr, phantom, bend right=25, "X\mathsf{N}^\lambda"] & & XIY \end{tikzcd} \quad = \quad \begin{tikzcd} XIIY \arrow[rr, bend left=50, "\rho_X IY"] \arrow[rr, "\rho_{XI}Y"] \arrow[rr, bend right=50, "XI\lambda_Y"'] \arrow[rr, phantom, bend left=25, "\mathsf{N}^\rho Y"] \arrow[rr, phantom, bend right=25, "\mathsf{K}_{XI,Y}"] & & XIY \end{tikzcd}$$

Proof. It is enough to establish this equation after post-whiskering with $X\lambda_Y$. The rest is a routine calculation, using on one side the definition of the cell N^λ, then naturality of K with respect to $f = X$ and $g = \lambda_Y$. On the other side, use the definition of N^ρ and then naturality of K with respect to $f = \rho_X$ and $g = Y$. In the end, two K-cells cancel. □

Combining the 2-cells described so far we get

$$\rho_I I \overset{\mathsf{K}^{-1}}{\Rightarrow} I\lambda_I \overset{\mathsf{N}^\lambda}{\Rightarrow} \lambda_{II} \overset{\mathsf{K}^\lambda}{\Rightarrow} \lambda_I I$$

and hence, by cancelling I on the right, an invertible 2-cell

$$\mathsf{P} : \rho_I \Rightarrow \lambda_I.$$

Now we could also define $\mathsf{Q} : \rho_I \Rightarrow \lambda_I$ in terms of

$$I\rho_I \overset{\mathsf{K}^\rho}{\Rightarrow} \rho_{II} \overset{\mathsf{N}^\rho}{\Rightarrow} \rho_I I \overset{\mathsf{K}^{-1}}{\Rightarrow} I\lambda_I.$$

Finally, in analogy with Axiom (1) for monoidal categories:

6.9 Lemma. *We have* $\mathsf{P} = \mathsf{Q}$. *(This is true independently of Axioms* (17) *and* (18)*.)*

Proof. Since I is cancellable, it is enough to show $I\mathsf{P}I = I\mathsf{Q}I$. To establish this equation, use the constructions of P and Q, then substitute the characterising Equations (19) for the auxiliary cells K^λ and K^ρ, and finally use Lemma 6.8. □

6.10. The 2-category of GPS units. For short we shall say *GPS unit* for the notion of unit just introduced. In summary, a GPS unit is a quadruple $(I, \lambda, \rho, \mathsf{K})$ where I is an object, λ_X and ρ_X are natural transformations with equi-arrow components, and $\mathsf{K} : X\lambda_Y \Rightarrow \rho_X Y$ is a coherent Kelly cell (natural in X and Y, of course).
A *morphism of GPS units* from $(I, \lambda, \rho, \mathsf{K})$ to (J, ℓ, r, H) is an arrow $u : I \to J$ equipped with natural families of invertible 2-cells

$$
\begin{tikzcd}
IX \arrow[r, "uX"] \arrow[d, "\lambda_X"'] \arrow[dr, phantom, "\mathsf{U}^{\text{left}}_X"] & JX \arrow[d, "\ell_X"] \\
X \arrow[r, "X"'] & X
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
XI \arrow[r, "Xu"] \arrow[d, "\rho_X"'] \arrow[dr, phantom, "\mathsf{U}^{\text{right}}_X"] & XJ \arrow[d, "r_X"] \\
X \arrow[r, "X"'] & X
\end{tikzcd}
$$

satisfying the equation

$$
\begin{tikzcd}
XIY \arrow[rr, "XuY"] \arrow[d, bend right=60, "X\lambda_Y"'] \arrow[d, bend left=60, "\rho_X Y"] \arrow[d, phantom, "\mathsf{K}"] \arrow[drr, phantom, "\mathsf{U}^{\text{right}}_X Y"] & & XJY \arrow[d, bend left=40, "r_X Y"] \\
XY \arrow[rr, "XY"'] & & XY
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
XIY \arrow[rr, "XuY"] \arrow[d, bend right=40, "X\lambda_Y"'] \arrow[drr, phantom, "X\mathsf{U}^{\text{left}}_Y"] & & XJY \arrow[d, bend right=60, "X\ell_Y"'] \arrow[d, bend left=60, "r_X Y"] \arrow[d, phantom, "\mathsf{H}"] \\
XY \arrow[rr, "XY"'] & & XY
\end{tikzcd}
\tag{20}
$$

Finally, a 2-*morphism of GPS unit maps* is a 2-cell $\mathsf{T} : u \Rightarrow v$ satisfying the compatibility conditions (15) and (16) of Lemma 5.2.

6.11. Remarks. Note first that u is automatically an equi-arrow. Observe also that U^{left} and $\mathsf{U}^{\text{right}}$ completely determine each other by Equation (20), as is easily seen by taking on the one hand X to be a left cancellable object and on the other hand Y to be a right cancellable object. Finally note that there are two further equations, expressing compatibility with K^λ and K^ρ, but they can be deduced from Equation (20), independently of the coherence Axioms (17) and (18). Here is the one for K^λ for future reference:

$$
\begin{tikzcd}
IXY \arrow[rr, "uXY"] \arrow[d, bend right=60, "\lambda_{XY}"'] \arrow[d, bend left=60, "\lambda_X Y"] \arrow[d, phantom, "\mathsf{K}^\lambda"] \arrow[drr, phantom, "\mathsf{U}^{\text{left}}_X Y"] & & JXY \arrow[d, bend left=40, "\ell_X Y"] \\
XY \arrow[rr, "XY"'] & & XY
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
IXY \arrow[rr, "uXY"] \arrow[d, bend right=40, "\lambda_{XY}"'] \arrow[drr, phantom, "\mathsf{U}^{\text{left}}_{XY}"] & & JXY \arrow[d, bend right=60, "\ell_{XY}"'] \arrow[d, bend left=60, "\ell_X Y"] \arrow[d, phantom, "\mathsf{H}^\ell"] \\
XY \arrow[rr, "XY"'] & & XY
\end{tikzcd}
\tag{21}
$$

7 Comparison with classical theory (Theorem E)

In this section we prove the equivalence between the two notions of unit.

7.1. From cancellable-idempotent units to GPS units. We fix a unit object (I, α). We also assume chosen a left constraint $\lambda_X : IX \to X$ with $\mathsf{L}_X : I\lambda_X \Rightarrow \alpha X$, and a right constraint $\rho_X : XI \to X$ with $\mathsf{R}_X : X\alpha \Rightarrow \rho_X I$. First of all, in analogy with Axioms (2) and (3) of monoidal categories we have:

7.2 Lemma. *In the situation of 7.1, there are unique natural invertible 2-cells*

$$\mathsf{K}^\lambda : \lambda_{XY} \Rightarrow \lambda_X Y$$
$$\mathsf{K}^\rho : X\rho_Y \Rightarrow \rho_{XY}$$

satisfying

$$\begin{tikzcd} IIXY \arrow[r, bend left, "I\lambda_X Y"] \arrow[r, bend right, "I\lambda_{XY}"'] \arrow[r, phantom, "I\mathsf{K}^\lambda"] & IXY \end{tikzcd} \quad = \quad \begin{tikzcd} IIXY \arrow[r, bend left=50, "I\lambda_X Y"] \arrow[r, "\alpha XY" description] \arrow[r, bend right=50, "I\lambda_{XY}"'] \arrow[r, phantom, shift left=3, "\mathsf{L}^{-1}Y"] \arrow[r, phantom, shift right=3, "\mathsf{L}"] & IXY \end{tikzcd} \tag{22}$$

$$\begin{tikzcd} XYII \arrow[r, bend left, "\rho_{XY} I"] \arrow[r, bend right, "X\rho_Y I"'] \arrow[r, phantom, "\mathsf{K}^\rho I"] & XYI \end{tikzcd} \quad = \quad \begin{tikzcd} XYII \arrow[r, bend left=50, "\rho_{XY} I"] \arrow[r, "XY\alpha" description] \arrow[r, bend right=50, "X\rho_Y I"'] \arrow[r, phantom, shift left=3, "\mathsf{R}"] \arrow[r, phantom, shift right=3, "X\mathsf{R}^{-1}"] & XYI \end{tikzcd} \tag{23}$$

Proof. The conditions precisely define the 2-cells, since I is cancellable. $\square$

7.3 Lemma. *In the situation of 7.1, there is a canonical family of invertible 2-cells (the Kelly cell)*

$$\mathsf{K} : X\lambda_Y \Rightarrow \rho_X Y,$$

natural in X and Y.

Proof. This is analogous to the construction of the associator: K is defined as the unique 2-cell $\mathsf{K} : X\lambda_Y \Rightarrow \rho_X Y$ satisfying the equation

$$\begin{tikzcd} XIIY \arrow[r, bend left, "X\alpha Y"] \arrow[r, bend right, "XI\lambda_Y"'] \arrow[r, phantom, "X\mathsf{L}"] \arrow[d, bend right, "X\alpha Y"'] \arrow[d, bend left, "\rho_X IY"] \arrow[d, phantom, "\mathsf{R}Y"] \arrow[dr, phantom, "©"] & XIY \arrow[d, "\rho_X Y"] \\ XIY \arrow[r, "X\lambda_Y"'] & XY \end{tikzcd} \quad = \quad \begin{tikzcd} XIIY \arrow[d, "X\alpha Y"] \\ XIY \arrow[d, bend left, "\rho_X Y"] \arrow[d, bend right, "X\lambda_Y"'] \arrow[d, phantom, "\mathsf{K}"] \\ XY \end{tikzcd} \tag{24}$$

This makes sense since $X\alpha Y$ is an equi-arrow, so we can cancel it away. Clearly K is invertible since L and R are. $\square$

We constructed K^λ and K^ρ directly from L, and R. Meanwhile we also constructed K, and we know from classical theory (6.5) that this cell determines the two others. The following proposition shows that all these constructions match up, and in particular that the constructed Kelly cell is coherent on both sides:

7.4 Proposition. *In the situation of 7.1, the families of* 2*-cells* K, K^λ *and* K^ρ *(constructed in 7.2 and 7.3) satisfy the GPS unit axioms* (17) *and* (18)*:*

$$\begin{tikzcd}
X\lambda_{YZ} \arrow[rr, Rightarrow, "X\mathsf{K}^\lambda_{Y,Z}"] \arrow[dr, Rightarrow, "\mathsf{K}_{X,YZ}"'] & & X\lambda_Y Z \arrow[dl, Rightarrow, "\mathsf{K}_{X,Y}Z"] \\
& \rho_X YZ &
\end{tikzcd}
\qquad
\begin{tikzcd}
X\rho_Y Z \arrow[rr, Rightarrow, "\mathsf{K}^\rho_{X,Y}Z"] & & \rho_{XY}Z \\
& XY\lambda_Z \arrow[ul, Rightarrow, "X\mathsf{K}_{Y,Z}"] \arrow[ur, Rightarrow, "\mathsf{K}_{XY,Z}"'] &
\end{tikzcd}$$

Proof. We treat the left constraint (the right constraint being completely analogous). We need to establish

$$\begin{tikzcd}
XIYZ \arrow[r, bend left=50, "\rho_X YZ"] \arrow[r, "X\lambda_Y Z" description] \arrow[r, bend right=50, "X\lambda_{YZ}"'] \arrow[r, phantom, shift left=3ex, "\mathsf{K}_{X,Y}Z"] \arrow[r, phantom, shift right=3ex, "X\mathsf{K}^\lambda_{Y,Z}"] & XYZ
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
XIYZ \arrow[r, bend left, "\rho_X YZ"] \arrow[r, bend right, "X\lambda_{YZ}"'] \arrow[r, phantom, "\mathsf{K}_{X,YZ}"] & XYZ
\end{tikzcd}$$

and it is enough to establish this equation pre-whiskered with $X\alpha YZ$. In the diagram resulting from the left-hand side:

$$\begin{tikzcd}
XIIYZ \arrow[r, "X\alpha YZ"] & XIYZ \arrow[r, bend left=50, "\rho_X YZ"] \arrow[r, "X\lambda_Y Z" description] \arrow[r, bend right=50, "X\lambda_{YZ}"'] \arrow[r, phantom, shift left=3ex, "\mathsf{K}_{X,Y}Z"] \arrow[r, phantom, shift right=3ex, "X\mathsf{K}^\lambda_{Y,Z}"] & XYZ
\end{tikzcd}$$

we can replace $(X\alpha YZ)\#(\mathsf{K}_{X,Y}Z)$ by the expression that defined $\mathsf{K}_{X,Y}Z$

(cf. (24)), yielding altogether

$$\begin{tikzcd}
& XIYZ \arrow[dr, "\rho_X YZ"] & \\
XIIYZ \arrow[ur, bend left, "X\alpha YZ"] \arrow[ur, bend right, "XI\lambda_Y Z"'] \arrow[ur, phantom, "X\mathsf{L}_Y Z"] \arrow[rr, phantom, "\copyright"] \arrow[dr, bend left, "\rho_X IYZ"] \arrow[dr, bend right, "X\alpha YZ"'] \arrow[dr, phantom, "\mathsf{R}_X YZ"] & & XYZ \\
& XIYZ \arrow[ur, bend left, "X\lambda_Y Z"] \arrow[ur, bend right, "X\lambda_{YZ}"'] \arrow[ur, phantom, "X\mathsf{K}^{\lambda}_{Y,Z}"] &
\end{tikzcd}$$

Here we can move the cell $X\mathsf{K}^{\lambda}_{Y,Z}$ across the square, where it becomes $XI\mathsf{K}^{\lambda}_{Y,Z}$ and combines with $X\mathsf{L}_Y Z$ to give altogether $X\mathsf{L}_{YZ}$ (cf. (22)). The resulting diagram

$$\begin{tikzcd}
& XIYZ \arrow[dr, "\rho_X YZ"] & \\
XIIYZ \arrow[ur, bend left, "X\alpha YZ"] \arrow[ur, bend right, "XI\lambda_{YZ}"'] \arrow[ur, phantom, "X\mathsf{L}_{YZ}"] \arrow[rr, phantom, "\copyright"] \arrow[dr, bend left, "\rho_X IYZ"] \arrow[dr, bend right, "X\alpha YZ"'] \arrow[dr, phantom, "\mathsf{R}_X YZ"] & & XYZ \\
& XIYZ \arrow[ur, bend right, "X\lambda_{YZ}"'] &
\end{tikzcd}$$

is nothing but

$$\begin{tikzcd}
XIIYZ \arrow[r, "X\alpha YZ"] & XIYZ \arrow[r, bend left, "\rho_X YZ"] \arrow[r, bend right, "X\lambda_{YZ}"'] \arrow[r, phantom, "\mathsf{K}_{X,YZ}"] & XYZ
\end{tikzcd}$$

(by Equation (24) again) which is what we wanted to establish. □

Hereby we have concluded the construction of a GPS unit from (I, α). We will also need a result for morphisms:

7.5 Proposition. *Let $(u, \mathsf{U}) : (I, \alpha) \to (J, \beta)$ be a morphism of units in the sense of 2.9, and consider the two canonical* 2*-cells* U^{left} *and* $\mathsf{U}^{\text{right}}$ *constructed* *in Lemma 5.1. Then Equation* (20) *holds:*

$$
\begin{tikzcd}
XIY \arrow[r, "XuY"] \arrow[d, bend right=60, "X\lambda_Y"'] \arrow[d, bend left=60, "\rho_X Y"] \arrow[d, phantom, "\mathsf{K}"] \arrow[dr, phantom, "\mathsf{U}^{\text{right}}_X Y"] & XJY \arrow[d, bend left=40, "r_X Y"] \\
XY \arrow[r, "XY"'] & XY
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
XIY \arrow[r, "XuY"] \arrow[d, bend right=40, "X\lambda_Y"'] \arrow[dr, phantom, "X\mathsf{U}^{\text{left}}_Y"] & XJY \arrow[d, bend right=60, "X\ell_Y"'] \arrow[d, bend left=60, "r_X Y"] \arrow[d, phantom, "\mathsf{H}"] \\
XY \arrow[r, "XY"'] & XY
\end{tikzcd}
$$

(Hence $(u, \mathsf{U}^{\text{left}}, \mathsf{U}^{\text{right}})$ *is a morphism of GPS units.)*

Proof. It is enough to prove the equation obtained by pasting the 2-cell $X\mathsf{U}Y$ on top of each side of the equation. This enables us to use the characterising equation for K and H. After this rewriting, we are in position to apply Equations (13) and (14), and after cancelling R and L cells, the resulting equation amounts to a cube, where it is easy to see that each side is just $\mathsf{U}^{\text{right}}_X \mathsf{U}^{\text{left}}_Y$. $\square$

7.6. From GPS units to cancellable-idempotent units. Given a GPS unit $(I, \lambda, \rho, \mathsf{K})$, just put

$$\alpha := \lambda_I,$$

then (I, α) is a unit object in the sense of 2.9. Indeed, we already observed that I is cancellable (6.4), and from the outset λ_I is an equi-arrow. That's all! To construct it we didn't even need the Kelly cell, or any of the auxiliary cells or their axioms.

7.7. Left and right actions from the Kelly cell. Start with natural left and right constraints λ and ρ and a Kelly cell $\mathsf{K} : X\lambda_Y \Rightarrow \rho_X Y$ (not required to be coherent on either side). Construct K^λ as in 6.5, put $\alpha := \lambda_I$, and define left and right actions as follows. We define L_X as

$$I\lambda_X \overset{\mathsf{N}^\lambda}{\Rightarrow} \lambda_{IX} \overset{\mathsf{K}^\lambda}{\Rightarrow} \lambda_I X = \alpha X,$$

while we define R_X simply as

$$X\alpha = X\lambda_I \overset{\mathsf{K}_{X,I}}{\Rightarrow} \rho_X I.$$

7.8 Proposition. *For fixed* $(I, \lambda, \rho, \mathsf{K})$*, the following are equivalent:*

(i) The left and right 2*-cells* L *and* R *just constructed in 7.7 are compatible with the Kelly cell in the sense of Equation* (24).

(ii) The Kelly cell K *is coherent on the left (i.e. satisfies Axiom* (17)*).*

(ii') The Kelly cell K *is coherent on the right (i.e. satisfies Axiom* (18)*).*

Proof. Proposition 7.4 already says that (i) implies both (ii) and (ii'). To prove (ii)⇒(i), we start with an auxiliary observation: by massaging the naturality equation

$$
\begin{tikzcd}[column sep=large, row sep=large]
XIIY \arrow[r, bend left, "\rho_X IY"] \arrow[r, bend right, "X\lambda_{IY}"'] \arrow[r, phantom, "\mathsf{K}_{X,IY}"] \arrow[d, "XI\lambda_Y"'] \arrow[dr, phantom, "X\lambda_{\lambda_Y}"] & XIY \arrow[d, "X\lambda_Y"] \\
XIY \arrow[r, bend right, "X\lambda_Y"'] & XY
\end{tikzcd}
=
\begin{tikzcd}[column sep=large, row sep=large]
XIIY \arrow[r, bend left, "\rho_X IY"] \arrow[d, "XI\lambda_Y"'] \arrow[dr, phantom, "©"] & XIY \arrow[d, "X\lambda_Y"] \\
XIY \arrow[r, bend left, "\rho_X Y"] \arrow[r, bend right, "X\lambda_Y"'] \arrow[r, phantom, "\mathsf{K}_{X,Y}"] & XY
\end{tikzcd}
$$

we find the equation

$$
\begin{tikzcd}[column sep=large, row sep=large]
XIIY \arrow[r, bend left, "X\lambda_{IY}"] \arrow[r, bend right, "XI\lambda_Y"'] \arrow[r, phantom, "X\mathsf{N}^\lambda"] \arrow[d, "\rho_X IY"'] \arrow[dr, phantom, "©"] & XIY \arrow[d, "\rho_X Y"] \\
XIY \arrow[r, bend right, "X\lambda_Y"'] & XY
\end{tikzcd}
=
\begin{tikzcd}[column sep=large, row sep=large]
XIIY \arrow[r, bend left, "X\lambda_{IY}"] \arrow[r, bend right, "\rho_X IY"'] \arrow[r, phantom, "\mathsf{K}^{-1}_{X,IY}"] \arrow[d, "\rho_X IY"'] \arrow[dr, phantom, "©"] & XIY \arrow[d, bend right, "X\lambda_Y"'] \arrow[d, bend left, "\rho_X Y"] \arrow[d, phantom, "\mathsf{K}_{X,Y}"] \\
XIY \arrow[r, bend right, "X\lambda_Y"'] & XY,
\end{tikzcd}
\tag{25}
$$

tailor-made to a substitution we shall perform in a moment.

Now for the main computation, assuming first that K is coherent on the left, i.e. that Axiom (17) holds. Start with the left-hand side of Equation (24), and insert the definitions we made for L and R to arrive at

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
XIIY \arrow[r, bend left=40, "X\lambda_I Y"] \arrow[r, "X\lambda_{IY}" description] \arrow[r, phantom, shift left=3ex, "X\mathsf{K}^\lambda"] \arrow[r, phantom, shift right=3ex, "X\mathsf{N}^\lambda"] \arrow[r, bend right=40, "XI\lambda_Y"'] \arrow[d, bend right, "X\lambda_I Y"'] \arrow[d, bend left, "\rho_X IY"] \arrow[d, phantom, "\mathsf{K}_{X,I}Y"] \arrow[dr, phantom, "©"] & XIY \arrow[d, "\rho_X Y"] \\
XIY \arrow[r, "X\lambda_Y"'] & XY
\end{tikzcd}
$$

in which we can now substitute (25) to get

$$\begin{tikzcd}[column sep=huge, row sep=huge]
XIIY \arrow[r, bend left=40, "X\lambda_I Y"] \arrow[r, "X\lambda_{IY}" description] \arrow[r, bend right=40, "\rho_X IY"'] \arrow[r, phantom, shift left=3ex, "X\mathsf{K}^\lambda"] \arrow[r, phantom, shift right=3ex, "\mathsf{K}^{-1}_{X,IY}"] \arrow[d, bend right=30, "X\lambda_I Y"'] \arrow[d, bend left=30, "\rho_X IY"] \arrow[d, phantom, "\mathsf{K}_{X,I}Y"] \arrow[dr, phantom, "\copyright"] & XIY \arrow[d, bend right=30, "X\lambda_Y"'] \arrow[d, bend left=30, "\rho_X Y"] \arrow[d, phantom, "\mathsf{K}_{X,Y}"] \\
XIY \arrow[r, "X\lambda_Y"'] & XY
\end{tikzcd}$$

Here finally the three 2-cells incident to the $XIIY$ vertex cancel each other out, thanks to Axiom (17), and in the end, remembering $\alpha = \lambda_I$, we get

$$\begin{tikzcd}
XIIY \arrow[dr, "X\alpha Y"] & \\
& XIY \arrow[dr, bend left=30, "\rho_X Y"] \arrow[dr, bend right=30, "X\lambda_Y"'] \arrow[dr, phantom, "\mathsf{K}"] & \\
& & XY
\end{tikzcd}$$

as required to establish that K satisfies Equation (24). Hence we have proved (ii)⇒(i), and therefore altogether (ii)⇒(ii'). The converse, (ii')⇒(ii) follows now by left-right symmetry of the statements. (But note that the proof via (i) is not symmetric, since it relies on the definition $\alpha = \lambda_I$. To spell out a proof of (ii')⇒(ii), use rather $\alpha = \rho_I$, observing that the intermediate result (i) would refer to different L and R.) □

We have now given a construction in each direction, but both constructions involved choices. With careful choices, applying one construction after the other in either way gets us back where we started. It is clear that this should constitute an equivalence of 2-categories. However, the involved choices make it awkward to make the correspondence functorial directly. (In technical terms, the constructions are ana-2-functors.) We circumvent this by introducing an intermediate 2-category dominating both. With this auxiliary 2-category, the results we already proved readily imply the equivalence.

7.9. A correspondence of 2-categories of units. Let $\mathscr{U}$ be following 2-category. Its objects are septuples

$$(I, \alpha, \lambda, \rho, \mathsf{L}, \mathsf{R}, \mathsf{K}),$$

with equi-arrows

$$\alpha : II \to I, \quad \lambda_X : IX \to X, \quad \rho_X : XI \to X,$$

(and accompanying naturality 2-cell data), and natural invertible 2-cells

$$\mathsf{L} : I\lambda_X \Rightarrow \alpha X, \quad \mathsf{R} : X\alpha \Rightarrow \rho_X I, \quad \mathsf{K} : X\lambda_Y \Rightarrow \rho_X Y.$$

These data are required to satisfy Equation (24) (compatibility of K with L and R).

The arrows in $\mathscr{U}$ from $(I, \alpha, \lambda, \rho, \mathsf{L}, \mathsf{R}, \mathsf{K})$ to $(J, \beta, \ell, r, \mathsf{L}', \mathsf{R}', \mathsf{H})$ are quadruples

$$(u, \mathsf{U}^{\text{left}}, \mathsf{U}^{\text{right}}, \mathsf{U}),$$

where $u : I \to J$ is an arrow in $\mathscr{C}$, U^{left} and $\mathsf{U}^{\text{right}}$ are as in 6.10, and U is a morphism of pseudo-idempotents from (I, α) to (J, β). These data are required to satisfy Equation (20) (compatibility with Kelly cells) as well as Equations (13) and (14) in Lemma 5.1 (compatibility with the left and right 2-cells).

Finally a 2-cell from $(u, \mathsf{U}^{\text{left}}, \mathsf{U}^{\text{right}}, \mathsf{U})$ to $(v, \mathsf{V}^{\text{left}}, \mathsf{V}^{\text{right}}, \mathsf{V})$ is a 2-cell

$$\mathsf{T} : u \Rightarrow v$$

required to be a 2-morphism of pseudo-idempotents (compatibility with U and V as in 2.5), and to satisfy Equation (15) (compatibility with U^{left} and V^{left}) as well as Equation (16) (compatibility with $\mathsf{U}^{\text{right}}$ and $\mathsf{V}^{\text{right}}$).

Let $\mathscr{E}$ denote the 2-category of cancellable-idempotent units introduced in 2.9, and let $\mathscr{G}$ denote the 2-category of GPS units of 6.10. We have evident forgetful (strict) 2-functors

$$\begin{tikzcd} & \mathscr{U} \arrow[dl, "\Phi"'] \arrow[dr, "\Psi"] & \\ \mathscr{E} & & \mathscr{G}. \end{tikzcd}$$

Theorem E (Equivalence). *The 2-functors Φ and Ψ are 2-equivalences. More precisely they are surjective on objects and strongly fully faithful (i.e. isomorphisms on hom categories).*

Proof. The 2-functor Φ is surjective on objects by Lemma 3.1 and Proposition 7.4. Given an arrow (u, U) in $\mathscr{E}$ and overlying objects in $\mathscr{U}$, Lemma 5.1 says there are unique U^{left} and $\mathsf{U}^{\text{right}}$, and Proposition 7.5 ensures the required compatibility with Kelly cells (Equation (20)). Hence Φ induces a bijection on objects in the hom categories. Lemma 5.2 says we also have a bijection on the level of 2-cells, hence Φ is an isomorphism on hom categories. On the other hand, Ψ is surjective on objects by 7.7 and Proposition 7.8. Given an arrow $(u, \mathsf{U}^{\text{left}}, \mathsf{U}^{\text{right}})$ in $\mathscr{G}$, Lemma 7.10 below says that for fixed overlying objects in $\mathscr{U}$ there is a unique associated U, hence Ψ induces a bijection on objects in the hom categories. Finally, Lemma 5.2 gives also a bijection of 2-cells, hence Ψ is strongly fully faithful. ☐

7.10 Lemma. *Given a morphism of GPS units*

$$(I, \lambda, \rho, \mathsf{K}) \xrightarrow{(u, \mathsf{U}^{\text{left}}, \mathsf{U}^{\text{right}})} (J, \ell, r, \mathsf{H})$$

fix an equi-arrow $\alpha : II \overset{\sim}{\to} I$ with natural families $\mathsf{L}_X : I\lambda_X \Rightarrow \alpha X$ and $\mathsf{R}_X : \alpha X \Rightarrow \rho_X I$ satisfying Equation (24) (compatibility with K), and fix an equi-arrow $\beta : JJ \overset{\sim}{\to} J$ with natural families $\mathsf{L}'_X : I\ell_X \Rightarrow \beta X$ and $\mathsf{R}'_X : \beta X \Rightarrow r_X I$ also satisfying Equation (24) (compatibility with H). Then there is a unique 2-cell

$$\begin{tikzcd} II \arrow[r, "uu"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{U}"] & JJ \arrow[d, "\beta"] \\ I \arrow[r, "u"'] & J \end{tikzcd}$$

satisfying Equations (13) and (14) (compatibility with U^{left} and the left 2-cells, as well as compatibility with $\mathsf{U}^{\text{right}}$ and the right 2-cells).

Proof. Working first with left 2-cells, define a family W_X by the equation

$$\begin{tikzcd} IIX \arrow[r, "uuX"] \arrow[d, bend right=60, "I\lambda_X"'] \arrow[d, bend left=60, "\alpha X"] \arrow[d, phantom, "\mathsf{L}_X"] \arrow[dr, phantom, "\mathsf{W}_X"] & JJX \arrow[d, bend left=40, "\beta X"] \\ IX \arrow[r, "uX"'] & JX \end{tikzcd} = \begin{tikzcd} IIX \arrow[r, "uuX"] \arrow[d, bend right=40, "I\lambda_X"'] \arrow[dr, phantom, "u\mathsf{U}^{\text{left}}_X"] & JJX \arrow[d, bend right=60, "J\ell_X"'] \arrow[d, bend left=60, "\beta X"] \arrow[d, phantom, "\mathsf{L}'_X"] \\ IX \arrow[r, "uX"'] & JX \end{tikzcd}$$

It follows readily from Equation (21) that the family has the property

$$\mathsf{W}_{XY} = \mathsf{W}_X Y$$

for all X, Y, and it is a standard argument that since a unit object exists, for example (I, λ_I), this implies that

$$\mathsf{W}_X = \mathsf{U}X$$

for a unique 2-cell

$$\begin{tikzcd} II \arrow[r, "uu"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\mathsf{U}"] & JJ \arrow[d, "\beta"] \\ I \arrow[r, "u"'] & J, \end{tikzcd}$$

and by construction this 2-cell has the required compatibility with U^{left} and the left constraints. To see that this U is also compatible with $\mathsf{U}^{\text{right}}$ and the right constraints we reason backwards: (u, U) is now a morphisms of units from (I, α) to (J, β) to which we apply the right-hand version of Lemma 5.1 to construct a new $\mathsf{U}^{\text{right}}$, characterised by the compatibility condition. By Proposition 7.5 this new $\mathsf{U}^{\text{right}}$ is compatible with U^{left} and the Kelly cells K and H (Equation (20)), and hence it must in fact be the original $\mathsf{U}^{\text{right}}$ (remembering from 6.10 that U^{left} and $\mathsf{U}^{\text{right}}$ determine each other via (20)). So the 2-cell U does satisfy both the required compatibilities. ☐

References

[1] Jean Bénabou. *Introduction to bicategories.* In *Reports of the Midwest Category Seminar*, no. 47 in Lecture Notes in Mathematics, pp. 1–77. Springer-Verlag, Berlin, 1967.

[2] Robert Gordon, John Power, and Ross Street. *Coherence for tricategories.* Mem. Amer. Math. Soc. 117 (1995), vi+81pp.

[3] Nick Gurski. *An algebraic theory of tricategories.* PhD thesis, University of Chicago, 2006. Available from `http://www.math.yale.edu/~mg622/tricats.pdf`.

[4] André Joyal and Joachim Kock. *Weak units and homotopy 3-types.* In A. Davydov, M. Batanin, M. Johnson, S. Lack, and A. Neeman, editors, *Ross Street Festschrift: Categories in algebra, geometry and mathematical physics*, vol. 431 of Contemp. Math., pp. 257–276. Amer. Math. Soc., Providence, RI, 2007. ArXiv:math.CT/0602084.

[5] Max Kelly. *On MacLane's conditions for coherence of natural associativities, commutativities, etc.* J. Algebra 1 (1964), 397–402.

[6] Max Kelly and Ross Street. *Review of the elements of 2-categories.* In *Category Seminar (Proc. Sem., Sydney, 1972/1973)*, no. 420 in Lecture Notes in Mathematics, pp. 75–103. Springer-Verlag, Berlin, 1974.

[7] Joachim Kock. *Weak identity arrows in higher categories.* IMRP Internat. Math. Res. Papers 2006 (2006), 1–54. ArXiv:math.CT/0507116.

[8] Joachim Kock. *Elementary remarks on units in monoidal categories.* Math. Proc. Cambridge Philos. Soc. 144 (2008), 53–76. ArXiv:math.CT/0507349.

[9] Saunders Mac Lane. *Natural associativity and commutativity.* Rice Univ. Studies 49 (1963), 28–46.

[10] Neantro Saavedra Rivano. *Catégories Tannakiennes.* No. 265 in Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1972.

[11] Carlos Simpson. *Homotopy types of strict 3-groupoids.* ArXiv:math.CT/9810059.

[12] Jim Stasheff. *Homotopy associativity of H-spaces. I, II.* Trans. Amer. Math. Soc. 108 (1963), 275–292; 293–312.

André Joyal
Département de Mathématiques
Université du Québec à Montréal
joyal.andre@uqam.ca

Joachim Kock
Departament de Matemàtiques
Univ. Autònoma de Barcelona
kock@mat.uab.cat