# A form of feasible interpolation for constant depth Frege systems

Jan Krajíček*

Charles University and Academy of Sciences, Prague

**Abstract**

Let $L$ be a first-order language and $\Phi$ and $\Psi$ two $\Sigma^1_1$ $L$-sentences that cannot be satisfied simultaneously in any finite $L$-structure. Then obviously the following principle $Chain_{L,\Phi,\Psi}(n,m)$ holds: For any chain of finite $L$-structures $C_1,\ldots,C_m$ with the universe $[n]$ one of the following conditions must fail:

1. $C_1 \models \Phi$,
2. $C_i \cong C_{i+1}$, for $i = 1,\ldots,m-1$,
3. $C_m \models \Psi$.

For each fixed $L$ and parameters $n,m$ the principle $Chain_{L,\Phi,\Psi}(n,m)$ can be encoded into a propositional DNF formula of size polynomial in $n,m$.

For any language $L$ containing only constants and unary predicates we show that there is a constant $c_L$ such that the following holds: If a constant depth Frege system in DeMorgan language proves $Chain_{L,\Phi,\Psi}(n, c_L \cdot n)$ by a size $s$ proof then the class of finite $L$-structures with universe $[n]$ satisfying $\Phi$ can be separated from the class of those $L$-structures on $[n]$ satisfying $\Psi$ by a depth 3 formula of size $2^{\log(s)^{O(1)}}$ and with bottom fan-in $\log(s)^{O(1)}$.

A proof system P admits feasible interpolation if from a P-proof $\pi$ of

$$\neg\alpha_n(\overline{x},\overline{y}) \vee \neg\beta_n(\overline{x},\overline{z}) \ ,$$

$\overline{x} = (x_1,\ldots,x_n)$ and $\overline{x},\overline{y},\overline{z}$ disjoint tuples of variables, one can infer some algorithmic information about separating two sets

$$U_n := \{\overline{a} \in \{0,1\}^n \mid \exists \overline{y}\alpha(\overline{a},\overline{y})\}$$

and

$$V_n := \{\overline{a} \in \{0,1\}^n \mid \exists \overline{z}\beta(\overline{a},\overline{z})\} \ .$$

*Supported in part by grants IAA100190902, AV0Z10190503, MSM0021620839, LC505 (Eduard Čech Center) and by a grant from the John Templeton Foundation.

Most often this means that there is an algorithm, using $\pi$ as an advice string and running in time polynomial in the length of $\pi$, that separates the two sets. But sometimes a different type of information is deduced (a monotone circuit, a winning strategy for a two-player game, a span program, a linear program, etc.).

Feasible interpolation was proposed ([10]) and developed ([23, 3, 12]) primarily as a method for proving lengths-of-proofs lower bounds: Any two disjoint $\mathcal{NP}$-sets that it is not possible to separate by an algorithm specified in the particular interpolation theorem for P give rise to formulas $\alpha_n$ and $\beta_n$ such that any P-proof of the disjunction $\neg\alpha_n \vee \neg\beta_n$ must be long for $n >> 0$. And as a lower bound method feasible interpolation is very successful, applying to the widest range of various proof systems among all lower bound methods: from logical systems like resolution R ([12]), to geometric proof systems like cutting planes CP or LK(CP) ([20, 13]), to algebraic systems like the Nullstellensatz system NS or Polynomial calculus PC ([22]), to the OBDD proof system ([14]), and others. In fact, even when the method fails for a proof system P, meaning that no feasible separating algorithm can be deduced from the existence of a short proof, it still provides a valuable information. Namely such a failure implies that the proof search for P cannot be done feasibly either (the so called non-automatizability of P, cf. [4]).

To show that a proof system P does not admit feasible interpolation one looks for two disjoint $\mathcal{NP}$-sets $U$ and $V$ that it is hard to separate (e.g. there is no polynomial-time separating algorithm) but for which the formulas expressing the disjointness of $U_n := U \cap \{0,1\}^n$ with $V_n := V \cap \{0,1\}^n$ have short P-proofs. No disjoint $\mathcal{NP}$-pair $U, V$ is proved to be hard to separate in this sense at present (it would imply, in particular, that $\mathcal{P} \neq \mathcal{NP}$) but there are several pairs conjectured to have this property. These are derived from various encryption schemes and the hardness of their separation follows from the conjectured security of the schemes.

This argument towards the impossibility of feasible interpolation was applied first to Extended Frege system EF in [15] (using RSA encryption scheme) and was then modified for Frege system F ([4]), and eventually even for constant depth Frege systems $F_d$ in [2] (they used Diffie-Hellman scheme). There are differences, however. While the results on EF and F show the (conditional) impossibility of even sub-exponential interpolating circuits, the result on $F_d$ yields only quasi-polynomial lower bound on such circuits.

In this paper we formulate a form of feasible interpolation for constant-depth Frege systems. The interpolating circuits will have quasi-polynomial size but also constant depth. It is easy to prove (unconditionally, using the parity lower bound of [9, 25]) that $F_d$ does not admit sub-exponential size constant depth interpolating circuits.

As strong lower bounds for systems $F_d$ are already known (and, in fact, we utilize the hardness of $\mathsf{PHP}$ for $F_d$'s in our construction) our aim is mainly to make a first step towards resurrecting feasible interpolation in some form for strong proof systems.

This is a paper in proof complexity and we shall assume that the reader is acquainted with the most basic and well-established notions of the area but we will not assume any prior knowledge of feasible interpolation results. The basic notions whose knowledge is assumed include constant depth Frege systems in DeMorgan language, bounded arithmetic theory $V^0_1$ of [5] (denoted $V^0$ in [7]) and the Paris-Wilkie [17] translation of bounded formulas to a sequence of constant-depth propositional formulas. We will also use the hardness of the pigeon-hole principle PHP for systems $F_d$ but we will recall explicitly the specific result (in the language of model-theory) used. Relevant background can be found in [11, 7], the definition of proof systems $F_d$ and of PHP also in [8, 21].

Notation: $[n]$ stands for $\{1, \ldots, n\}$.

# 1 The isomorphism-chain principle

In this section we consider first-order and propositional formalizations of the principle we will study. Let $L$ be a first-order language. We will consider only finite relational languages that may contain also constants but not functions symbols of non-zero arity. This is chiefly because functions are represented for the propositional translation by their graphs anyway.

Let $\Phi = \exists \overline{X} \phi(X)$ and $\Psi = \exists \overline{Y} \psi(Y)$ be two $\Sigma^1_1$ $L$-sentences that cannot be satisfied simultaneously in any finite $L$-structure. Then obviously the following *chain principle* holds for any $n, m \in \mathbf{N}$:

> *For any chain of finite $L$-structures $C_1, \ldots, C_m$ with the universe $[n]$ one of the following conditions must fail:*
>
> 1. $C_1 \models \Phi$,
> 2. $C_i \cong C_{i+1}$, *for* $i = 1, \ldots, m-1$,
> 3. $C_m \models \Psi$.

For any fixed $L$ and parameters $x, y$ the chain principle can be formulated by a bounded first-order formula $Chain_{L,\Phi,\Psi}(x, y)$ as follows. The formula uses

- names for all second-order witnesses $\overline{S}$ and $\overline{T}$ of the both $\Sigma^1_1$-sentences $\Phi$ and $\Psi$ respectively,
- indexed names $c_i$ and $R_i$ for all constant symbols $c$ and relation symbols $R$ in $L$ (intended to describe the $i$-th $L$-structure $C_i$) for $i \leq y$, and
- indexed names for binary relations $H_i$ (intended to be the graphs of the isomorphism between $C_i$ and $C_{i+1}$) for $i < y$.

and says that

- either the structure $C_1$ with the universe $[x]$ and language $L$ interpreted by $c_1, \ldots, R_1, \ldots$ does not satisfy $\phi(\overline{S})$,

- or the structure $C_y$ with the universe $[x]$ and language $L$ interpreted by $c_y, \ldots, R_y, \ldots$ does not satisfy $\psi(\overline{T})$,
- or there is $1 \leq i < y$ such that $H_i$ is not a graph of an isomorphism from $C_i$ onto $C_{i+1}$.

Incorporating suitable Skolem functions among the second-order witnesses $\overline{X}$ and $\overline{Y}$ we may assume without a loss of generality that both formulas $\phi$ and $\psi$ are universal formulas with the open kernels in a CNF.

For any fixed $n, m$ the instance $Chain_{L,\Phi,\Psi}(n, m)$ can be encoded into a propositional DNF formula of size polynomial in $n, m$. This is the standard Paris-Wilkie translation (see [11, Chpt.9] or [7]). The propositional formula, denoted $\langle Chain_{L,\Phi,\Psi}\rangle_{n,m}$, is built from atoms corresponding to atomic sentences in the language of formula $Chain_{L,\Phi,\Psi}(n, m)$ with parameters from $[n]$ or $[m]$.

# 2 The idea of chain feasible interpolation

We shall now describe the idea of the form of feasible interpolation we will study. We will call it chain feasible interpolation or simply *chain interpolation*.

Assume you want to prove the chain principle $Chain_{L,\Phi,\Psi}(x, y)$, with $L$, $\Phi$ and $\Psi$ fixed. One way how to prove it is to show, by induction on $i \leq y$, that $C_i \models \Phi$. Case $i = 1$ is condition (1) of $\neg Chain_{L,\Phi,\Psi}(x, y)$. Condition (2) then implies the induction step

$$C_i \models \Phi \rightarrow C_{i+1} \models \Phi$$

as an isomorphism between $C_i$ and $C_{i+1}$ maps a witness to $\Phi$ in $C_i$ to a witness in $C_{i+1}$. Hence if we had induction for the formula $C_i \models \Phi$ (with the induction parameter $i$) then $C_y \models \Phi$ would follow, and we could conclude the proof by brining this to a contradiction with condition (3). However, the formula

$$C_i \models \Phi$$

is a $\Sigma^1_1$-property of $i$, definable by a bounded $\Sigma^{1,b}_1$-formula. The argument just outlined can be thus run through in a theory having induction for $\Sigma^{1,b}_1$-formulas; this is theory $V^1_1$ of [5] (denoted $V^1$ in [7]), see also [11]. This theory corresponds to Extended Frege systems EF and hence the chain principle has polynomial size EF-proofs as long as the incompatibility of $\Phi$ and $\Psi$:

$$A \models \neg\Phi \vee \neg\Psi$$

($A$ is a free second order object ranging over $L$-structures on $[x]$) is provable in $V^1_1$. In particular, the example from [15] that shows (conditioning upon the security of RSA) that EF does not admit (ordinary) feasible interpolation also shows that it admits chain interpolation neither.

Another possibility how to prove the chain principle is to explicitly define, in terms of a witness $\overline{S}$ for $\Phi$ in $C_1$ and isomorphisms $H_1, \ldots, H_{i-1}$, a witness for

$\Phi$ in $C_i$. However, the depth of formulas defining such a witness grows linearly with $i$ and hence this can be performed again in EF but presumably not even in Frege systems F. In fact, using the standard witnessing argument these two proof strategies are essentially equivalent.

There is yet another way how to prove the chain principle that comes to a mind, and it is this one that is potentially formalizable in Frege or in constant depth Frege systems. Assume we can find a first order $L$-sentence $\gamma$ such that for all $L$-structures $A$:

$$A \models \Phi \quad \Rightarrow \quad A \models \gamma$$

and also

$$A \models \Psi \quad \Rightarrow \quad A \models \neg\gamma \ .$$

That is, the property of structures $A$ to satisfy $\gamma$ separates the class of structures satisfying $\Phi$ from the class of those satisfying $\Psi$.

It is straightforward to see that theory $V^0_1$ proves for any first-order $L$-sentence $\gamma$ that $C_i \cong C_{i+1}$ implies

$$C_i \models \gamma \ \Rightarrow \ C_{i+1} \models \gamma$$

Hence as long as we can prove in $V^0_1$ instances of the interpolation implications above:

$$C_1 \models \Phi \ \Rightarrow \ C_1 \models \gamma$$

and

$$C_y \models \Psi \ \Rightarrow \ C_y \models \neg\gamma \ ,$$

we can refute the chain principle in the theory.

For any fixed first-order $L$-sentence $\gamma$ the property of structures $A$ with the universe $[n]$ that $A \models \gamma$ is expressible as a symmetric $AC^0$-property of $A$. Hence a little bit more generally we could take an $AC^0$-property separating the class of structures satisfying $\Phi$ from the class of those satisfying $\Psi$ in place of $\gamma$, and run the argument through in a constant depth Frege system.

The (somewhat naive) idea behind the chain feasible interpolation is that this ought to be essentially the only way how constant depth Frege systems can prove the chain principle shortly. We now formulate this working conjecture as an open problem in a form weaker than the informal account above. In particular, we do not ask for the provability of the interpolating implications but only for their validity, and we also do not require the symmetry of the interpolating property.

**Open problem 2.1** *Let $L$ be a first-order language containing relation symbols or constants. Let $\Phi$ and $\Psi$ by two $\Sigma^1_1$ $L$-sentences that cannot be simultaneously satisfied in any finite $L$-structure.*

*Is it true that for every $d \geq 1$ there is $d' \geq 1$ such that the following holds for all sufficiently large $m \geq n \geq 1$:*

- *If there is a size $s$ $F_d$-proof of $\langle Chain_{L,\Phi,\Psi}\rangle_{n,m}$ then there is a property of $L$-structures with the universe $[n]$ definable by a depth $d'$ circuit of size $2^{s^{o(1)}}$ that separates the class of $L$-structures on $[n]$ satisfying $\Phi$ from those satisfying $\Psi$?*

# 3 The unary case

In this section we answer affirmatively Problem 2.1 for the case of languages $L$ containing constants and unary predicate symbols but not relational symbols of arity bigger than 1. In this case even $d' = 3$ irrespective of what $d$ is and the size bound is even quasi-polynomial $2^{\log(s)^{O(1)}}$.

The unary case already allows one to define interesting pairs of $\Sigma^1_1$ sentences as the next couple of examples demonstrates.

**Examples:**

**(1)** Let $L$ consists of one unary predicate $W$. Formula $\Phi$ says that the cardinality of $W$ is even by stating the existence of a complete pairing on $W$ while $\Psi$ says similarly (there exists a pairing on $W$ leaving exactly one point out) that it is odd. The chain principle is then related to the parity principle, and the exponential lower bound for parity principle in $F_d$ (implied by [16, 19]) implies that the chain principle has no short $F_d$ proofs even for chains of length 1.

**(2)** For the same $L$ as in (1) one can $\Sigma^1_1$-define that the cardinality of $W$ is at east $n/2$ (by saying that there is a 2-to-1 map from the universe into $W$) and at most $n/4$ (by saying analogously that there is a 1-to-4 map from $W$ into the universe). The chain principle then relates to the weak pigeonhole principle which is known to have quasi-polynomial size proofs in $F_d$ for $d$ sufficiently large ([18, 11]). Hence the chain principle has also quasi-polynomial size $F_d$ proofs for chains of length 1 but, as will follow from our theorem, not for general chains.

**Theorem 3.1** *Let $L$ be a first-order language containing unary relation symbols or constants. Let $\Phi$ and $\Psi$ by two $\Sigma^1_1$ $L$-sentences that cannot be simultaneously satisfied in any finite $L$-structure.*

*Then there is constant $c_L \geq 1$ such that for every $d \geq 1$ there is $c \geq 1$ such that the following holds for all $n$ large enough:*

- *If there is a size $s_n$ $F_d$-proof of $\langle Chain_{L,\Phi,\Psi}\rangle_{n,c_L\cdot n}$ then there is a property of $L$-structures with the universe $[n]$ definable by a depth 3 formula of size $2^{\log(s_n)^c}$ and bottom fan-in $\log(s_n)^c$ that separates the class of $L$-structures on $[n]$ satisfying $\Phi$ from the class of those satisfying $\Psi$.*

**Proof :**

A constant $e$ in language $L$ can be represented by a unary predicate representing $x = e$ and so without a loss of generality we may assume that $L$ contains no constants but only $\ell$ unary predicates $P_1, \ldots, P_\ell$. Put $c_L := 2^\ell$.

Fix $d \geq 1$. Assume for the sake of a contradiction that for all $c \geq 1$ there are arbitrarily large $k$ such that $\langle Chain_{L,\Phi,\Psi}\rangle_{k,c_L\cdot k}$ has a size $s_k$ $F_d$-proof but no depth 3 formula of size at most $2^{\log(s_k)^c}$ and bottom fan-in $\log(s_k)^c$ separates the class of $L$-structures on $[k]$ satisfying $\Phi$ from the class of those satisfying $\Psi$.

By compactness there is a non-standard model $\mathbf{N}^*$ of true arithmetic and a non-standard element $n \in \mathbf{N}^* \setminus \mathbf{N}$ such that, in $\mathbf{N}^*$, there is a size $s_n$ $F_d$-proof $\Pi$ of $\langle Chain_{L,\Phi,\Psi}\rangle_{n,c_L\cdot n}$ but for all standard $c \in \mathbf{N}$ no depth 3 formula of size

at most $2^{\log(s_n)^c}$ and bottom fan-in $\log(s_n)^c$ separates the class of structures on $[n]$ satisfying $\Phi$ from the class of those satisfying $\Psi$.

By a 1-type $\alpha(x)$ we shall mean any conjunction of the form

$$\eta_1 \wedge \ldots \wedge \eta_\ell$$

where $\eta_i$ is either $P_i(x)$ or $\neg P_i(x)$. 1-types are in an obvious 1-to-1 correspondence with elements of $\{0,1\}^\ell$.

For a 1-type $\alpha(x)$ and an $L$-structure $A$ let $r_\alpha(A)$ be the number of elements of the $A$ satisfying the type $\alpha(x)$. Clearly the sum $\sum_\alpha r_\alpha(A)$, $\alpha \in \{0,1\}^\ell$, equals to the cardinality of the universe of $A$, and two finite structures $A$ and $B$ with universes of the same cardinality are isomorphic iff $r_\alpha(A) = r_\alpha(B)$ for all 1-types $\alpha$.

For $t \geq 0$ and a 1-type $\alpha$, $\kappa_{\alpha,t}$ is an $L$-sentence formalizing "$r_\alpha(A) \geq t$", meaning that for all structures $A$, $A \models \kappa_{\alpha,t}$ iff $r_\alpha(A) \geq t$. Sentence $\kappa_{\alpha,t}$ is defined as follows:

$$\exists x_1, \ldots, x_t \, [\bigwedge_{i \neq j} x_i \neq x_j \ \wedge \ \bigwedge_i \alpha(x_i)]$$

where $1 \leq i, j \leq t$. The propositional translation $\langle \kappa_{\alpha,t} \rangle_n$ of the sentence is of the form

$$\bigvee_{a_1,\ldots,a_t} [\bigwedge_i \alpha(a_i)]$$

where $a_1, \ldots, a_t$ range over subsets of $[n]$ of $t$ different elements. The conjunction $\bigwedge_i \alpha(a_i)$ has $\ell \cdot t$ literals (as each $\alpha(a_i)$ is an $\ell$-term). Hence $\langle \kappa_{\alpha,t} \rangle_n$ is a DNF formula with top fan-in $\binom{n}{t} \leq n^t$, bottom fan-in $O(t)$ and total size bounded above by $O(tn^t)$.

For $c \geq 1$ define the $c$-spectrum $\Delta_c(A)$ of a structure $A$ with universe $[n]$ to be the unique conjunction

$$\bigwedge_{\alpha \in \{0,1\}^\ell} \Gamma_\alpha$$

valid in $A$ where each $\Gamma_\alpha$ is

- either $\langle \kappa_{\alpha,t} \rangle_n \wedge \neg \langle \kappa_{\alpha,t+1} \rangle_n$, for some $t < \log(s_n)^c$,
- or $\langle \kappa_{\alpha,\log(s_n)^c} \rangle_n$.

In other words, the $c$-spectrum counts exactly the number of elements of each type as long as it is less than $\log(s_n)^c$ or says that it is at least $\log(s_n)^c$.

Note that each $c$-spectrum is a constant size conjunction of DNFs and CNFs, each with top fan-in $\leq n^t$ and bottom fan-in $O(t)$. Note also that the number of different $c$-spectra is bounded above by

$$(\log(s_n)^c)^{2^\ell} = \log(s_n)^{O(c)} \ .$$

**Claim 1:** *In $\mathbf{N}^*$ there is a non-standard $c^*$ and two $L$-structures $A$ and $B$ with universe $[n]$ such that:*

1. $A \models \Phi$.

2. $B \models \Psi$.

3. *Structures $A$ and $B$ have the same $c^*$-spectrum:* $\Delta_{c^*}(A) = \Delta_{c^*}(B)$.

To prove the claim assume for the sake of contradiction that the $c^*$-spectra of any pair of structures $A$, $B$ with universe $[n]$ satisfying $\Phi$ and $\Psi$ respectively are different, for any non-standard $c^*$. By overspill this must be true also for some standard $c$. But then formula

$$\bigvee_{A'} \Delta_c(A')$$

where $A'$ ranges over $L$-structures on $[n]$ satisfying $\Phi$ is a formula separating $L$-structures on $[n]$ satisfying $\Phi$ from those satisfying $\Psi$. Applying DeMorgan rules this formula can be transformed into a depth 3 formula of bottom fan-in $\log(s_n)^{O(c)}$. As the number of different $c$ spectra is $\log(s_n)^{O(c)}$, the size of the formula is $2^{\log(s_n)^{O(c)}}$.

This contradicts our assumption about $n$ and the claim follows.

Fix a non-standard element $c^*$ and two $L$-structures $A$ and $B$ with universe $[n]$ obeying the claim. Also fix any non-standard element $w < \log(s_n)^{c^*}$ but satisfying for all standard $c$:

$$\log(s_n)^c < w \ .$$

Note that for all 1-types $\alpha$ it holds that $r_\alpha(A) = r_\alpha(B)$ as long $r_\alpha(A) < w$ or $r_\alpha(B) < w$.

Let us enumerate the 1-types as $\alpha_1, \ldots, \alpha_{2^\ell}$ in such a way that for some $1 \leq i_0 \leq 2^\ell$ it holds

$$r_{\alpha_i}(A) \geq w \ \text{ iff } \ i \leq i_0 \ .$$

By the remark above then also

$$r_{\alpha_i}(B) \geq w \ \text{ iff } \ i \leq i_0$$

and

$$r_{\alpha_i}(A) \ = \ r_{\alpha_i}(B) \ , \ \text{ for } \ i > i_0 \ .$$

Let us simplify the notation and for a structure $D$ denote $r_{\alpha_i}(D)$ simply $r_i(D)$.

We are going to define a chain of $L$-structures $C_1, \ldots, C_{c_L \cdot n}$ on $[n]$. Each structure will be be uniquely determined by a $2^\ell$-tuple $r_1, \ldots, r_\ell \geq 0$ of numbers such that $\sum_i r_i = n$ as follows: Element $a \in [n]$ has type $\alpha_i$ iff

$$r_1 + \ldots + r_{i-1} < a \leq r_1 + \ldots + r_i \ .$$

We define first two auxiliary chain $D_1, \ldots, D_p$ and $E_1, \ldots, E_q$ in steps $0, \ldots, 2^\ell$. In step 0 put $D_1$ to be a permutation of $A$ so that the elements are ordered in

accordance of the ordering $\alpha_1, \ldots, \alpha_{2^\ell}$ of the 1-types, and likewise let $E_1$ be a suitable reordering of $B$.

For $i = 1, \ldots, 2^\ell$ we assume that chains $D_1, \ldots, D_u$ and $E_1, \ldots, E_v$ we defined in steps prior to step $i$ and that they satisfy:

$$r_j(D_u) = r_j(E_v) \ , \quad \text{for all} \ \ j < i$$

Consider three cases:

1. $r_i(D_u) = r_i(E_v)$: then neither chain is extended and we move to step $i+1$ if $i < 2^\ell$ or stop if $i = 2^\ell$.

2. $r_i(D_u) > r_i(E_v)$: Extend the $D$-chain by structures $D_{u+1}, \ldots, D_{u+t}$ where $t = r_i(D_u) - r_i(E_v)$, with $D_{u+j}$ being determined by the $2^\ell$-tuple

   $$r_1(D_u), \ldots, r_{i-1}(D_u), r_i(D_u) - j, r_{i+1}(D_u) + j, r_{i+2}(D_u), \ldots, r_{2^\ell}(D_u) \ .$$

   Hence we will have $r_i(D_{u+t}) = r_i(E_v)$. Note that necessarily $r_i(E_v) \geq w$.

3. $r_i(D_u) < r_i(E_v)$: Dually to case 2. extend the $E$-chain by structures $E_{v+1}, \ldots, E_{v+t}$ where $t = r_i(E_v) - r_i(D_u)$, with $E_{v+j}$ being determined by the $2^\ell$-tuple

   $$r_1(E_v), \ldots, r_{i-1}(E_v), r_i(E_v) - j, r_{i+1}(E_v) + j, r_{i+2}(E_v), \ldots, r_{2^\ell}(E_v) \ .$$

   Hence we will again have $r_i(D_u) = r_i(E_{v+t})$, and that necessarily $r_i(D_u) \geq w$.

Assume that the final chains are

$$D_1, \ldots, D_p \ \text{ and } \ E_1, \ldots, E_q \ .$$

Define the chain $C_1, \ldots, C_{c_L \cdot n}$ to be

$$A, D_1, \ldots, D_p, E_q, \ldots, E_1, B, B, \ldots, B$$

with suitably many copies of $B$ at the end. In the construction of the two chains above new structures are added in step $i$ only if the $i$-th type has at least $w$ elements in $A$ and $B$. Each element gets moved in the worst case through all 1-types. Hence we may certainly estimate that $p + q \leq 2^\ell \cdot (n - w)$ and hence $p + q + 2 < c_L \cdot n$ too and the chain is well-defined. (We could have made this more economical and construct a chain of length $n$ rather than $c_L \cdot n$ but at the expense of a much more cumbersome definition of $C_i$ and $H_i$.)

Now we shall define a model of bounded arithmetic that we will extend shortly to another one. Let $I \subsetneq_e \mathbf{N}^*$ be a cut in $\mathbf{N}^*$ defined as

$$I \ := \ \{u \in \mathbf{N}^* \mid u < 2^{\log(s_n)^k} \ \text{ for some standard } k\} \ .$$

Note that for every $u \in I$ it holds in $\mathbf{N}^*$:

$$u < 2^{w^{1/k}}, \quad \text{for all standard } k \ .$$

Define $\mathbf{M} = (I, \mathcal{X})$ to be the following two-sorted model of theory $V^0_1$: Its first-order part is the cut $I$ and its second-order part $\mathcal{X}$ are all subsets of $I$ coded in $\mathbf{N}^*$. In particular, the $F_d$-proof $\Pi$ we started with as well as the chain of structures $C_i$ are in $\mathcal{X}$. The structure $\mathbf{M}$ clearly satisfies $V^0_1$.

**Claim 2:** *There is $\mathcal{X}^* \supseteq \mathcal{X}$ such that for $\mathbf{M}^* = (I, \mathcal{X}^*)$ the following properties hold:*

1. *$\mathbf{M}^*$ is a model of $V^0_1$.*
2. *There is $f \in \mathcal{X}^*$ such that in $\mathbf{M}^*$ $f$ is a bijection from $[w+1]$ onto $[w]$. In particular, the bijective pigeonhole principle $\mathsf{PHP}_w$ fails in $\mathbf{M}^*$.*

To prove the claim recall from above that every number $u$ in $I$ satisfies

$$u < 2^{w^{1/k}}, \quad \text{for all standard } k \ .$$

We may abbreviate this as $u < 2^{w^{o(1)}}$.

Hence the length of any set in $\mathcal{X}$ is $2^{w^{o(1)}}$ as well and, in particular, any $F_{d'}$-proof in $\mathcal{X}$ (any $d'$) has size $2^{w^{o(1)}}$ too. By the exponential lower bound for constant depth Frege proofs of $\mathsf{PHP}$ in [1, 16, 19] this implies that $\mathcal{X}^*$ contains no constant depth Frege proof of the bijective $\mathsf{PHP}_w$.

By the standard model-extension results this yields the existence $\mathcal{X}^*$ with the required properties. See [11] for details. This proves the claim.

Working inside $\mathbf{M}^*$ we are going to define (by a bounded formula) binary relations $H_i(x, y)$ that will be isomorphisms between $C_i$ and $C_{i+1}$. For the first pair $C_1 = A$ and $C_2 = D_1$ let $H_1$ be the graph of an automorphism of $A$ (definable in $\mathbf{N}^*$, hence in $\mathcal{X}$) reordering the elements of $A$ as required in the construction of the chain. Analogously, for the pair $C_i = E_1$ and $C_{i+1} = B$, let $H_i$ be the graph of the inverse of a suitable automorphism reordering $B$ into $E_1$. For the pairs $B, B$ in the last segment of the chain $H_i$ is the graph of the identity.

For the remaining non-trivial case of a pair $C_i, C_{i+1}$ from the subchain $D_1, \ldots, E_1$ we proceed as follows. By the construction of the chain the two structures are given by two $2^\ell$-tuples differing only slightly. If $C_i$ is given by

$$r_1, \ldots, r_{t-1}, r_t, r_{t+1}, r_{t+2}, \ldots, r_{2^\ell}$$

then $C_{i+1}$ is given by

$$r_1, \ldots, r_{t-1}, r_t + 1, r_{t+1} - 1, r_{t+2}, \ldots, r_{2^\ell}$$

or by

$$r_1, \ldots, r_{t-1}, r_t - 1, r_{t+1} + 1, r_{t+2}, \ldots, r_{2^\ell} \ .$$

We shall assume the former. Moreover, we know that both $r_t$ and $r_{t+1} - 1$ are at least $w$.

Define $H_i$ to be the graph of mapping $h_i : [n] \to [n]$ defined as follows. For $b_i := r_1 + \ldots + r_t$ (definable from $C_i$ and $C_{i+1}$) put:

1. $h_i(a) := a$, if $a \leq b_i - w$ or $a > b_i + w + 1$,
2. $h_i(a) := b_i + 1 + f(a - b_i)$, if $b_i < a \leq b_i + w + 1$,
3. $h_i(a) := b_i - w + f^{(-1)}(a + w - b_i)$, if $b_i - w < a \leq b_i$.

We are ready to complete the proof of the theorem. Assuming that no interpolating formula with the parameters described in the theorem exists we have constructed a model $\mathbf{M}^*$ of $V^0_1$ in which the chain principle $Chain_{L,\Phi,\Psi}(n, c_L \cdot n)$ fails, as witnessed by the chain $C_1, \ldots, C_{c_L \cdot n}$ and relations $H_i$. Because theory $V^0_1$ proves the soundness of all systems $F_d$, any standard $d$, there cannot be an $F_d$-proof of $\langle Chain_{L,\Phi,\Psi}\rangle_{n,c_L \cdot n}$ in $\mathcal{X}^*$. In particular, there is no such proof in $\mathcal{X}$ either. By the definition of $\mathcal{X}$ this means that in $\mathbf{N}^*$ no $F_d$-proof of $\langle Chain_{L,\Phi,\Psi}\rangle_{n,c_L \cdot n}$ has size less than $s_n$ (as $s_n \in I$).

This proves the theorem.

**q.e.d.**

It would be interesting to avoid the use of the PHP lower bound for constant depth Frege systems in the proof and to extend the theorem to the non-unary languages.

Both these advances are necessary should we hope to extend the chain interpolation in some form to Frege systems. This is because Frege systems can count and thus can count the isomorphism invariants $r_\alpha(A)$. For the same reason F also admits polynomial size proofs of PHP, cf.[6]. But at present I am not aware of an argument ruling out, for any $L$, a chain interpolation for Frege systems with the interpolating property being definable by a subexponential circuit, dropping the condition of constant depth. In fact, this can be proved for unary languages but they do not define the interesting disjoint $\mathcal{NP}$-pairs discussed in the introduction.

For extending the theorem to the non-unary case observe that $d'$ from Problem 2.1 cannot be fixed anymore. The following example was pointed out by Neil Thapen; it is a special case of a principle defined in [24]. Take a $k$-ary relation symbol $R(x_1, \ldots, x_k)$ and define $\Phi$ to be $\exists x_1 \forall x_2 \ldots R(x_1, \ldots, x_k)$, and $\Psi := \neg\Phi$. Any interpolating property thus defines $\Phi$ and the lower bound of [25, 9] implies that no depth $k - 1$ circuit can have size $2^{n^{o(1)}}$.


# Acknowledgements:

I am indebted to Emil Jeřábek (Prague) for comments on a draft of the paper. I thank for discussions on the topic to the French Café circle: Phuong Nguyen, Pavel Pudlák, Neil Thapen and Iddo Tzameret.

# References


[1] M. Ajtai, The complexity of the pigeonhole principle, in: *Proc. IEEE 29$^{th}$ Annual Symp. on Foundation of Computer Science*, (1988), pp. 346-355.

[2] M. L. Bonet, C. Domingo, R. Gavalda, A. Maciel, and T. Pitassi, Non-automatizability of bounded-depth Frege proofs, *Computational complexity*, **13**, (2004), pp.47-68.

[3] M. L. Bonet, T. Pitassi, and R. Raz, Lower bounds for cutting planes proofs with small coefficients, *J. of Symbolic Logic*,(1997), pp.708-728.

[4] M. L. Bonet, T. Pitassi, and R. Raz, On interpolation and automatization for Frege systems, *SIAM J,Computing*, **29(6)**, (2000), pp.1939-1967.

[5] S. R. Buss, *Bounded Arithmetic*, (1986). Naples, Bibliopolis.

[6] S. R. Buss, The propositional pigeonhole principle has polynomial size Frege proofs, *J. Symbolic Logic*, **52**, (1987), pp.916-927.

[7] S. A. Cook and P. Nguyen, *Foundations of Proof Complexity: Bounded Arithmetic and Propositional Translations*, book in preparation (a draft is available via web: `http://www.cs.toronto.edu/ sacook/csc2429h/book`).

[8] S. A. Cook, and A. R. Reckhow, The relative efficiency of propositional proof systems, *J. Symbolic Logic*,**44(1)**, (1979), pp.36-50.

[9] J. Hastad, Almost optimal lower bounds for small depth circuits. in: *Randomness and Computation*, ed. S.Micali, Ser.Adv.Comp.Res., (1989), **5**: 143-170. JAI Pres.

[10] J. Krajíček, Lower bounds to the size of constant-depth propositional proofs, *Journal of Symbolic Logic*, **59(1)**, (1994), pp.73-86.

[11] J. Krajíček, *Bounded arithmetic, propositional logic, and complexity theory*, Encyclopedia of Mathematics and Its Applications, Vol. **60**, Cambridge University Press, (1995).

[12] J. Krajíček, Interpolation theorems, lower bounds for proof systems, and independence results for bounded arithmetic, *J. of Symbolic Logic*, **62(2)**, (1997), pp.457-486.

[13] J. Krajíček, Discretely ordered modules as a first-order extension of the cutting planes proof system, *J. Symbolic Logic*, **63(4)**, (1998), pp.1582-1596.

[14] J. Krajíček, An exponential lower bound for a constraint propagation proof system based on ordered binary decision diagrams, *J. of Symbolic Logic*, **73(1)**, (2008), pp.227-237.

[15] J. Krajíček, and P. Pudlák, Some consequences of cryptographical conjectures for $S^1_2$ and $EF$", in: *Logic and Computational Complexity* (Proc. of the meeting held in Indianapolis, October 1994), Ed. D. Leivant, Springer-Verlag, Lecture Notes in Computer Science, Vol. **960**, (1995), pp.210-220.

Revised version in: *Information and Computation*, Vol. **140 (1)**, (January 10, 1998), pp.82-94.

[16] J. Krajíček, P. Pudlák,and A. Woods, An Exponential Lower Bound to the Size of Bounded Depth Frege Proofs of the Pigeonhole principle", *Random Structures and Algorithms*, **7(1)**, (1995), pp.15-39.

[17] J. B. Paris and A. J. Wilkie, Counting problems in bounded arithmetic, in: *Methods in Mathematical Logic*, LNM 1130, (1985), pp. 317-340. Springer-Verlag.

[18] J. B. Paris, A. J. Wilkie, and A. R. Woods, Provability of the pigeonhole principle and the existence of infinitely many primes, *Journal of Symbolic Logic*, **53**, (1988), pp. 1235-1244.

[19] T. Pitassi, P. Beame, and R. Impagliazzo, Exponential lower bounds for the pigeonhole principle, *Computational complexity*, **3**, (1993), pp.97-308.

[20] P. Pudlák, Lower bounds for resolution and cutting plane proofs and monotone computations, *J. of Symbolic Logic*,(1997), pp.981-998.

[21] P. Pudlák, The lengths of proofs, in: Handbook of Proof Theory, S.R. Buss ed., Elsevier, (1998), pp.547-637.

[22] P. Pudlák and J. Sgall, Algebraic models of computation and interpolation for algebraic proof systems, in: *Proof complexity and feasible arithmetic*, eds.P. Beame and S. R. Buss, DIMACS Ser. Discrete Math and Comp.Sci. **39**, AMS, (1998), pp. 179-205.

[23] A. A. Razborov, Unprovability of lower bounds on the circuit size in certain fragments of bounded arithmetic, *Izvestiya of the R.A.N.*, **59(1)**, (1995), pp.201-224.

[24] A. Skelley and N. Thapen, The provably total search problems of bounded arithmetic, submitted (preprint 2008).

[25] A. Yao, Separating the polynomial-time hierarchy by oracles, in: *Proc. 26th Ann. IEEE Symp. on Found. of Comp. Sci.*, (1985), pp. 1-10.

**Mailing address:**
Department of Algebra
Faculty of Mathematics and Physics
Charles University
Sokolovská 83, Prague 8, CZ - 186 75
The Czech Republic
krajicek@math.cas.cz