A Systematic Proof Theory for Several Modal Logics

Charles Stewart and Phiniki Stouppa

ABSTRACT. The family of normal propositional modal logic systems is given a very systematic organisation by their model theory. This model theory is generally given using frame semantics, and it is systematic in the sense that for the most important systems we have a clean, exact correspondence between their constitutive axioms as they are usually given in a Hilbert-Lewis style and conditions on the accessibility relation on frames.

By contrast, the usual structural proof theory of modal logic, as given in Gentzen systems, is ad-hoc. While we can formulate several modal logics in the sequent calculus that enjoy cut-elimination, their formalisation arises through system-by-system fine tuning to ensure that the cut-elimination holds, and the correspondence to the axioms of the Hilbert-Lewis systems becomes opaque.

This paper introduces a systematic presentation for the systems **K**, **D**, **M**, **S**4, and **S**5 in the calculus of structures, a structural proof theory that employs *deep inference*. Because of this, we are able to axiomatise the modal logics in a manner directly analogous to the Hilbert-Lewis axiomatisation. We show that the calculus possesses a cut-elimination property directly analogous to cut-elimination for the sequent calculus for these systems, and we discuss the extension to several other modal logics.

1 Introduction

Modal logic has been one of the most fundamental advances in logic since Frege's invention of the quantifier, and indeed arguably the most important innovation of modern logic. We might attribute this success to:

- The ability of modal notations to naturally accomodate key concepts needed in formal description, most fundamentally concepts of mode and tense. While many propositional modal logics can be encoded in first-order predicate logic in a straightforward way, for many applications the modal systems are simpler and more useful;

- Most of the widely used modal logics being decidable;

- The normal modal logics possession of an elementary model theory in the form of frame semantics (due to Kripke, Hintikka and Kanger, see [8]) that provides a systematic correspondence between their constitutive axioms as they are usually given in Hilbert style and conditions on the accessibility relation on frames[1].

Let's look at the toolkit available to the classical logician for propositional modal logic compared to that for propositional logic. In propositional logic we characterise consequence relations using Hilbert deductive systems, give our semantics using Boolean algebra valued models, provide decision procedures using tableau methods, and provide techniques for proof analysis using sequent calculi in the mould of Gentzen.

In modal logic, close substitutes for the first three of these tools can be found: Hilbert systems are extended in a standard way with rules to provide what we might call Hilbert-Lewis deductive systems; Kripke or Beth frame semantics provides a sufficiently rich universe to provide models for almost all modal logics we care about; and tableau systems can be extended to accomodate modal logics. However, when it comes to proof analysis the situation is more fraught. A relatively natural extension of Gentzen-style sequent calculi provides cut-free characterisations of several important modal logics; however it fails

[1] See Garson [12], and Blackburn, de Rijke and Venema [3] for readable accounts of this relationship.

to provide characterisations of others, and even in the cases where sequent systems can be provided, these are arrived at in an ad-hoc manner[2], and the relationship to the Hilbert-Lewis calculus becomes opaque. It's not an exaggeration to say that, by and large, modal logicians have not found proof analysis to be worth the effort; this, no doubt, is responsible for the omission of structural proof theory in texts such as Blackburn, de Rijke and Venema [3].

In this paper we will provide a new proof theoretic basis for modal logic using the calculus of structures introduced by Guglielmi [14]. This proof calculus provides a structural proof theory in the sense of Gentzen for logical systems because it possesses a notion of cut-free proof directly analogous to that for the sequent calculus. It possesses several properties that give proof theory carried out in the calculus of structures important advantages over the traditional proof theory carried out in structural proof theory, and which extend to the modal logic systems we cover. It is our hope that these new possibilities for proof analysis can reinvigorate this neglected corner of the modal logician's toolkit.

The structure of the rest of this paper is as follows. In the next section we recall the characterisation of a much studied 'cube' of 15 normal modal logics in Hilbert-Lewis deductive systems, and describe the sequent calculus for just one of these, the system M (also known as system T). In section 3 we present the calculus of structures for classical logic, introduce its extension to system M, and talk about some of the most important analytical properties the calculus has in detail, including cut-elimination. In section 4 we discuss the generalisation of the structural proof theory to the other modal logics of the cube in both sequent calculus and the calculus of structures. In section 5 we provide a critical and slightly philosophical discussion comparing the achievements of this calculus to the rival calculus that is best placed to challenge us in providing a highly general approach to proof analysis, namely Belnap's display logic as applied to modal

[2] I do not mean by this that the designer of sequent systems has no methodology available other than trial and error, rather I mean that this methodology consists of a complex set of heuristics of low generality, and so falls short of being systematic.

logic by Wansing, as well as the more conservative but nevertheless highly interesting approach of Pottinger's hypersequent calculus as investigated by Avron. Finally, section six provides a brief, critical summary of what has been achieved in this paper.

2 Proof theory of system M

Let us begin by establishing some terminology. A *logical system* is defined over a formal language by a deductive system that determines its *theory*, a subset of the sentences of the language. Each formula in the theory of a logical system is called a *theorem*. Logical systems are *schematic*, that is they allow schematic letters that may stand in for arbitrary propositions, so the theory of a logical system for a more restrictive language can naturally be projected to richer languages by application of substitution. Two logical systems are *equivalent* if their theories are the same, they are *distinct* otherwise. If the theory of one logical system strictly contains the theory of another, then we call the former system the stronger system.

We begin by showing how propositional modal logic is defined using Hilbert-Lewis deductive systems, as these proof calculi give the oldest and most common axiomatisations, and so we may consider these axiomatisations as constituting a standard. The sentences of classical propositional logic contain an inexhaustible supply of schematic letters, letters for truth and falsity, which we indicate $\mathbf{tt}$ and $\mathbf{ff}$, and are closed under the binary operators $\supset$, $\leftrightarrow$, $\wedge$ and $\vee$ and the unary operator $\neg$. The Hilbert style axiomatisation of this logical system proceeds in the standard way, where any substitution of propositions for schematic letters in the axioms of the system may be used to start a derivation or add a new theorem to the derivation, and non-axiomatic propositions may be obtained by modus ponens.

Hilbert-Lewis systems differ from Hilbert systems in that they possess (at least) a pair of dual modalities, and characteristically are schematic not through the schematicity of the axioms, but via the rule *uniform substitution*, which allows the schematic letters to be replaced by any proposition; the consequences of this difference are that deductions are normally a little longer, but it is slightly easier to define some metalogical apparatus; in any case the consequence relation should be unaffected by this difference, and in this paper we will be casual about the difference, defining axioms as sentences in the spirit of Hilbert-Lewis systems, but applying them as if they were schematic. In addition, all of the Hilbert-Lewis systems in which we are here interested will have the single further rule of necessitation: if $\vdash p$ then $\vdash \Box p$, and will have the duality captured by the axiom **DM**: $\neg \Box p \leftrightarrow \Diamond \neg p$.

Then the weakest normal modal logic **K** is axiomatised as a Hilbert-Lewis system by extending classical propositional logic with the axiom **K**: $\Box(p \supset q) \supset (\Box p \supset \Box q)$. Naturally, this corresponds to the class of frame models[3] where there are no conditions imposed on the frame accessibility relation.

We then obtain the stronger modal logics we are interested in by adding further axioms. All of the most studied systems of modal logic arise by adding some subset of the following axioms:

DEFINITION 1

1. *Axiom* ***D****:* $\Box p \supset \Diamond p$
2. *Axiom* ***T****:* $\Box p \supset p$
3. *Axiom* ***4****:* $\Box p \supset \Box \Box p$
4. *Axiom* ***B****:* $p \supset \Box \Diamond p$
5. *Axiom* ***5****:* $\Diamond p \supset \Box \Diamond p$

which correspond to constraining the accessibility relation to be serial (ie. there are always nodes accessible from any node), reflexive, transitive, symmetric and Euclidean respectively. There are a total of fifteen distinct logical systems obtainable, which we may organise

[3]We avoid the terminology of possible worlds, since we consider model systems lacking the T axiom, that cannot reasonably be considered to be about possibility or contingency, and further we agree with Thomas Forster that the language of possible world semantics is intoxicating to the careless and is best avoided[11]. Hence we call Hintikka/Kripke/Kanger style semantics, frame semantics; we avoid formalising these but will need to talk about the frame accessibility relation, which is given over nodes.

in a lattice according to the 'stronger than' relation, with the most important of these being:

1. **K** itself;
2. **D**, obtained by extending system **K** with rule **D**;
3. **M**, obtained by extending system **K** with rule **T**: note this system is often named **T**, a nomenclature we avoid to prevent confusion with Gödel's system T.
4. **S4**, obtained by extending system **M** with rule **4**;
5. **B**, obtained by extending system **M** with rule **B**;
6. **S5**, obtained by extending system **M** with rule **5**, or equivalently by extending system **M** with rules **B** and **4**.

The strongest system from these modal logics that is perfectly straightforward to formulate in a sequent system and to prove cut-free is system **G-M** (for Gentzen system **M**): we formulate[4] this using Schütte's approach of one-sided sequents. We make use of a *dualising function* in formulating the syntax of the calculus: $\overline{\phi}$ is defined recursively to be the De Morgan dual of the formula ϕ:

$$\begin{array}{cccc} \overline{\neg A} = A & \overline{A \wedge B} = \overline{A} \vee \overline{B} & \overline{\Box A} = \Diamond \overline{A} & \overline{\mathbf{tt}} = \mathbf{ff} \\ \overline{p} = \neg p & \overline{A \vee B} = \overline{A} \wedge \overline{B} & \overline{\Diamond A} = \Box \overline{A} & \overline{\mathbf{ff}} = \mathbf{tt} \end{array}$$

where p is a schematic letter, and $A \supset B$ and $A \leftrightarrow B$ are treated as derived connectives using the standard encodings. The inferences of the system are given by trees generated by the following inference rules, where the nodes are multisets of formulae (indicated by Γ, Δ, with the notations $\Box\Gamma, \Diamond\Delta$ indicating the result of prefixing each formula of the multiset appropriately):

[4] It is possible to show that, given liberal constraints on the form of the sequent calculus, 2-sided (Gentzen style) systems and 1-sided (Schütte style) systems characterise exactly the same consequence relations in the presence of De Morgan dualities [21], and this demonstration generalises to hypersequents.

Axiom and cut:

$$\frac{}{\vdash \Gamma, A, \overline{A}}\ \mathrm{ax} \qquad \frac{\vdash \Gamma, A \qquad \vdash \Delta, \overline{A}}{\vdash \Gamma, \Delta}\ \mathrm{cut}$$

Contraction and weakening:

$$\frac{\vdash \Gamma, A, A}{\vdash \Gamma, A}\ \mathrm{contr} \qquad \frac{\vdash \Gamma}{\vdash \Gamma, A}\ \mathrm{wk}$$

Truth:

$$\frac{}{\vdash \mathbf{tt}}\ \mathbf{tt}$$

Logical connectives and modal operators:

$$\frac{\vdash \Gamma, A, B}{\vdash \Gamma, A \vee B}\ \vee \qquad \frac{\vdash \Gamma, A \qquad \vdash \Gamma, B}{\vdash \Gamma, A \wedge B}\ \wedge$$

$$\frac{\vdash \Gamma, \overline{A}}{\vdash \Gamma, \neg A}\ \neg$$

$$\frac{\vdash \Gamma, A}{\vdash \diamond\Gamma, \Box A}\ \Box_1 \qquad \frac{\vdash \Gamma, A}{\vdash \Gamma, \diamond A}\ \diamond_1$$

The theorems of this system then are given by the concluding formulae of inferences where the concluding sequent contains exactly one formula. It is easily shown that the sequent formulation contains all the theorems of the Hilbert-Lewis formulation, and with the lightest touch of ingenuity, modelling the multisets by disjunctions, the reverse containment can also be demonstrated. We can also show the following theorem, where an inference rule or set of inference rules are *admissible* if all theorems of the system may be proven without their use:

THEOREM 2 *The cut rule is admissible in the sequent calculus formalisation of system* $\boldsymbol{M}$.

This cut-elimination theorem is not really more complex to prove than that for its subsystem of classical logic, though there is a novelty: since the inference rules for the modal operators introduce 'side-effects'; the rule $\Box_1$ adds $\diamond$ modalities to side assumptions, which can then be used in cuts. This rule doesn't interfere with permutability of cuts, however. These side effects are are characteristic of an important sense in which these 'modal Gentzen calculi' go beyond traditional Gentzen calculi: with the traditional calculi only a formula in the consequence, and some subset of its subformulae in the premisses, are 'touched' by the logical rule, hence they are what we call *strongly focussed*; modal Gentzen calculi with rules such as $\Box_1$ we call *strongly unfocussed*, where we say a system is *weakly unfocussed* if it is not strongly focussed, and is *weakly focussed* if it is not strongly unfocussed[5]. Later we will encounter weakly unfocussed calculi.

3 The calculus of structures

Now we introduce a characterisation of the system $\mathbf{M}$ in the calculus of structures. As for the single-sided sequent calculus, we make use of De Morgan dualities between connectives in the formulation of the system. By convention, the calculus has a rather different notation for formulae than is used in Hilbert-Lewis and Gentzen proof calculi: the calculus of structures uses a bracket notation for the propositional logical connectives to suggest the associativity and commutativity properties: the calculus treats *structures* that are equivalence classes of the syntactic representations quotiented over these relations, and also quotiented over rules for units. Conjunction is represented by "$(\ldots)$", so (R, S) stands for the conjunction of R and S; likewise disjunction is represented by "$[\ldots]$". We also introduce a dualising operator in the same way as for the single-sided sequent calculus.

So the *formulae* of the calculus of structures are built up from the units $\mathbf{tt}$, $\mathbf{ff}$, the schematic letters, where for each schematic letter a we admit the complement $\overline{a}$ as a formula, and whenever $R_1, \ldots, R_n$ are formulae, so are $[R_1, \ldots, R_n]$ and $(R_1, \ldots, R_n)$. A *formula context* $S\{-\}$ is obtained from a formula by replacing any leaf (eg. a schematic letter) by '$-$'; then $S\{R\}$ is the formula obtained by re-

[5]The point of the terminology of focussing is to say that a calculus is focussed if each rule that deals with a connective is only about that connective. As such this is related to the properties such as separation as defined by Wansing [27], but neither strong nor weak focussedness is expressible in terms of Wansing's properties, and our interest here is technical, concerned with permutability of cut, rather that the meaning theoretic issues that motivated Wansing.

placing this $-$ by the formula R. Curly braces are ommited when the formula R is precisely of the form $(R_1, \ldots, R_n)$ or $[R_1, \ldots, R_n]$.

The *duals* of formulae are defined recursively, where the dual of each schematic letter is its complement and vica versa, and

$$\begin{array}{c}\overline{(R_1, \ldots, R_n)} = [\overline{R_1}, \ldots, \overline{R_n}] \\ \overline{[R_1, \ldots, R_n]} = (\overline{R_1}, \ldots, \overline{R_n}) \\ \overline{\mathbf{tt}} = \mathbf{ff} \qquad \overline{\mathbf{ff}} = \mathbf{tt}\end{array}$$

so for any formula R, $R = \overline{\overline{R}}$.

The *structures* are the equivalence classes of formulae obtained by quotienting over

1. (Associativity)
   $[R_1, \ldots, R_i, [T_1, \ldots, T_j], U_1, \ldots, U_k] = [R_1, \ldots, R_i, T_1, \ldots, T_j, U_1, \ldots, U_k]$
   and
   $(R_1, \ldots, R_i, (T_1, \ldots, T_j), U_1, \ldots, U_k) = (R_1, \ldots, R_i, T_1, \ldots, T_j, U_1, \ldots, U_k)$
2. (Congruence) If $R_1 = R_2$ then $S\{R_1\} = S\{R_2\}$, for any formula context $S\{-\}$;
3. (Commutativity) $[R_1, R_2] = [R_2, R_1]$ and $(R_1, R_2) = (R_2, R_1)$;
4. (Identities) $R = [R, \mathbf{ff}]$ and $R = (R, \mathbf{tt})$.

The equations can be considered as being in the spirit of the equations Schütte introduced over propositional formulae that ensure each formula is equal to its negation normal form, but they go further, representing more complex, though still tractable, equivalences over formulae. Their presence makes the calculus rather more difficult to master than the sequent calculus (especially the identities present difficulties) but they make working with the system much easier once this initial hurdle is passed, and normally ensure that important theorems avoid the trivial issues of syntax that they would otherwise be burdened with.

Inferences are chains of applications of inference rules, where each inference rule relating one premise, given above the rule, to one conclusion, below, whiere premise and conclusion are structures rather than formulae. All of the inference rules are *deep*, which means that each rule is given by a pair of formulae specifying the premise and conclusion that are both given with the same formula context. The inference rules for classical logic, the system **SKSg** are given:

Interaction and cut rules:

$$i\downarrow \frac{S\{\mathbf{tt}\}}{S[R,\overline{R}]} \qquad i\uparrow \frac{S(R,\overline{R})}{S\{\mathbf{ff}\}}$$

The switch rule:

$$s\,\frac{S([R,T],U)}{S[(R,U),T]}$$

The weakening rules:

$$w\downarrow \frac{S\{\mathbf{ff}\}}{S\{R\}} \qquad w\uparrow \frac{S\{R\}}{S\{\mathbf{tt}\}}$$

And the contraction rules:

$$c\downarrow \frac{S[R,R]}{S\{R\}} \qquad c\uparrow \frac{S\{R\}}{S(R,R)}$$

The theorems of this system are the formulae belonging to the structures that occur as conclusions of inferences whose premise is the structure $\mathbf{tt}$.

This system has many desirable properties:

1. The system is self dual: each rule labelled with $\uparrow$ takes form:

   $$*\uparrow \frac{S\{R\}}{S\{T\}}$$

   matching by De Morgan duality the form of the corresponding rule labelled with $\downarrow$:

   $$*\downarrow \frac{S\{\overline{T}\}}{S\{\overline{R}\}}$$

   where rules not labelled with $\uparrow$ or $\downarrow$ are self-dual (so far we have seen just the switch rule).

2. The entire *up-fragment*, ie. the rules labelled with $\uparrow$, is admissible; that is, the full system is equivalent to the system obtained by removing the whole up-fragment. This is shown by Bruennler [6] and discussed in more detail below, by means of a translation from cut-free proofs of the sequent calculus into proofs of system **KSg**, that is, system **SKSg** without the rules of the up-fragment. Because the rule $i\uparrow$ closely models the cut rule, it is natural to describe this admissibility result as cut-elimination for the calculus of structures.

3. We can restrict the interaction, cut and weakening rules to atoms, by which we mean that the applications of the rules using formulae R can be restricted to the case where R is an atom (ie. a or $\overline{a}$) both for system **SKSg** and the cut-free system **KSg**. Furthermore, when adding to system **SKSg** the *medial rule*
$$m\,\frac{S[(R,U),(T,V)]}{S([R,T],[U,V])}$$
we can also restrict the contraction rule to atoms. These restrictions are achieved by simple local transformations on inferences, as opposed to the complex global transformation that we associate with cut-elimination. The system with these restrictions we call *atomic* **SKSg** (analogously atomic **KSg**), or just **SKS** (analogously **KS**). An analogous restriction to atoms can be achieved for the sequent calculus by similarly local transformations in the case of the axiom and weakening rules, but for fundamental reasons of the shallowness of inference cannot be achieved at all for contraction, and only by means of global transformations in the case of cut.

4. System **SKS** enjoys important computational properties: it is *local*, and so too is its subsystem **KS**, in the sense that looking at the inferences going either up or down, structure is rearranged, or atoms introduced, abandoned or duplicated, but arbitrarily large substructures are never introduced, abandoned or duplicated. Brünnler also discusses [op cit] an important advantage corollary to locality and atomicity of cut: using deep

inference he gets a *finitary* variant of **SKS** by simple means (i.e. without cut elimination); this means that there are only finitely different ways rules of the system can be applied to obtain a given conclusion, which in principle means the calculus can be used for proof search. However, it is fair to point out that these benefits do carry a cost: by contrast to the sequent calculus there are a great many possible rules that one may apply during proof search, so one cannot read off a tableaux algorithm from **KS** in the way one can for cut-free **LK**. Devising efficient proof search algorithms that can take advantage of these properties is an important goal of the program of research in the calculus of structures.

A thorough mathematical and conceptual examination of these properties and their importance is given in Bruennler's [6]; a shorter discussion appears in [5].

We can extend this calculus to obtain the system **M** by allowing formulae of the form $\Box R$, and $\diamond R$, extending the equivalences defining structures with $\mathbf{tt} = \Box\mathbf{tt}$ and $\mathbf{ff} = \diamond\mathbf{ff}$ and extending the set of inferences as follows:

$$k\downarrow\frac{S\{\Box[R,T]\}}{S[\Box R,\diamond T]} \qquad t\downarrow\frac{S\{\Box R\}}{S\{R\}}$$

We call the system that extends system **KSg** with the above down rules **KSg-M**, and the symmetric system extending system **SKSg** with both the down and up rules **SKSg-M**.

We show the equivalence of system **SKSg-M** to **M** in two steps. Firstly we show we can map inferences of **SKSg-M** onto inferences of **M**:

DEFINITION 3 *We define the map $-^s$ from formulae of **M** onto structures of **SKSg-M** recursively:*

$$p^s = p \text{ for } p \text{ one of } \mathbf{tt},\ \mathbf{ff},\ \text{or a schematic letter;}$$

$$(\neg A)^s = \overline{A^s} \qquad (A\wedge B)^s = (A^s, B^s) \qquad (\Box A)^s = \Box A^s$$

$$(A\vee B)^s = [A^s, B^s] \qquad (\diamond A)^s = \diamond A^s$$

PROPOSITION 4

1. *There is a map* $-^h$ *from structures of* ***SKSg-M*** *to formulae of* ***M*** *with the properties that* $-^{hs}$ *is the identity map on structures, and for all formulae* A*,* B*, if* A *is a subformula of* B *then either* A^{sh} *is a subformula of* B^{sh} *or* $(\neg A)^{sh}$ *is a subformula of* $(\neg B)^{sh}$*.*

2. *If* $A^s = B^s$ *then* $\vdash A \leftrightarrow B$ *is a theorem of* ***M****;*

3. *If* $R = A^s$ *and* A *is a theorem of* ***M****, then* R *is a theorem of* ***SKSg-M****;*

Proof Part 1 follows from defining the obvious recursive map from the formulae of **SKSg-M**, and observing that for any structure we may choose a lexically simplest representative formula (given some total order on atoms and their complements), to obtain the reverse map which is easily verified to have the desired properties.

Part 2 follows easily from the observation that $-^s$ is bijective when restricted to disjunctive normal forms.

To prove part 3, we must first prove that each axiom of **M** maps onto a theorem of **SKSg-M**, which is a straightforward exercise in the construction of inferences in the calculus of structures: for didactic reasons we recommend the reader works through at least two cases, for example $(A \supset B \supset C) \supset (A \supset C) \supset B \supset C$, and axiom **K**, and then show that inferences of the Hilbert-Lewis calculus map onto inferences of **SKSg-M** by considering the three inference rules. To see that inferences making use of uniform substitutivity are modelled in **SKSg-M** we need simply observe that the inference rules of **SKSg-M** are schematic. In the other two rules proceed by an induction on the length of proofs. $\square$

THEOREM 5 ***M*** *and* ***SKSg-M*** *are equivalent.*

Proof For each inference rule of **KSg-M**, which takes the general form:

$$* \frac{S\{R\}}{S\{T\}}$$

we show that $R^h \supset T^h$ is a theorem of **M**. The dual rules map onto these from the theorem of classical logic: $(A \supset B) \leftrightarrow (\neg B \supset \neg A)$.

The following lemma is a straightforward exercise in theoremhood over **K**:

LEMMA 6 *If $A \supset B$ is a theorem of **M**, then so are:*

1. $A \wedge C \supset B \wedge C$;
2. $A \vee C \supset B \vee C$;
3. $\Box A \supset \Box B$;
4. $\Diamond A \supset \Diamond B$.

from which, by an induction on the makeup of formula contexts $S\{-\}$, we obtain for each inference rule $(S\{R\})^h \supset (S\{T\})^h$. Then inferences of **KSg-M** can be mapped onto chains of applications of modus ponens, starting from the theorem $\mathbf{tt}$. $\Box$

THEOREM 7 *The up-fragment of **SKSg-M** is admissible.*

Proof We show this by mapping cut-free proofs of **G-M** onto proofs of **KSg-M**, following just the same technique as [5], that is, for each rule of **G-M**

$$\frac{\Gamma_1 \qquad \dots \qquad \Gamma_n}{\Delta}$$

we must find an inference of **KSg-M** with premise $\Box^i(\Gamma_1^s, \dots, \Gamma_n^s)$ for some i and conclusion Δ^s, where we extend the mapping $-^s$ from formulae to sequents by considering sequents to be disjunctions of their constituent formulae, and where the notation $\Box^i A$ indicates the formulae obtained from A by prefixing $\Box$ to it i times. These proofs are straightforward; the only case where we need i to be nonzero is mapping $\Box_1$; the rule then maps onto a number of applications of the $k\downarrow$ rule equal to the number of formulae in Γ. We can then use these inferences to construct our map on cut-free proofs recursively. $\Box$

Lastly if we add the following rules:

$$j\downarrow \frac{S[\Diamond R, \Diamond T]}{S\{\Diamond[R,T]\}} \qquad uw\downarrow \frac{S\{\mathbf{ff}\}}{S\{\Box\mathbf{ff}\}} \qquad l\downarrow \frac{S[\Box R, \Box T]}{S\{\Box[R,T]\}}$$

we can restrict cut, interaction, contraction and weakening to atoms, and so obtain a local system, which we call **KS-M**.

4 Systems K, D, S4, and S5

We can extend the account of the structural proof theory that we have given of **M** to the systems **K**, **D**, and **S4** quite analogously. First we consider the axiomatisation of these systems in the sequent calculus: define two new inference rules:

$$\frac{\vdash \diamond\Gamma, A}{\vdash \diamond\Gamma, \Box A}\,\Box_2 \qquad \frac{\vdash \Gamma, A}{\vdash \diamond\Gamma, \diamond A}\,\diamond_2$$

Then we obtain our axiomatisations as follows:

1. **G-K** is obtained by dropping rule $\diamond_1$ from **G-M**;
2. **G-D** is obtained from **G-M** by substituting rule $\diamond_2$ for $\diamond_1$;
3. **G-S4** is obtained from **G-M** by substituting rule $\Box_2$ for $\Box_1$;

THEOREM 8 *Each of these systems is equivalent to the matching Hilbert system, and for each, the cut rule is admissible.*

PROOF For **G-K** and **G-D** the proofs are quite analogous to that for **G-M**. **G-S4** is similarly shown to be equivalent to the matching system **S4**; the cut-elimination proof is more subtle because many of the usual permutations on cuts fail, a fact that is unfortunately rather glossed over in the literature. Cf. Ohnishi and Matsumoto, and Valenti [17, 24, 25]. □

Our axiomatisation in the calculus of structures proceeds by defining the inference rules:

$$d\,\frac{S\{\Box R\}}{S\{\diamond R\}} \qquad 4\downarrow\frac{S\{\diamond\diamond R\}}{S\{\diamond R\}}$$

Then we obtain the equivalent systems:

1. **KSg-K** is **KSg-M** without the $t\downarrow$ rule;
2. **KSg-D** is **KSg-K** together with the $d\downarrow$ rule;
3. **KSg-S4** is **KSg-K** and the $t\downarrow$ and $4\downarrow$ rules;

4. **SKSg-K**, **SKSg-D**, and **SKSg-S4** are the symmetric versions of the above.

Note that inclusion of an axiom in the Hilbert system corresponds to inclusion of a down rule in the corresponding system in the calculus of structures.

THEOREM 9 *The analogs of Proposition 4 and Theorem 5 hold for each of* ***SKSg-K****,* ***SKSg-D****, and* ***SKSg-S4****.*

The axiomatisation of systems **S5** and **B** in the one-sided sequent calculus present serious difficulties, however. For example, it is possible to axiomatise **S5** by replacing the rule $\Box_1$ of **G-M** by any one of the following three rules[6]:

$$\frac{\vdash \Diamond\Gamma, \Box\Delta, A}{\vdash \Diamond\Gamma, \Box\Delta, \Box A}\,\Box_3$$
$$\frac{\vdash \Gamma, \Box\Delta, A}{\vdash \Diamond\Gamma, \Box\Delta, \Box A}\,\Box_4$$
$$\frac{\vdash \Diamond\Gamma, \Delta, A}{\vdash \Diamond\Gamma, \Box\Delta, \Box A}\,\Box_5$$

This system can be mapped onto a reasonable axiomatisation of **KSg-S5** which we axiomatise as **KSg-S4** plus the following rule:

$$b\downarrow \frac{S\{\Diamond \Box R\}}{S\{R\}}$$

We can show that the theories of the systems **S5**, the three possible formulations of **G-S5** and **SKSg-S5** are equivalent, and that cut-free proofs of **G-S5** map onto **KSg-S5**, but unfortunately cut-elimination for all three variants of **G-S5** fails:

PROPOSITION 10

1. *The axiomatisations of* ***G-S5*** *formulated using rules* $\Box_3$ *and* $\Box_5$ *have no up-fragment free proof of* $\neg A \vee \Box \Diamond A$*, a theorem of* ***S5****.*

[6] The first of these three alternatives, using $\Box_3$ is the one-sided analogue of the two-sided sequent formulation of **S5** due to Ohnishi and Matsumoto, [17].

2. *The axiomatisation of **G-S5** formulated using rule $\Box_4$ has no cut-free proof of $\neg\Box\Box A \vee \diamond A$, a theorem of **S5**.*

PROOF The are cut-bearing proofs for part 1 with cut formula $\Box A$, and for part 2 with cut formula $\Box \diamond A$, for which in each case all the side formulae have $\Box$ as their main operator, which in each case is the principal conclusion of the appropriate $\Box_i$ rule. □

While cut-free sequent systems for **S5** have appeared in the literature; for example [4], these formulations do not take the simple form of the sequent systems treated in this paper, possessing either sophisticated rules that expose proof structure (either explicitly as in Braüner's connections between formula, or tacitly by some form of extra-logical labelling), or introduce some kind of deep inference that involves judgements with more sophisticated structure than Gentzen's sequents. We have not heard of any truly cut-free sequent formulation of **B** in a Gentzen-style sequent calculus.

We can axiomatise system **B** in the calculus of structure by adding $b\downarrow$ to **KSg-M** to obtain **KSg-B**, and likewise **SKSg-B**. Here our inference rules dealing with the modal operators find themselves in a one-one relationship to four of the five axioms given in the lattice of Hilbert systems described in Definition 1; we can extend this to rule **5** with the inference rule:

$$5\downarrow \frac{S\{\diamond\Box R\}}{S\{\Box R\}}$$

CONJECTURE 11 *For all of the systems of the cube characterised by adding any subset of the modal down rules we have described above, cut is admissible.*

Note that though there are fifteen systems of the cube, there are thirty-two possible ways of characterising them; all of these characterisations are conjectured to allow cut elimination.

In particular we make an observation.

THEOREM 12 ***KSg-S5** is cut-free.*

We will discuss this result in section five, which provides the strongest of several pices of evidence for the conjecture, when we treat hypersequents.

The approach to proofs we have so far outlined are proofs by translation which leverage known cut elimination proofs for the sequent calculus. It is also possible to prove cut elimination directly, although non-constructively, by semantic means. The advantage of this method is that it is as systematic as the formalisation in the calculus of structures; the disadvantages are, besides non-constructivity, that such methods give only limited proof-theoretic insight, and so far are restricted to fewer axioms than are needed to describe the systems of the cube.

Let us introduce two new Hilbert-Lewis axioms:

DEFINITION 13

1. *Axiom **W5**:* $\diamond \Box p \supset \Box \Box p$

2. *Axiom **C4**:* $\Box \Box p \supset \Box p$

Then we have $\mathbf{W5}, \mathbf{C4} \vdash \mathbf{5}$, $\mathbf{5} \vdash \mathbf{W5}$, and $\mathbf{T} \vdash \mathbf{C4}$; these new axioms then are weak versions of **5** and **T** that combine to provide the inferential strength of axiom **5**. Each of these can be incorporated into the calculus of structures as follows:

$$w5\downarrow \frac{S\{\diamond \Box R\}}{S\{\Box \Box R\}} \qquad c4\downarrow \frac{S\{\Box \Box R\}}{S\{\Box R\}}$$

CONJECTURE 14 *For all 128 possibilities of adding some subset of the seven rules $d\downarrow$, $t\downarrow$, $4\downarrow$, $b\downarrow$, $c4\downarrow$, $5\downarrow$ and $w5\downarrow$ to **KSg** the up-fragment is admissible.*

Obviously this conjecture is stronger than the first, and so far the evidence for it is weaker. At least a subsystem generated by four of the above seven rules is known to be cut-free by a semantic argument. It appears to be the case that the semantic argument works for systems described by axioms that correspond to *rooted conditions* on the frame accessibility relation; that is, conditions that can be described by a Π^0_2 formula where the universal quantifiers describe a tree of arcs across nodes, and the existential quantifiers assert the existence of arcs each of which either originates from the root of the tree, or which originates from a node that is existentially asserted and distinct from any in the original tree. The rules that correspond to rooted conditions we call rooted, these that don't we call unrooted.

THEOREM 15 *For each of the 10 systems described by some subset of the four unrooted rules* $d\downarrow$, $t\downarrow$, $4\downarrow$, *and* $c4\downarrow$, *the up-fragment is admissible.*

To repeat, this theorem is proven by semantic means that we believe extend to all systems described by rooted rules. Additionally the six systems that can be formalised without the $c4\downarrow$ rule can all be proven by means of translation from known systems in the sequent calculus (four of which are the systems discussed in most depth at the beginning of this section).

5 Display logic and hypersequents

As remarked before, the most obvious rival to our calculus is the application of display logic to modal logic as investigated by Wansing [27]. Modal display logic is modular in the sense that we seek, and it has a syntactic cut elimination that ensures any properly presented system (that is, a 'properly displayable' system) is cut-free, in the true sense of guaranteeing analytic normal forms. It embraces classes of calculi that we do not know how to express satisfactorily in the calculus of structures, namely calculi without De Morgan duality, such as intuitionistic logic[7]. And indeed these strengths come from a property display logic possesses which is similar to a property of the calculus of structures, namely, the manipulations on structures allow the logical rules to be applied at effectively unbounded depths, or in other words, display logic is also a calculus of deep inference.

Given all this, one might reasonably ask – why pursue another calculus of deep inference? Our answer is that we believe that for the purposes of proof analysis, the design of the calculus of structures will ultimately lead to technically better results, essentially for the following reason: display logic wishes to combine the sophisticated notion of proof structure one obtains with deep inference with the traditional

[7]Though there are logics expressible in the calculus of structures which it appears that display logic cannot express at all, such as system NEL [13], due to the branching nature of proofs in display logic

approach to proof analysis based on the subformula property. The result is that display logic leads a double life, with the subformula property holding on formulae, the leaves of the structures display logic manipulates, but not able to say anything about the fully fledged structures which is really where all the action happens. A second consequence is that display logic seeks to embrace constraints on its presentations, such as Došen's principle, that appear to be necessary to get reasonable results from subformla-property-based proof analysis. Even so, the proof analysis does not seem to be very fruitful, when judged against the paradigmatic stadard of Gentzen's formalisations: in the sequent calculus one can simply read of a tableu mechanism from a cut-free sequent calculus; in display logic, the procedure, or perhaps better heuristic, is more fraught. To be blunt, some of the technical advantages that flow from the subformula property in the context of the sequent calculus do not flow from the property that display logicians call 'subformula property'; perhaps, by analogy, the meaning-theoretic parallels depend upon a certain amount of charitable optimistism.

By contrast the calculus of structures makes no distinction between logical and structural rules, which brings simplicity; constraints such as Došen's principle, or some of the properties of systems described by Wansing like separation, symmetry and explicitness simply make no sense in this context; and seeks to substitute a wholly novel methodology of proof analysis, with some striking properties that improve on Gentzen-family calculi, such as atomicity.

Additionally there is a technical disadvantage of display logic for our investigation into systems of modal logic: display logic appears to be most naturally a tense logic, in that in introducing a modality, one obtains not only introduction rules for that modality and its De Morgan dual, but also the reverse modalities in the tense logical sense (so four modal operators, rather than the expected two). Hence, whereas when one has a proof in **KS-S5** each structure occuring in the inference corresponds to a theorem of **S5**[8], in display logic the axiom dealing for 5 is expressed in terms of tense logic (due to the conversion of modal axioms to rules involving primitive tense formula,

[8]By means of the translation $-^h$ described in section three.

following the results of Marcus Kracht [15]), and so the judgements appearing in the tree of a proof may be assertions not of **S5** but of **S5t**, its tensed extension. By conservativity, we know that we have the right theorems, but conservativity of theoremhood does not map into conservativity of cut-free provability; there is still a problem unsolved. In particular, if we consider these kind of results adequate, then the problem of cut elimination for modal logic becomes mostly a solved problem: we simply use calculi that embed modal logics in sequent calculi for geometric theories, such as pioneered by Alex Simpson[20].

Certainly there is room for dispute over the relative merits of the two approaches to providing a structural proof theory for modal logic; however we think it is important to recognise firstly that the ultimate test of health of a proof calculus will lie in its usefulness to logicians not interested in proof theory for proof theory's sake; secondly that in providing a toolkit for proof analysis display logic and the calculus of structures are not necessarily rivals, but may instead be complementary formalisms.

We also wish to discuss hypersequents as they have been treated by Arnon Avron. These extend the sequent calculus, allowing several strands of inference to accur in parallel, and to communicate by means of certain 'external' structural rules (the familiar structural rules are called 'internal', we will call rules that act locally upon a sequent internal also). We note that for the modal logics in which we are interested, the 2-sided hypersequents Avron treats are equivalent to 1-sided hypersequents; thus we may describe the general hypersequent $\vdash \Gamma_1 | \ldots | \vdash \Gamma_n$ by the structure $[\Box[\Gamma_1^s], \ldots, \Box[\Gamma_1^s]]$. Avron captures the system **S5** in the hypersequent system **HS5** by giving internal rules corresponding to the rules for **S4**, together with, in addition to external contraction and weakening, a modalised splitting rule (we give our 1-sided variant):

$$\mathbf{MS}\,\frac{G| \vdash \Box\Gamma_1, \diamond\Gamma_2, \Gamma_3}{G| \vdash \Box\Gamma_1, \diamond\Gamma_2 | \vdash \Gamma_3}$$

All internal rules map easily onto the rules for **S4**, just in a deeper context of the form $[\Box\{-\}, R]$, and the external weakening and contraction rules are straightforwardly modelled by $w\downarrow$ and $c\downarrow$; all that remains is to capture the modalised splitting rule, which can be modelled by the $4\downarrow$, $5\downarrow$ and s rules. By these means, cut free proofs of system **HS5** are mapped to cut-free proofs of **KSg-S5**, thus establishing the theorem described in section four.

Let us note, though, that the modalised splitting rule does not fix all of the problems arising due to lack of expressiveness of the cut-free sequent calculus. In particular, the modalised splitting rule gives the inferential strength of the 5 axiom only by also giving the inferential strength of the 4 axiom as well. To be more precise, hypersequent systems with the modalised splitting rule cannot be used to axiomatise systems **B**, **KB**, **DB**, **K5** or **D5**, and it has not been shown how the inferential strength of unrooted axioms can be achieved in a hypersequent calculus other than by means of modalised splitting. Thus, while hypersequents achieve an important result, one that we depend upon for the results of this paper, the means by which it achieves the goals appears to be fundamentally limited.

6 Review

The most important achievement of this paper bas been to provide a novel, modular approach to the proof theory of modal logic that begins to allow the kind of powerful proof analysis described by Brünnler for several important modal logics (those for which we can prove up-fragment admissibility), and for which we have reason to hope can be extended to all the modal logic systems we have described. Furthermore we hope that our readers will find the system to be elegant and provocative.

The most serious defect of the account given here from the point of view of formal aesthetics is that our proofs are external: our methods for showing cut-elimination are so far a rather underpowered semantical technique or by means of translations from a quite separate theory, so though we could pedantically claim to have provided a syntactic proof of cut-elimination, the proof is quite as external as with semantical proofs of cut-elimination. Furthermore, we are mostly left without resources to deal with the cases where we do not have appropriate cut-free axiomatisations, such as for system **B**. Nonetheless, it must be emphasised that these are defects of our knowledge of how to prove results; to the best of this limited knowledge, the family of proof systems are in excellent shape: no rival calculus appears to describe as many systems so simply.

There are three principal avenues to explore in search of the missing proof techniques. The first is to try to extend the semantic proofs; Avron's paper on hypersequents succeeds in providing a semantical proof of cut elimination for **HS5** and it is possible the technique can be adpated and generalised. Second, Guglielmi [14] introduces a technique, called splitting, that may be regarded as the preferred way to give internal, syntactic proofs of cut-elimination in the calculus of structures, and in joint work with one of the authors (in preparation) has shown how this technique may be extended to deal with the case of SKS. Lastly, a very syntactically involved technique of permutability of rules can be applied to prove cut elimination; two proofs of this nature are due to Strassburger [23], as well as a proof by Brünnler [6]. Of these approaches, a splitting proof would be the most valuable.

A further issue relates to the role of proof analysis in the toolkit of the modal logician. The single most important application of proof analysis in propositional logic has been its crucial role in influencing the design of tableau methods. With modal logics, the ad hoc nature of sequent characterisations has been reflected by an equally ad hoc methodology for the design of modal tableau. Furthermore the designers of modal tableau have found it necessary to resort to ugly techniques equivalent to 'analytic' cut (a proof with an analytic cut is by virtue of this cut not analytic; this is perhaps the most misleading piece of nomenclature in proof theory), so indeed the degree of disorder among modal tableau is greater than among modal sequent calculi. The most obvious and useful test of a claim to have provided a worthy modal proof theory is leverage the theory to providing a principled approach to modal tableau. The application of the calculus of structures to the design of such systems is under intensive investigation, making use of insights flowing from Dale Miller's 'proof search as computation' slogan [7, 16].

BIBLIOGRAPHY

[1] A. Avron The method of hypersequents in the proof theory of propositional non-classical logics. In W. Hodges et al. (eds.), *Logic: From Foundations to Applications*,

pages 1–32. Oxford University Press, 1996.

[2] N. Belnap. Display Logic. *Journal of Philosophical Logic*, 11:375–417, 1982.

[3] P. Blackburn, M. de Rijke, Y. Venema. *Modal Logic.* Cambridge University Press, 2001.

[4] T. Braüner. A cut-free Gentzen formulation of the modal logic S5. In the *Logic Journal of the Interest Group in Pure and Applied Logics*, volume 8(5), pages 629–643, 2000.

[5] K. Brünnler. Locality for classical logic. Submitted to *Archive for Mathematical Logic*, 2003. Preprint available `http://www.wv.inf.tu-dresden.de/ kai/LocalityClassical.pdf`.

[6] K. Brünnler. Deep Inference and Symmetry in Classical Logic. Logos Verlag, Berlin, 2004.

[7] P. Bruscoli. A Purely Logical Account of Sequentiality in Proof Search. In *Proc. ICLP 2002*, LNCS 2401:302–316. Springer-Verlag, 2002.

[8] R. A. Bull and K. Segerberg. Basic Modal Logic. In *The Handbook of Philosophical Logic*, volume 2, pages 1–88. Kluwer, 1984.

[9] L. Cardelli and A. D. Gordon. Anytime, anywhere: Modal logics for mobile ambients. In *Proc. Principles of Programming Languages 2000*, pages 365–377. ACM Press, 2000.

[10] K. Došen. Sequent-Systems for Modal Logic. In *Journal of Symbolic Logic*, volume 50(1), pages 149–168, 1985.

[11] T. Forster. The modal aether. Manuscript.

[12] J. Garson. Modal Logic. In E. N. Zalta, editor, *The Stanford Encyclopaedia of Philosophy*, Winter 2001 edition. Available `http://plato.stanford.edu/archives/win2001/entries/logic-modal`.

[13] A. Guglielmi and L. Straßburger. Non-commutativity and MELL in the Calculus of Structures. In *Proc. Computer Science Logic 2001*, LNCS 2142:54–68. Springer-Verlag, 2001.

[14] A. Guglielmi. A System of Interaction and Structure. Submitted to *ACM Transactions on Computational Logic*, 2002. Preprint available `http://www.ki.inf.tu-dresden.de/~ guglielmi/Research/Gug/Gug.pdf`.

[15] M. Kracht Power and weakness of the modal display calculus. In, H. Wansing (ed.), *Proof Theory of Modal Logic*, pages 93–121. Kluwer, Dordrecht, 1996.

[16] R. Milner. *Communicating and mobile systems: the pi-calculus.* Cambridge University Press, 1999.

[17] M. Ohnishi and K. Matsumoto. Gentzen method in modal calculi, parts I and II. In *Osaka Mathematical Journal*, volume 9, pages 113–130, 1957, and volume 11, pages 115–120, 1959.

[18] K. Schütte. *Proof Theory.* North-Holland, 1962.

[19] H. Schwichtenberg and A. Troelstra. *Basic Proof Theory.* Cambridge University Press, 1998.

[20] A. Simpson *Proof theory of intuitionistic modal logic.* PhD thesis, University of Edinburgh, 1994.

[21] C. A. Stewart. Which 2-sided sequent systems are equivalent to 1-sided systems? Unpublished manuscript, 2004.

[22] P. Stouppa. The design of modal proof theories. MSc dissertation, Technische Universität Dresden. In preparation.

[23] L. Straßburger. A local system for linear logic. In *LPAR 02*, 2002.

[24] S. Valenti. Cut-elimination in a modal sequent calculus for K. In *Bolletino dell'Unione Mathematica Italiana*, volume 1B, pages 119–130, 1982.

[25] S. Valenti. The sequent calculus for the modal logic D. In *Bolletino dell'Unione Mathematica Italiana*, volume 7A, pages 455–460, 1993.

[26] H. Wansing. Sequent Calculi for Normal Modal Propositional Logics. In *Journal of Logic and Computation*, volume 4(2), pages 125–142, 1994.

[27] H. Wansing. *Displaying Modal Logic.* Kluwer, Dordrecht, 1998.

Charles Stewart
International Centre for Computational Logic, Technische Universität Dresden.
cas@linearity.org

Phiniki Stouppa
Institut für Informatik und angewandte Mathematik, University of Bern.
stouppa@iam.unibe.ch