

Arthur Merlin Games

If the tosses are public, this is called Arthur Merlin (AM) games.

Arthur is the verifier and Merlin is the prover.

$MAM \dots$ denotes the sequence of communication where M starts first, then A tosses coins, then M responds, and so on. At the end verifier accepts or rejects.

So, for example in MAM , on input x , prover generates a string y , then Arthur generates a random string z and then prover generates a string w and then at the end Arthur runs a polynomial time computable function $P(x, y, z, w)$ to accept or reject.

$AM[k]$ denotes $AMAMAM \dots AM$ (AM appears k times).

Theorem: $MA \subseteq AM$

Proof: First Boost probability. Then note that most of the “guesses” work for every “bad” merlin move.

Suppose the check made by the verifier is $R(x, y, z)$, where y is the string provided by Merlin and z is a random guess.

Without loss of generality assume that $|y| = p(|x|)$ and probability of error is $< 2^{-p(|x|)-2}$.

Note that if $x \in L$, then Merlin can always give the correct y even if Arthur goes first. So the interesting case is when $x \notin L$. Thus, the fraction of strings z which fail to catch the prover for any particular y provided by Merlin on input x is at most $2^{-p(|x|)-2}$. Thus, the total fraction of strings z which fail for some y is at most 2^{-2} . So, even if Arthur goes first, $1 - 2^{-2}$ fraction of z will catch a cheating Merlin.

Theorem: $MAM \subseteq AM$

Proof: Similar idea as used for $MA \subseteq AM$.

Suppose the check made by the verifier is $R(x, y, z, w)$, where y is the string provided by Merlin followed by z as the random guess and then w as the string provided by Merlin.

Without loss of generality assume that $|y| = p(|x|)$ and probability of error is $2^{-p(|x|)-2}$. Note that w is anyway provided by Merlin after knowing z .

Again, if $x \in L$, then Merlin can provide the two strings y and w honestly. If $x \notin L$, then the fraction of strings z which fail to catch Merlin for any particular y , for some w , is at most $2^{-p(|x|)-2}$. Thus, the total fraction of strings z which fail for some y for some w is at most 2^{-2} .

Theorem: $AM[k] \subseteq AM$

Proof: Repeatedly using above method (to replace MAM at the end to AM). Also note that two consecutive AA can be converted to A, as one can just do both the coin tosses together.

Theorem: $AM \subseteq \Pi_2^p$.

Proof: Suppose $L \in AM$. Assume the error probability is reduced.

Now note that in the proof of $BPP \subseteq \Sigma_2^p$, we only consider the “successful” and “unsuccessful” strings with respect to acceptance. Furthermore, BPP is closed under complementation. Thus, the same proof also shows $BPP \subseteq \Pi_2^p$.

Thus, using the same trick we can replace the verifier check by universal followed by existential quantifier.

Theorem: $MA \subseteq \Sigma_2^p \cap \Pi_2^p$.

$MA \subseteq \Pi_2^p$ follows from $MA \subseteq AM$.

$MA \subseteq \Sigma_2^p$ follows using the $BPP \subseteq \Sigma_2^p$ trick.

Theorem: If $coNP \subseteq AM$, then $coAM \subseteq AM$.

Proof: Languages in $coAM$ are of form:

$x \in L$ implies (for most z)(for all y)[$R(x, y, z)$]

$x \notin L$ implies (for most z) \neg (for all y)[$R(x, y, z)$]

Now using $coNP \subseteq AM$, we can replace above by

$x \in L$ implies (for most z)(for most w)(exists u)[$R'(x, z, w, u)$]

$x \notin L$ implies (for most z)(for most w) \neg (exists

u)[$R'(x, z, w, u)$]

which is in AM .

Theorem: If $coNP \subseteq AM$, then $PH \subseteq AM \subseteq \Pi_2^p$.

Proof:

By induction we show that $\Sigma_k^p \subseteq AM$.

For $k = 1$, this is already done.

For $k > 1$, suppose we have shown that $\Sigma_{k-1}^p \subseteq AM$.

Then consider $L \in \Sigma_k^p$. This can be expressed as $x \in L$ iff $(\exists y)[(x, y) \in L']$ for some $L' \in \Pi_{k-1}^p$.

Here $L' \in coAM$ (as $\Sigma_{k-1}^p \subseteq AM$).

Thus, $L' \in AM$. Which gives $L \in MAM \subseteq AM$.

Public vs Private Tosses:

Theorem: $IP(q(n)) \subseteq AM(q(n) + 2)$